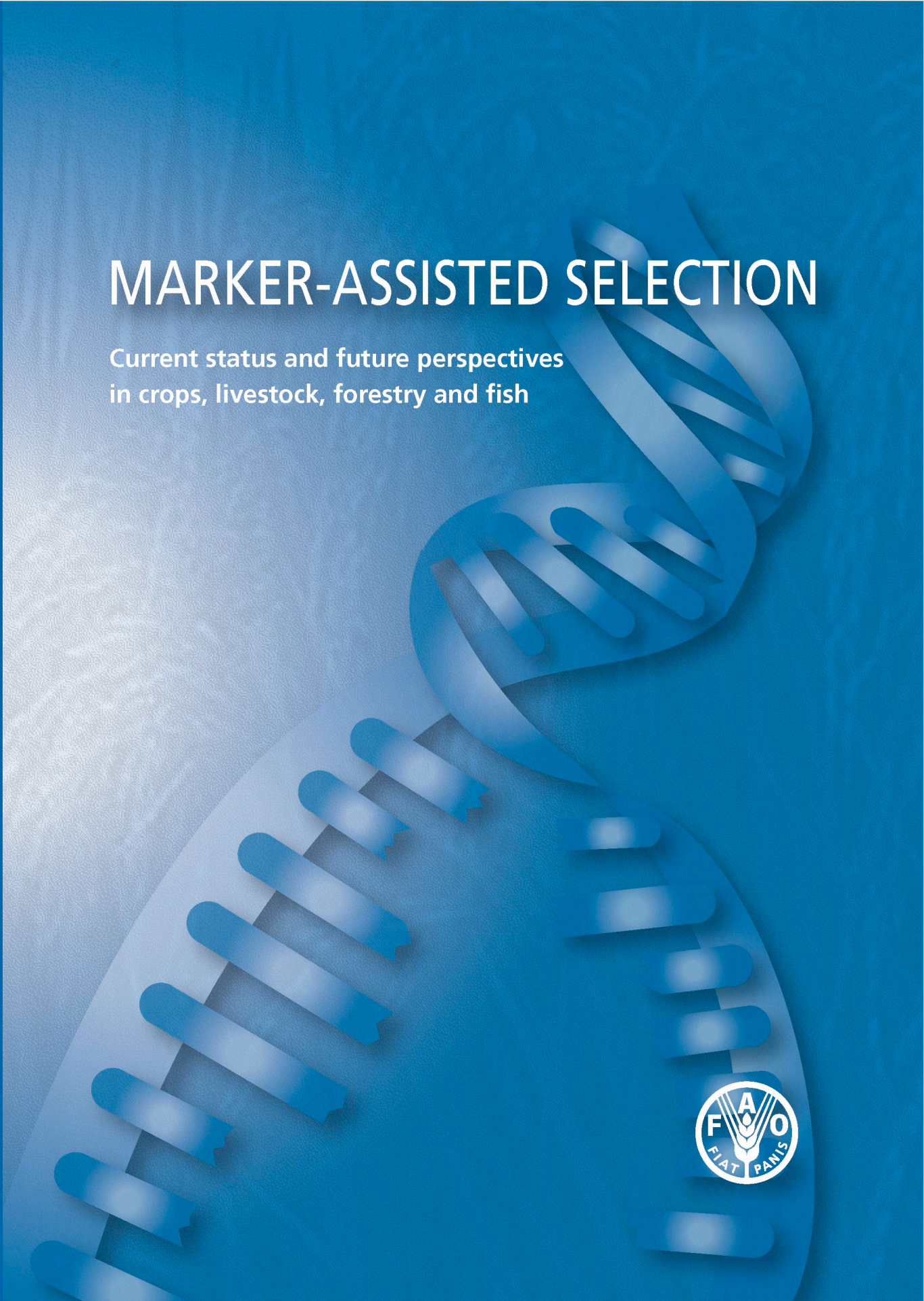
MARKER-ASSISTED SELECTION

Current status and future perspectives in crops, livestock, forestry and fish



MARKER-ASSISTED SELECTION

Current status and future perspectives in crops, livestock, forestry and fish

Edited by

Elcio P. Guimarães
John Ruane
Beate D. Scherf
Andrea Sonnino
James D. Dargie

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2007

Reprint 2009

ISBN 978-92-5-105717-9

Contents

Acknowledgements

Preparation of this major publication on marker-assisted selection would not have been possible without the committed support of a large number of individuals from the time that the idea of this publication was initially raised in early 2004. First and foremost among these are the members of the FAO Working Group on Biotechnology, in particular its former Chairperson, Jim Dargie, and his successor, Shivaji Pandey. Three of the members, Devin Bartley (FAO Aquaculture Management and Conservation Service), Pierre Sigaud (formerly of FAO Forest Resources Development Service) and Nuria Urquia (FAO Seed and Plant Genetic Resources Service), as well as Paul Boettcher (Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture) and Eric Hallerman (Virginia Polytechnic Institute and State University, United States of America) are particularly thanked for commenting on individual chapters in the book.

Thanks are also extended to Adrianna Gabrielli, Michela Paganini and Chrissi Redfern for their assistance with editorial and layout aspects of this publication.

Finally, special thanks go to all the authors for dedicating their time and efforts to the preparation of these chapters in an exemplary and timely fashion and for their meticulous attention to detail in the editing process.

Foreword

Since almost the beginning of human civilization, exploiting variation in the characteristics of the plant and animal genetic resources that are used for producing food and other agricultural products through breeding has been at the heart of efforts to increase and diversify agricultural production and productivity, enhance food security and incomes, and adapt farming to changing environmental conditions and social needs. Initially, this was achieved simply by selecting and reproducing preferred individuals or spontaneous variants, and indeed this practice remains important today as the basis for producing new generations of cultivated landraces and indigenous breeds. However, the crops, trees, livestock and fish that are farmed today have arisen largely from the introduction of scientific breeding at the beginning of the twentieth century, with the inclusion of crosses into breeding schemes prior to artificial selection and application of Mendel's laws of inheritance to improve both simple and quantitative traits providing the foundations for modern genetics.

Today, thanks to continuing investments made in research and technology development, the process of producing improved varieties, clones, breeds and strains of agriculturally important species has become progressively more accurate, reliable and efficient. Nevertheless, one of the continuing technical constraints to more effective breeding is that selecting material with one or a combination of the characteristics required by farmers, foresters, industry and consumers still relies mainly on physical and agronomic attributes (phenotype). Some of these characteristics are influenced by the environment and are therefore not necessarily a good guide to the actual heritable genetic composition (genotype) of the material in question. Others may not be visible or may only be detected in mature plants and animals. Others again may be difficult or very costly to screen, and many characters such as drought tolerance and milk composition are controlled by a large number of genes whose mode of action as well as their interaction with each other and with various environmental triggers is mainly unknown. Improving the identification, selection and monitoring of specific characters in plants and animals through breeding schemes is therefore a critical need to secure future improvements in genetic resources for food and agriculture.

Since the first description of DNA structure over 50 years ago, scientists have made tremendous strides in identifying genes and gene functions, making it increasingly possible to detect genetic differences (DNA polymorphisms) for traits among individual plants and animals in a much more direct way, thereby assisting in the selection of desired traits. The central technology involved is molecular marker-assisted selection (MAS), using sequences and/or banding patterns of DNA that have been shown through linkage mapping to be located in or near genes that affect the phenotype. These molecular markers can then be used to assist breeders track whether

the specific gene or chromosome segment(s) known to affect the phenotype of interest is present in the individuals or populations of interest.

Although the ultimate goal of identifying the location, function and most favourable alleles of each gene through genome sequence and post-genomics research, and then using markers to select for economically important genes in breeding programmes, is still decades away, in recent years the use of MAS in agriculture has moved progressively from theory to practical application. In the process, it has generated both high expectations for increasing genetic progress through breeding, and raised a number of unresolved challenges. These include: selection of the most appropriate methods and tools for MAS among the many now available for the task at hand, analysing and managing the data produced given the increasing trend towards high-throughput techniques and the constraints imposed by suboptimal levels of resources currently attached to breeding and science and technology including biotechnology, and dealing with intellectual property rights, especially in developing countries.

Since its foundation, FAO has recognized that the biological basis for sustainable agricultural production, fighting hunger and world food security lies in the genetic resources used for food and agriculture. It has also recognized the enormous contributions that have been made to the improvement of these resources through both traditional and more advanced breeding, as well as the ever-increasing role played by biotechnology in improving breeding processes and products. As a knowledge organization, one of FAO's major roles is to provide its Members and their institutions with factual, comprehensive and current information relevant to sound stewardship of crops, livestock, forestry and fisheries, thereby ensuring its availability as a global public good. This book, by providing a comprehensive description and assessment of the use of MAS for increasing the rate of genetic gain in crops, livestock, forestry and farmed fish, including the related policy, organizational and resource considerations, continues the Organization's tradition of dealing with issues of importance to agricultural and economic development in a multidisciplinary and cross-sectoral manner. As such it is hoped that the information and options presented and the suggestions made will provide valuable guidance to scientists and breeders in both the public and private sectors, as well as to government and institutional policy- and decision-makers.

Shivaji Pandey
Chairperson
FAO Working Group on Biotechnology

Abbreviations and acronyms

AATF	African Agricultural Technology Foundation
AB-QTL	Advanced backcross QTL
ACMV	African cassava mosaic virus
AFLP	Amplified fragment length polymorphism
AI	Artificial insemination
AMBIONET	Asian Maize Biotechnology Network
AMMANET	African Molecular Marker Applications Network
AnGR	Animal genetic resources
ASARECA	Association for Strengthening Agricultural Research in Eastern and Central Africa
BAC	Bacterial artificial chromosome
BCMNV	Bean common mosaic necrotic virus
BCMV	Bean common mosaic virus
BecA	Biosciences eastern and central Africa
BGYMV	Bean golden yellow mosaic virus
BIO-EARN	East African Regional Programme and Research Network for Biotechnology, Biosafety and Biotechnology Policy Development
BLUP	Best linear unbiased prediction
bp	Base pairs
BSAPs	Biodiversity Strategies and Action Plans
Bt	*Bacillus thuriengensis*
BTA	*Bos taurus* chromosome
BYDV	Barley yellow dwarf virus
CAADP	Comprehensive Africa Agriculture Development Programme
CAPS	Cleaved amplified polymorphic sequences
CBB	Cassava bacterial blight
CBS	Cassava brown streak
CBSD	Cassava brown streak disease
CCN	Cereal cyst nematode
cDNA	Complementary DNA
CGIAR	Consultative Group on International Agricultural Research
CGM	Cassava green mite
CI	Confidence interval

CIAT	International Center for Tropical Agriculture (Centro Internacional de Agricultura Tropical)
CIMMYT	International Maize and Wheat Improvement Center (Centro Internacional de Mejoramiento de Maíz y Trigo)
CIP	International Potato Center Centro (Internacional de la Papa)
CIRAD	French Agricultural Research Centre for International Development (Centre de coopération internationale en recherche agronomique pour le développement)
cM	Centi-Morgan
CMD	Cassava mosaic disease
CMV	Cassava mosaic virus
CORPOICA	Colombian Agricultural Research Corporation (Corporación Colombiana de Investigación Agropecuaria)
CR	Country report
CT	Computer tomography
DArT	Diversity array technology
DFID	United Kingdom's Department for International Development
DH	Double-haploid
DHPLC	Denaturing high pressure liquid chromatography
DMC	Dry matter content
DNA	Deoxyribonucleic acid
DYD	Daughter yield deviation
EACMV	East Africa cassava mosaic virus
EBV	Estimated breeding value
EC	European Commission
ECOSOC	Economic and Social Council of the United Nations
eQTL	Expressed gene QTL
EST	Expressed sequence tag
ESTP	Expressed sequence tagged polymorphism
EU	European Union
EUCAGEN	*Eucalyptus* Genome Network
F_1	First filial generation
F_2	Second filial generation
FAO	Food and Agriculture Organization of the United Nations
FAO-BioDeC	FAO Biotechnology in Developing Countries
FHB	*Fusarium* head blight

FIVIMS	Food Insecurity and Vulnerability Information and Mapping Systems
FNP	Functional nucleotide polymorphism
FSC	Forest Stewardship Council
FSIL	Full-sib intercross line
GABI	Genome analysis of the plant biological system
GAS	Gene-assisted selection
GCA	General combining ability
GCP	Generation Challenge Programme
GDP	Gross domestic product
GE	Genetic engineering
GH	Growth hormone
GIS	Geographical information systems
GMOs	Genetically modified organisms
GRDC	Grains Research and Development Corporation
GRFA	Genetic resources for food and agriculture
GRM	Gametic relationship matrix
h^2	Heritability
HIPC	Heavily indebted poor countries
HWE	Hardy-Weinberg equilibrium
IAC	InterAcademy Council
IAP	InterAcademy Panel
IARCs	International agricultural research centres
IBD	Identity by descent
ICAR	Indian Council for Agricultural Research
ICMV	Indian cassava mosaic virus
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
ICSU	International Council for Science
IFPRI	International Food Policy Research Institute
IHN	Infectious haematopoietic necrosis
IITA	International Institute of Tropical Agriculture
ILRI	International Livestock Research Institute
INIFAP	National Institute for Forestry, Agriculture and Livestock Research (Instituto Nacional de Investigaciones Forestales y Agropecuarias)
IP	Intellectual property
IPGRI	International Plant Genetic Resource Institute
IPR	Intellectual property right
IRR	Internal rate of return
IRRI	International Rice Research Institute
ISAG	International Society for Animal Genetics

ISNAR	International Service for National Agricultural Research
ISSR	Inter-simple sequence repeats
ITPGRFA	International Treaty on Plant Genetic Resources for Food and Agriculture
JGI	Joint Genome Institute
KARI	Kenya Agricultural Research Institute
LD	Linkage disequilibrium
LDL	Linkage disequilibrium and linkage
LD-MAS	Linkage disequilibrium MAS
LE	Linkage equilibrium
LE-MAS	Linkage equilibrium MAS
LIMS	Laboratory information management system
LOD	Logarithm of the odds ratio
MABC	Marker-assisted back-crossing
MA-BLUP	Marker-assisted best linear unbiased prediction
MAI	Marker-assisted introgression
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
MARS	Marker-assisted recurrent selection
MAS	Marker-assisted selection
MBL	Medical biotechnology laboratories
MC	Molecular characterization
MD	Marek's disease
MDG	Millennium Development Goals
MFA	Microfibril angle
MHC	Major histocompatibility complex
miRNA	MicroRNA
ML	Maximum likelihood
MoDAD	Measurement of domestic animal diversity
mRNA	Messenger RNA
MSV	Maize streak virus
MTA	Material Transfer Agreement
NARES	National agricultural research and extension systems
NARS	National agricultural research systems
NDA	Non-disclosure agreement
NEPAD	New Partnership for Africa's Development
NGO	Non-governmental organization
NIRS	Near infrared reflectance spectroscopy
NPV	Net present value
NUE	Nitrogen use efficiency

OBM	Orange blossom midge
OECD	Organisation for Economic Co-operation and Development
OIE	World Organisation for Animal Health
OPV	Open-pollinated variety
PAGE	Polyacrylamide gel electrophoresis
PBRs	Plant breeders' rights
PCR	Polymerase chain reaction
PGRFA	Plant genetic resources for food and agriculture
PIC	Polymorphic information content
PPB	Participatory plant breeding
PPD	Post-harvest physiological deterioration
PRSPs	Poverty reduction strategy papers
PT	Progeny test
PVP	Plant variety protection
QPM	Quality protein maize
QTL	Quantitative trait loci (or locus)
QTL-NILs	Near isogenic lines for QTL
QTN	Quantitative trait nucleotide
R&D	Research and development
RAPD	Random amplified polymorphic DNA
rDNA	Ribosomal DNA
RFLP	Restriction fragment length polymorphism
RGA	Resistance gene analogues
RNA	Ribonucleic acid
RRA	Rapid rural appraisal
S&T	Science and technology
SACMV	South African cassava mosaic virus
SAGE	Serial analysis of gene expression
SBMV	Soil-borne mosaic virus
SCA	Specific combining ability
SCAR	Sequence characterized amplified region
SCN	Soybean cyst nematode
SCS	Somatic cell score
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
siRNA	Short interfering RNA
SLS-MAS	Single large-scale MAS
SLU	Swedish University of Agricultural Sciences
SMA	Simple marker analysis
SNP	Single nucleotide polymorphism

SoW-AnGR	State of the World's Animal Genetic Resources
SPS Agreement	WTO Agreement on the Application of Sanitary and Phytosanitary Measures
SSCP	Single strand conformation polymorphism
SSLP	Simple sequence length polymorphism
SSR	Simple sequence repeat (syn. microsatelllite)
STB	*Septoria tritici* blotch
STS	Sequence-tagged sites
SW	Seed weight
SWaps	Sector-wide approaches
TBT Agreement	WTO Agreement on Technical Barriers to Trade
TC	Tissue culture
TEs	Transposable elements
TMV	Tobacco mosaic virus
ToMV	Tomato mottle virus
TRIPS Agreement	WTO Agreement on Trade-Related Aspects of Intellectual Property Rights
TSWV	Tomato spotted wilt virus
TUA	Technology Use Agreement
TYLCV	Tomato yellow leaf curl virus
UN	United Nations
UPOV	International Union for the Protection of New Varieties of Plants
USAID	United States Agency for International Development
USDA	United States Department of Agriculture
WEC	Worm egg count
WFS	World Food Summit
WIPO	World Intellectual Property Organization
WRI	World Resources Institute
WSC	Wood specific consumption
WTO	World Trade Organization
YMV	Yellow mosaic virus

Contributors

Amalia Barone
Professor in Plant Genetics
University of Naples "Federico II"
Department of Soil, Plant, Environmental and Animal Production Sciences
Via Universita' 100
80055 Portici, Italy
e-mail: ambarone@unina.it

Roswitha Baumung
Researcher, Division Livestock Sciences
University of Natural Resources and Applied Life Sciences (BOKU),
Gregor Mendelstr. 33
1180 Vienna, Austria
e-mail: roswitha.baumung@boku.ac.at

Steve E. Beebe
Bean Breeder
Bean Project Manager
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali, Colombia
e-mail: s.beebe@cgiar.org

Mathew W. Blair
Germplasm Specialist/Bean Breeder
Biotechnology Unit
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali, Colombia
e-mail: m.blair@cgiar.org

Penny Butcher
Conservation Research Scientist
Kings Park and Botanic Garden
Botanic Gardens and Parks Authority
Fraser Avenue, West Perth 6005
Western Australia, Australia
e-mail: pbutcher@bgpa.wa.gov.au

Marcelo J. Carena
Associate Professor/Director
Corn Breeding and Genetics
North Dakota State University
Department of Plant Sciences
Loftsgard Hall 166
Fargo, ND 58105-5051, USA
e-mail: marcelo.carena@ndsu.edu

Hernán Ceballos
Cassava Breeding
Cassava Breeding Project
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali, Colombia
and Universidad Nacional de Colombia
e-mail: h.ceballos@cgiar.org

James D. Dargie
Former Director, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture
Brunnstubengasse 43
2102 Bisamberg, Austria
e-mail: j.dargie@aon.at

Jack C. M. Dekkers
Professor and Section Leader
Animal Breeding and Genetics
Department of Animal Science
239D Kildee Hall
Iowa State University
Ames, IA 50011-3150, USA
e-mail: jdekkers@iastate.edu

Dirk-Jan de Koning
Principal Investigator – Genetics and Genomics
The Roslin Institute
Roslin Biocentre
Roslin, Midlothian
EH25 9PS, UK
e-mail: DJ.deKoning@bbsrc.ac.uk

Christel Denneboom
Plant Research International
PO Box 16
6700 AA Wageningen
Netherlands
e-mail: christel.denneboom@wur.nl

Ab L.F. de Vos
Plant Research International
PO Box 16
6700 AA Wageningen
Netherlands
e-mail: ab.devos@wur.nl

Oene Dolstra
Senior Scientist
Plant Research International
PO Box 16
6700 AA Wageningen
Netherlands
e-mail: oene.dolstra@wur.nl

Jeremy D. Edwards
Research Associate
University of Arizona, Department of Plant Sciences
821B Marley Building, 1145 E 4 ST.
Tucson, AZ 85721, USA
e-mail: jdedw@cals.arizona.edu

Martin Fregene
Senior Scientist and Cassava Geneticist
Centro Internacional de Agricultura Tropical (CIAT)
Km 17, Recta Cali-Palmira
A.A. 6713, Cali
Colombia
e-mail: m.fregene@cgiar.org

Luigi Frusciante
Professor in Plant Genetics
University of Naples "Federico II"
Department of Soil, Plant, Environmental and Animal Production Sciences
Via Universita' 100
80055 Portici
Italy
e-mail: fruscian@unina.it

Marc Giband
Researcher - Cotton Molecular Genetics
CIRAD - UPR Systemes Cotonniers / EMBRAPA Algodão
Rua Osvaldo Cruz 1143
Centenario
58.107-720 Campina Grande, PB
Brazil
e-mail: marc.giband@cirad.fr

Dario Grattapaglia
Senior Scientist and Professor - Plant Genetics/Genomics
Brazilian Agricultural Research Corporation – Embrapa
Genetic Resources and Biotechnology Center
Av. W5 Norte - Final - Plano Piloto
70770-900 - Brasília, DF, Brazil
and
Genomic Science Program
Universidade Católica de Brasília
SGAN 916 Modulo B
70790-160 Brasília, DF, Brazil
e-mail: dario@cenargen.embrapa.br

Elcio Perpétuo Guimarães
Senior Officer - Cereals/Crops Breeding
Food and Agriculture Organization of the United Nations (FAO)
00153 Rome, Italy
e-mail: elcio.guimaraes@fao.org

Bernard Hau
Doctor in Plant breeding
CIRAD
UPR Systèmes Cotonniers
TA B-10/02
Avenue Agropolis
34980 Montpellier Cedex 5
France
e-mail: bernard.hau@cirad.fr

Victoria W.K. Henson-Apollonio
Senior Scientist, Project Manager
The CGIAR Central Advisory Service on Intellectual Property (CAS-IP)
Bioversity International
Via dei Tre Denari, 472/a
00057 Maccarese
Italy
e-mail: v.henson-apollonio@cgiar.org

Paul M. Hocking
Principal Investigator - Genetics and Genomics
The Roslin Institute
Roslin Biocentre, Roslin
Midlothian EH25 9PS, UK
e-mail: Paul.Hocking@bbsrc.ac.uk

David A. Hoisington
Global Theme Leader - Biotechnology
ICRISAT
Patancheru, 502 324 Andhra Pradesh, India
e-mail: d.hoisington@cgiar.org

Robert Koebner
Crop Breeding and Biotechnology Consultant
CropGen International
Mockbeggars, Townhouse Road
Old Costessey
Norwich NR8 5BX, UK
e-mail: mockbeggars@onetel.com

Jean-Marc Lacape
Cotton Molecular Breeding
CIRAD
UMR Développement et Amélioration des Plantes
TA A-96/03
Avenue Agropolis
34980 Montpellier Cedex 5, France
e-mail: marc.lacape@cirad.fr

Michael Lee
Professor of Plant Breeding and Genetics
Department of Agronomy
Iowa State University
Room 1553 Agronomy Hall
100 Osborn Drive
Ames, IA 50011-1010, USA
e-mail: mlee@iastate.edu

Maurício Antônio Lopes
Research Scientist - Plant Genetics
Brazilian Agricultural Research Corporation – Embrapa
Genetic Resources and Biotecnology Center
Av. W5 Norte - Final - Plano Piloto
70770-900 - Brasília, DF, Brazil
e-mail: mlopes@cenargen.embrapa.br

Victor Martinez
Assistant Professor
Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile
Avda. Santa Rosa 11735
La Pintana, Santiago, Chile
e-mail: vmartine@uchile.cl

Susan R. McCouch
Professor
Plant Breeding & Genetics
162 Emerson Hall
Cornell University
Ithaca, NY 14853-1901
USA
e-mail: srm4@cornell.edu

Michael Morris
Lead Agricultural Economist
The World Bank - MSN 6-602
1818 Street, NW.
Washington, DC 20433
USA
e-mail: mmorris3@worldbank.org

Trung-Bieu Nguyen
Cotton Breeding
CIRAD
UPR Systèmes Cotonniers
TA B-10/02
Avenue Agropolis
34980 Montpellier Cedex 5
France

Dafydd Pilling
Editor State of the World AnGR
Food and Agriculture Organization of
the United Nations (FAO)
00153 Rome
Italy
e-mail: dafydd.pilling@fao.org

Michel Ragot
Head - Genetic Information Management
Syngenta Seeds
12 chemin de l'Hobit
31790 Saint-Sauveur
France
e-mail: michel.ragot@syngenta.com

Barbara Rischkowsky
At that time:
Coordinator State of the World AnGR
Food and Agriculture Organization of
the United Nations (FAO)
00153 Rome, Italy
Currently:
Senior Livestock Scientist
International Center for Agricultural
Research in the Dry Areas (ICARDA)
PO Box 5466
Aleppo, Syrian Arab Republic
e-mail: b.rischkowsky@cgiar.org

Jonathan Robinson
Consultant
Tick-aho, Joroisniemenkehätie
79600 Joroinen, Finland
e-mail: jrobinson@tiscali.it

John Ruane
Agricultural Officer (Biotechnology)
FAO Working Group on Biotechnology
Food and Agriculture Organization of
the United Nations (FAO)
00153 Rome, Italy
e-mail: john.ruane@fao.org

Beate D. Scherf
Animal Production Officer
Animal Genetic Resources Group
Food and Agriculture Organization of
the United Nations (FAO)
00153 Rome, Italy
e-mail: beate.scherf@fao.org

Anna K. Sonesson
Senior Scientist, Genetics and Breeding
AKVAFORSK (Institute of Aquaculture
Research AS)
PO Box 5010
1432 Ås, Norway
e-mail: anna.sonesson@akvaforsk.no

Andrea Sonnino
Senior Agricultural Research Officer
Research and Technological
Development Service
Food and Agriculture Organization of
the United Nations (FAO)
00153 Rome, Italy
e-mail: andrea.sonnino@fao.org

Simon Southerton
Senior Research Scientist
Ensis Genetics
The joint forces of CSIRO and SCION
PO Box E4008 Kingston ACT 2604,
Australia
e-mail: simon.southerton@ensisjv.com

Richard Summers
Cereal Breeding Lead
RAGT Seeds Ltd
The Maris Centre
45 Hauxton Road
Trumpington
Cambridge CB2 2LQ, UK
e-mail: RSummers@ragt.fr

Julius H. J. van der Werf
Professor in Animal Breeding and
Genetics
School of Rural Science and Agriculture
University of New England
Armidale, NSW 2351, Australia
e-mail: julius.vanderwerf@une.edu.au

E.N. (Robert) van Loo
Plant Research International
PO Box 16
6700 AA Wageningen, Netherlands
e-mail: robert.vanloo@wur.nl

Marilyn Warburton
Senior Scientist, Molecular Geneticist
Genetic Resources Enhancement Unit
International Maize and Wheat
Improvement Center (CIMMYT)
Apdo. Postal 6-641
06600 Mexico, DF, Mexico
e-mail: m.warburton@cgiar.org

Joel Ira Weller
Research Scientist
Department of Cattle and Genetics
Institute of Animal Sciences
Agricultural Research Organization
The Volcani Center
PO Box 6
Bet Dagan 50250, Israel
e-mail: weller@agri.huji.ac.il

H. Manilal William
Senior Scientist, Global Wheat Program
Genetic Resources Enhancement Unit
International Maize and Wheat
Improvement Center (CIMMYT)
Apdo. Postal 6-641
06600 México, DF, Mexico
e-mail: m.william@cgiar.org

SECTION I

Introduction to marker-assisted selection

CHAPTER 1

Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: an overview of the issues

John Ruane and Andrea Sonnino

SUMMARY

This chapter provides an overview of the techniques, current status and issues involved in using marker-assisted selection (MAS) for genetic improvement in developing countries. Molecular marker maps, the necessary framework for any MAS programme, have been constructed for the majority of agriculturally important species, although the density of the maps varies considerably among species. Despite the considerable resources that have been invested in this field and despite the enormous potential it still represents, with few exceptions, MAS has not yet delivered its expected benefits in commercial breeding programmes for crops, livestock, forest trees or farmed fish in the developed world. When evaluating the potential merits of applying MAS as a tool for genetic improvement in developing countries, some of the issues that should be considered are its economic costs and benefits, its potential benefits compared with conventional breeding or with application of other biotechnologies, and the potential impact of intellectual property rights (IPRs) on the development and application of MAS.

INTRODUCTION

The potential benefits of using markers linked to genes of interest in breeding programmes, thus moving from phenotype-based towards genotype-based selection, have been obvious for many decades. However, realization of this potential has been limited by the lack of markers. With the advent of DNA-based genetic markers in the late 1970s, the situation changed and researchers could, for the first time, begin to identify large numbers of markers dispersed throughout the genetic material of any species of interest and use the markers to detect associations with traits of interest, thus allowing MAS finally to become a reality. This led to a whole new field of academic research, including the milestone paper by Paterson *et al*. (1988). This showed that with the availability of large numbers of genetic markers for their species of interest (tomato), the effects and location of marker-linked genes having an impact on a number of quantitative traits (fruit traits in their case) could be estimated using an approach that could be applied to dissect the genetic make-up of any physiological, morphological and behavioural trait in plants and animals.

Most of the traits considered in animal and plant genetic improvement programmes are quantitative, i.e. they are controlled by many genes together with environmental factors, and the underlying genes have small effects on the phenotype observed. Milk yield and growth rate in animals or yield and seed size in plants are typical examples of quantitative traits. In classical genetic improvement programmes, selection is carried out based on observable phenotypes of the candidates for selection and/or their relatives but without knowing which genes are actually being selected. The development of molecular markers was therefore greeted with great enthusiasm as it was seen as a major breakthrough promising to overcome this key limitation. As Young (1999) wrote: "Before the advent of DNA marker technology, the idea of rapidly uncovering the loci controlling complex, multigenic traits seemed like a dream. Suddenly, it was difficult to open a plant genetics journal without finding dozens of papers seeking to pinpoint many, if not most, agriculturally relevant genes." However, despite the considerable resources that have been invested in this field and despite the enormous potential it still represents, with few exceptions, MAS has not yet delivered its expected benefits in commercial breeding programmes for crops, livestock, forest trees or farmed fish in the developed world. In developing countries, where investments in molecular markers have been far smaller, delivery of benefits has lagged even further behind.

The focus of this chapter is on the use of molecular markers for genetic improvement of populations through MAS, including marker-assisted introgression. Its aim is to provide an easily understandable overview of the techniques, applications and issues involved in the use of DNA markers in MAS for genetic improvement of domestic plant and animal populations in developing countries. In the next section of the chapter, a brief description of the technical aspects of molecular markers and MAS is provided. The current status of the application of MAS in crops, forestry, livestock and fish is then summarized, while the final section

Note: This chapter is based on the Background Document to Conference 10 (on molecular marker-assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries) of the FAO Biotechnology Forum, 17 November–14 December 2003 (available at www.fao.org/biotech/C10doc.htm).

highlights issues that might be important to applications of MAS in developing countries. Although molecular markers may be used for a wide range of different tasks, such as to quantify the genetic diversity and relationships within and between agricultural populations (e.g. livestock breeds), to investigate biological processes (such as mating systems, pollen movement or seed dispersal in plants) or to identify specific genotypes (e.g. cloned forest trees), these applications are not considered here.

BACKGROUND TO MAS

Molecular markers

All living organisms are made up of cells that are programmed by genetic material called DNA. This molecule is made up of a long chain of nitrogen-containing bases (there are four different bases – adenine [A], cytosine [C], guanine [G] and thymine [T]). Only a small fraction of the DNA sequence typically makes up genes, i.e. that code for proteins, while the remaining and major share of the DNA represents non-coding sequences, the role of which is not yet clearly understood. The genetic material is organized into sets of chromosomes (e.g. five pairs in *Arabidopsis thaliana*; 30 pairs in *Bos taurus* [cow]), and the entire set is called the genome. In a diploid individual (i.e. where chromosomes are organized in pairs), there are two alleles of every gene – one from each parent.

Molecular markers should not be considered as normal genes as they usually do not have any biological effect. Instead, they can be thought of as constant landmarks in the genome. They are identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next. They rely on a DNA assay, in contrast to morphological markers that are based on visible traits, and biochemical markers that are based on proteins produced by genes.

Different kinds of molecular markers exist, such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs) markers, amplified fragment length polymorphisms (AFLPs), microsatellites and single nucleotide polymorphisms (SNPs). They may differ in a variety of ways – such as their technical requirements (e.g. whether they can be automated or require use of radioactivity); the amount of time, money and labour needed; the number of genetic markers that can be detected throughout the genome; and the amount of genetic variation found at each marker in a given population. The information provided to the breeder by the markers varies depending on the type of marker system used. Each has its advantages and disadvantages and, in the future, other systems are likely to be developed. More details on the individual marker systems are provided in Chapter 3.

From markers to MAS

The molecular marker systems described above allow high-density DNA marker maps (i.e. with many markers of known location, interspersed at relatively short intervals throughout the genome) to be constructed for a range of economically important agricultural species, thus providing the framework needed for eventual applications of MAS.

Using the marker map, putative genes affecting traits of interest can then be detected by testing for statistical associations between marker variants and any trait of interest. These traits might be genetically simple – for example, many traits for disease resistance in plants are controlled by one or a few genes (Young, 1999). Alternatively,

they could be genetically complex quantitative traits, involving many genes (i.e. so-called quantitative trait loci [QTL]) and environmental effects. Most economically important agronomic traits tend to fall into this latter category. For example, using 280 molecular markers (comprising 134 RFLPs, 131 AFLPs and 15 microsatellites) and recording populations of rice lines for various plant water stress indicators, phenology, plant biomass, yield and yield components under irrigated and water stress conditions, Babu *et al.* (2003) detected a number of putative QTL for drought resistance traits.

Having identified markers physically located beside or even within genes of interest, in the next step it is now possible to carry out MAS, i.e. to select identifiable marker variants (alleles) in order to select for non-identifiable favourable variants of the genes of interest. For example, consider a hypothetical situation where a molecular marker M (with two alleles M1 and M2), identified using a DNA assay, is known to be located on a chromosome close to a gene of interest Q (with a variant Q1 that increases yield and a variant Q2 that decreases yield), that is, as yet, unknown. If a given individual in the population has the alleles M1 and Q1 on one chromosome and M2 and Q2 on the other chromosome, then any of its progeny receiving the M1 allele will have a high probability (how high depends on how close M and Q are to each other on the chromosome) of also carrying the favourable Q1 allele, and thus would be preferred for selection purposes. On the other hand, those that inherit the M2 allele will tend to have inherited the unfavourable Q2 allele, and so would not be preferred for selection. With conventional selection which relies on phenotypic values, it is not possible to use this kind of information.

The success of MAS is influenced by the relationship between the markers and the genes of interest. Dekkers (2004) distinguished three kinds of relationship:

- The molecular marker is located within the gene of interest (i.e. within the gene Q, using the example above). In this situation, one can refer to gene-assisted selection (GAS). This is the most favourable situation for MAS since, by following inheritance of the M alleles, inheritance of the Q alleles is followed directly. On the other hand, these kinds of markers are the most uncommon and are thus the most difficult to find.
- The marker is in linkage disequilibrium (LD) with Q throughout the whole population. LD is the tendency of certain combinations of alleles (e.g. M1 and Q1) to be inherited together. Population-wide LD can be found when markers and genes of interest are physically very close to each other and/or when lines or breeds have been crossed in recent generations. Selection using these markers can be called LD-MAS.
- The marker is not in linkage disequilibrium (i.e. it is in linkage equilibrium [LE]) with Q throughout the whole population. Selection using these markers can be called LE-MAS. This is the most difficult situation for applying MAS.

The universal nature of DNA, molecular markers and genes means that MAS can, in theory, be applied to any agriculturally important species. Indeed, active research programmes have been devoted to building molecular marker maps and detecting QTLs for potential use in MAS programmes in a whole range of crop, livestock, forest tree and fish species. In addition, MAS can be applied to support existing conventional breeding programmes. These programmes use strategies such as: recurrent selection (i.e.

using within-breed or within-line selection, important in livestock); development of crossbreds or hybrids (by crossing several improved lines or breeds) and introgression (where a target gene is introduced from, for example, a low-productive line or breed (donor) into a productive line (recipient) that lacks the target gene (a strategy especially important in plants). See Dekkers and Hospital (2002) for more details. MAS can be incorporated into any one of these strategies (e.g. for marker-assisted introgression by using markers to accelerate introduction of the target gene). Alternatively, novel breeding strategies can be developed to harness the new possibilities that MAS raises.

CURRENT STATUS OF APPLICATIONS OF MAS IN AGRICULTURE

Below is a brief summary of the current status regarding application of MAS in the different agricultural sectors. For more details, a number of case studies for crops are presented in Section II of the book and for livestock, forestry and fish in Sections III, IV and V, respectively.

Crops

The promise of MAS has possibly been greeted with the most enthusiasm and expectation in this particular agricultural sector, stimulating tremendous investments in the development of molecular marker maps and research to detect associations between phenotypes and markers. Molecular marker maps have been constructed for a wide range of crop species. Information on major plant projects (such as the sequencing of the entire rice genome) can be found at www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html.

In a recent review, however, Dekkers and Hospital (2002) noted that "as theoretical and experimental results of QTL detection have accumulated, the initial enthusiasm for the potential genetic gains allowed by molecular genetics has been tempered by evidence for limits to the precision of the estimates of QTL effects", and that "overall, there are still few reports of successful MAS experiments or applications." They reported that marker-assisted introgression of known genes was widely used in plants, particularly by private breeding companies, whereas marker-assisted introgression of unknown genes had often proved to be less useful in practice than expected. As Young (1999) wrote: "even though marker-assisted selection now plays a prominent role in the field of plant breeding, examples of successful, practical outcomes are rare. It is clear that DNA markers hold great promise, but realizing that promise remains elusive."

There is also considerable divergence with respect to the applications of MAS among different crop species. For example, Koebner (2003) highlighted the relatively fast uptake of MAS in maize compared with wheat and barley, arguing that this largely reflected the breeding structure. Thus, whereas maize breeding is dominated in industrialized countries by a small number of large private companies that produce F_1 hybrids, a system allowing protection from farm-saved seed and competitor use, breeding for the other major cereal species is primarily by public sector organizations and most varieties are inbred pure breeding lines, a system allowing less protection over the released varieties. Progress in arable crops is nevertheless quite advanced compared with horticultural crop species such as apples and pears, where development of molecular marker maps has been slow and only few QTL have been detected (Tartarini, 2003), even if MAS can potentially be very

useful for genetic improvement of such long-cycle plants.

Livestock

Again, much effort has been put into the development of molecular marker maps in this sector. The first reported map in livestock was for chicken in 1992, which was quickly followed by the publication of maps for cattle, pigs and sheep. Since then, the search for useful markers has continued and further species have been targeted, including goat, horse, rabbit and turkey (see www.thearkdb.org/ for the current status regarding some major livestock species). Microsatellite markers have been of major importance.

Dekkers (2004) recently reviewed commercial applications of MAS in livestock and noted that several gene or marker tests are available on a commercial basis in different species and for different traits, and that the majority of uses involve GAS, where an important gene (e.g. responsible for a congenital defect) has been identified or, to a lesser degree, LD-MAS. He pointed out that documentation is poor since, although several genetic tests are available, the extent to which they are used in commercial applications is unclear, as is the manner in which they are used and whether their use leads to greater responses to selection. He concluded that "opportunities for the application of MAS exist, in particular for GAS and LD-MAS and, to a lesser degree, for LE-MAS because of greater implementation requirements. Regardless of the strategy, successful application of MAS requires a comprehensive integrated approach with continued emphasis on phenotypic recording programmes to enable QTL detection, estimation and confirmation of effects, and use of estimates in selection. Although initial expectations for the use of MAS were high, the current attitude is one of cautious optimism."

Forestry

As for crops, extensive efforts have been devoted to construction of molecular marker maps for the major commercial genera, such as eucalypts, pines and acacia. RFLPs, RAPDs, microsatellites and AFLPs have been extensively used. The Web site http://dendrome.ucdavis.edu/index.php provides updated information on the status regarding molecular marker maps in forestry.

Molecular maps have been used to locate markers associated with variation in forestry traits of commercial interest, such as growth, frost tolerance, wood properties, vegetative propagation, leaf oil composition and disease resistance. Since MAS allows early selection before traits of interest (e.g. wood quality) are expressed, a major incentive for using molecular techniques in tree breeding is to improve the rate of genetic gain by reducing the long generation interval However, Butcher (2003) noted that "MAS has yet to be incorporated in operational breeding programmes for plantation species" and she referred to the high costs of genotyping, the large family sizes required to detect QTL and the lack of knowledge of QTL interactions with genetic background, tree age and environment as explanatory factors.

In a recent review of biotechnology in forestry, Yanchuk (2002) also highlighted the potential advantage of early selection using MAS, but again pointed out that MAS is not yet being applied routinely in tree breeding programmes, largely "because of economic constraints (i.e. the additional genetic gains are generally not large enough to offset the costs of applying the technology). Thus it is likely that MAS will only be applied for a handful of species and situations, e.g. a few

of the major commercially used pine and *Eucalyptus* species. Molecular markers are therefore primarily an information tool and are used to locate DNA/genes that can be of interest for genetic transformation, or information on population structure, mating systems and pedigree confirmation."

Fish

Molecular marker maps have been constructed for a number of aquaculture species, e.g. tilapia, catfish, giant tiger prawn, kuruma prawn, Japanese flounder and Atlantic salmon, although their density is generally low. Density is high for the rainbow trout, where the map published in 2003 has over 1 300 markers spread throughout the genome – the vast majority are AFLPs but it also includes over 200 microsatellite markers (Nichols *et al.*, 2003). Some QTLs of interest have been detected (e.g. for cold and salinity tolerance in tilapia and for specific diseases in rainbow trout and salmon). In a recent review of MAS in fish breeding schemes, Sonesson (2003) suggested that MAS would be especially valuable for traits that are impossible to record on the candidates for selection such as disease resistance, fillet quality, feed efficiency and sexual maturation, and concluded that MAS is not used in fish breeding schemes today and that the lack of dense molecular maps is the limiting factor.

Conclusions

Molecular marker maps, the necessary framework for any MAS programme, have been constructed for the majority of agriculturally important species but the density of the maps varies considerably among species. Currently, MAS does not play a major role in genetic improvement programmes in any of the agricultural sectors. Enthusiasm and optimism remain concerning the potential contributions that MAS offers for genetic improvement. However, this seems to be tempered by the realization that it may be more difficult and therefore take longer than originally thought before genetic improvement of quantitative traits using MAS is realized. The conclusions from the review by Dekkers and Hospital (2002) are a good reflection of this: "Further advances in molecular technology and genome programmes will soon create a wealth of information that can be exploited for the genetic improvement of plants and animals. High-throughput genotyping, for example, will allow direct selection on marker information based on population-wide LD. Methods to effectively analyse and use this information in selection are still to be developed. The eventual application of these technologies in practical breeding programmes will be on the basis of economic grounds, which, along with cost-effective technology, will require further evidence of predictable and sustainable genetic advances using MAS. Until complex traits can be fully dissected, the application of MAS will be limited to genes of moderate-to-large effect and to applications that do not endanger the response to conventional selection. Until then, observable phenotype will remain an important component of genetic improvement programmes, because it takes account of the collective effect of all genes."

SOME FACTORS RELEVANT TO APPLYING MAS IN DEVELOPING COUNTRIES

In the debate on the role or value of MAS as a potential tool for genetic improvement in developing countries, some of the potential factors that should be considered are described briefly below, as they may influence applications of the technology.

Economic factors

As with any new technology promising increased benefits, the costs of application must also be considered. According to Dekkers and Hospital (2002), "economics is the key determinant for the application of molecular genetics in genetic improvement programmes. The use of markers in selection incurs the costs that are inherent to molecular techniques. Apart from the cost of QTL detection, which can be substantial, costs for MAS include the costs of DNA collection, genotyping and analysis." For example, Koebner (2003) suggested that the current costs of MAS would need to fall considerably before it would be used widely in wheat and barley breeding. In practice, therefore, although MAS may lead to increased genetic responses, decision-makers need to consider whether it may be cost-effective or whether the money and resources spent on developing and applying MAS might instead be more efficiently used on improving existing conventional breeding programmes or adopting other new technologies.

Little consideration has been given to this issue. Some results have, however, been published recently from studies at the International Maize and Wheat Improvement Center (CIMMYT) in Mexico on the relative cost-effectiveness of conventional selection and MAS for different maize breeding applications. One application considered by Morris *et al.* (2003) was the transfer of an elite allele at a single dominant gene from a donor line to a recipient line. Here, conventional breeding is less expensive but MAS is quicker. For situations like this, where the choice between conventional breeding and MAS involves a trade-off between time and money, they suggested that the cost-effectiveness of using MAS depends on four parameters: the relative cost of phenotypic versus marker screening; the time saved by MAS; the size and temporal distribution of benefits associated with accelerated release of improved germplasm and, finally, the availability to the breeding programme of operating capital. They conclude that "all four of these parameters can vary significantly between breeding projects, suggesting that detailed economic analysis may be needed to predict in advance which selection technology will be optimal for a given breeding project."

In the applications considered by CIMMYT, the costs of developing molecular markers associated with the trait of interest were not considered, as it was assumed that they were already available. There is a distinction between development costs (e.g. identifying molecular markers on the genome, detecting associations between markers and the traits of interest) and running costs (typing individuals for the appropriate markers in the selection programme) of MAS. Development costs can be considerable, so developing countries need to consider whether to develop their own technology or, alternatively, to import the technology developed elsewhere, if available.

Another aspect to be considered is how to evaluate the economic benefits of MAS. For a publicly-funded breeding programme, it should include economic benefits to farmers from genetic improvement of their plants or animals. For private companies on the other hand, the impacts of using MAS on their market share, and not on rates of genetic improvement, would be of greatest interest.

The economics of MAS are considered in more detail later, in particular in Chapter 19.

MAS versus conventional methods

Although conventional breeding programmes that rely on phenotypic records

have their limitations, they have shown over time that they can be highly successful. Application of MAS will not occur in a vacuum and the potential benefits (genetic, economic, etc.) of using MAS need to be compared with those achieved or expected from any existing conventional breeding programmes.

In the different agricultural sectors, this question has received much attention from researchers. There seems to be general consensus that the relative success of MAS compared with conventional breeding may depend on the kind of trait (or traits) to be genetically improved. If the trait is difficult to record or is not routinely recorded in conventional programmes, MAS will offer more advantages than if it is routinely recorded. Similarly, if the trait is sex-limited or can only be measured late in life then MAS is favoured, as marker information can be used in both sexes and at any age.

In considering the merits of MAS versus conventional breeding, it is also important to keep in mind that the existence of a strong breeding programme is a prerequisite for the application of advanced molecular technologies such as MAS. In situations where the infrastructure and capacity are insufficient to support a successful conventional breeding programme, MAS will not provide a shortcut to genetic improvement.

MAS versus other biotechnologies for genetic improvement

The relative costs and benefits of applying MAS should be compared not only with conventional breeding but also with the use of other new technologies that can potentially improve agricultural populations genetically. These include tissue culture in crops and forest trees, reproductive technologies (e.g. embryo transfer or cloning) in livestock and triploidization or sex-reversal in farmed fish. They also include genetic modification, a technology that can be applied to all sectors. Compared with genetic modification, regulation of MAS, be it at the level of research and development, field testing, commercial release or import/export of developed products, is more relaxed; in addition, public acceptance of the technology is not an issue.

Intellectual property rights issues

As discussed in Conference 6 of the FAO Biotechnology Forum (FAO, 2001), the issue of intellectual property rights (IPRs) is playing an ever greater role in food and agriculture in developing countries. Participants in that conference, *inter alia*, suggested that this issue was having a generally negative influence on the quality of agricultural research carried out and on the nature of research collaborations between the public and private sector and between developing and developed countries.

It is therefore obvious that IPRs may also have an impact on the development and application of MAS in developing countries. For example, the AFLP molecular marker mapping technique is patented. Molecular markers can be patented, although this can often be overcome by using other markers near the gene of interest. Individual genes can also be patented. With IPRs, however, there is nevertheless public disclosure of the invention or information. Non-disclosure of information, where patents are not sought but the information on markers or detected QTL is nevertheless kept secret, can also have negative impacts, by denying developing countries access to potentially useful information.

More details on IPRs and MAS can be found in Chapter 20.

REFERENCES

Babu, R.C., Nguyen, B.D., Chamarerk, V., Shanmugasundaram, P., Chezhian, P., Jeyaprakash, P., Ganesh, S.K., Palchamy, A., Sadasivam, S., Sarkarung, S., Wade, L.J. & Nguyen, H.T. 2003. Genetic analysis of drought resistance in rice by molecular markers: association between secondary traits and field performance. *Crop Sci.* 43: 1457–1469.

Butcher, P.A. 2003. Molecular breeding of tropical trees. *In* A. Rimbawanto & M. Susanta, eds. *Proc. Int. Conf. Adv. in Genetic Improvement of Tropical Tree Species*, 1–3 October 2002. Centre for Forest Biotechnology and Tree Improvement, Yogyakarta, Indonesia.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82: E313–E328 (available at http://jas.fass.org/cgi/reprint/82/13_suppl/E313)

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Revs. Genet.* 3: 22–32.

FAO. 2001. *Agricultural biotechnology for developing countries – results of an electronic forum,* by J. Ruane & M. Zimmermann. FAO Research and Technology Paper No. 8. Rome (also available at www.fao.org/docrep/004/Y2729E/Y2729E00.htm).

Koebner, R. 2003. MAS in cereals: green for maize, amber for rice, still red for wheat and barley. In *Proc. Int. Workshop on Marker-Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding?* (available at www.fao.org/biotech/docs/Koebner.pdf).

Morris, M., Dreher, K., Ribaut, J-M. & Khairallah, M. 2003. Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol. Breed.*11: 235–247.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, B., Bentzen, P., Spies, I., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A. & Thorgaard, G.H. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet.* 34: 102–115.

Paterson A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. & Tanksley, S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726.

Sonesson, A.K. 2003. Possibilities for marker-assisted selection in fish breeding schemes. In *Proc. Int. Workshop on Marker-Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding?* (available at www.fao.org/biotech/docs/Sonesson.pdf).

Tartarini, S. 2003. Marker-assisted selection in pome fruit breeding. In *Proc. Int. Workshop on Marker-Assisted Selection: A Fast Track to Increase Genetic Gain in Plant and Animal Breeding?* (available at www.fao.org/biotech/docs/Tartarini.pdf).

Yanchuk, A. 2002. The role and implications of biotechnology in forestry. *Forest Genetic Resources* 30: 18–22 (available at www.fao.org/docrep/005/Y4341E/Y4341E06.htm).

Young, N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Mol. Breed.* 5: 505–510.

CHAPTER 2

An assessment of the use of molecular markers in developing countries

Andrea Sonnino, Marcelo J. Carena, Elcio P. Guimarães, Roswitha Baumung, Dafydd Pilling and Barbara Rischkowsky






SUMMARY

Four different sources of information were analysed to assess the current uses of molecular markers in crops, forest trees and livestock in developing countries: the FAO Biotechnology in Developing Countries (FAO-BioDeC) database of biotechnology in developing countries; country reports evaluating the current status of applied plant breeding and related biotechnologies; country reports on animal genetic resources management for preparing the First Report on the State of the World's Animal Genetic Resources (SoW-AnGR); and the results of a questionnaire survey on animal genetic diversity studies. Even if still largely incomplete, the current data show that molecular markers are widely used for plant breeding in the developing world and most probably their use will increase in the future. In the animal sector the use of molecular markers seems less developed and limited or absent in most developing countries. Major differences exist among and within regions regarding the application of molecular marker techniques in plant and animal breeding and genetics. These can be explained by the relatively high investments in infrastructure and human resources necessary to undertake research in these fields. The spectrum of application of molecular markers in crop plants is quite wide, covering many plants relevant to the enhancement of food security, but other important plant species are still neglected. The practical results of marker-assisted selection (MAS) in the field are disappointingly modest, possibly due to: low levels of investment; limited coordination between biotechnologists and practical breeders; instable, non-focused or ill-addressed research projects; and the lack of linkages between research and farmers. Partnerships between developed and developing countries may be a means of better realizing the potential of molecular marker techniques for improving both animal and crop production.

INTRODUCTION

Assessments relating to the use of molecular markers in crop plants are based on two sources of information: (i) FAO-BioDeC, a searchable database of biotechnology products and techniques in use and in the pipeline in developing and transition countries (available at www.fao.org/biotech/inventory_admin/dep/default.asp); and (ii) FAO country reports produced by national agricultural research systems (NARS) as part of a survey of country information and trends in resources allocated for applied plant breeding and related biotechnology, with the aim of raising awareness, evaluating opportunities for investment and designing national, regional and/or global strategies to strengthen the capacity of national plant breeding programmes (Guimarães, Kueneman and Carena, 2006).

As the FAO-BioDeC database contains little information on the use of molecular markers in relation to animals, it is even more difficult to give a comprehensive overview of the situation with respect to livestock in developing countries than it is for crops. However, information on the use of molecular markers was drawn from the country reports on animal genetic resources (AnGR) management submitted to FAO as part of the preparation of the First Report on the State of the World's Animal Genetic Resources (SoW-AnGR) and from a questionnaire survey on genetic diversity studies. The country reports covered a wide variety of aspects of AnGR management and contain only quite general information about the role of molecular techniques. The questionnaire survey looked specifically at the use of molecular markers in livestock genetic diversity studies and was directed to researchers involved in such studies. As such, it gives an indication of where genetic diversity studies are being undertaken and which markers are primarily used, but it does not provide a complete picture.

While this book focuses on the use of markers to assist in genetic selection (MAS), it is often difficult to obtain specific information on the extent to which markers are used for this purpose in developing countries. For this reason, some of the data presented in this chapter cover the overall use of molecular markers in developing countries and do not allow discrimination between molecular markers used for selection from uses for other purposes, such as the descriptive studies of genetic diversity within populations or genetic distance between populations. Other data presented here describe the use of molecular markers for measuring genetic diversity only. In this case, the information can be considered as an indicator of the human capacity and infrastructure available for use of markers in MAS. For these reasons, and due to the incomplete nature of some of the information available, this overview should be considered preliminary, but still meaningful.

FAO-BioDeC

At the time of writing (September 2006), FAO-BioDeC includes 2 336 entries related to crops and 829 entries related to forest trees. The database currently covers 74 developing countries, including countries with economies in transition.

No quantitative information is available concerning the human capacity or funding involved in any research initiative. Activities carried out in developed countries or at international research centres, such as those that are part of the Consultative Group on International Agricultural Research (CGIAR), are not considered.

To compile the data in FAO-BioDeC, several sources of information were

TABLE 1
Number of research initiatives utilizing genetic markers in the crop and forestry sectors sorted by type of markers

Markers	Crop	Forestry	Total
RFLP	61	9	70
RAPD	158	15	173
SSRs/Microsatellites	68	19	87
AFLP	65	3	68
Isozymes	2	50	52
Chloroplast DNA markers	0	11	11
rDNA (ribosomal DNA sequences)	0	4	4
Other or not specified	135	77	212
Total	**489**	**188**	**677**

TABLE 2
Number of research initiatives utilizing genetic markers in the crop and forestry sectors according to the development stage of the technique or product

Phase	Crop	Forestry	Total
Experimental phase	344	179	523
Field tests	107	8	115
Commercial phase	4	1	5
Unspecified	34	0	34
Total	**489**	**188**	**677**

consulted (for a complete description see FAO, 2005). In particular, information on plant biotechnology products and techniques was gathered from a survey undertaken in Latin America by the International Service for National Agricultural Research (ISNAR) and from country biotechnology status assessment reports prepared for FAO in South and Southeast Asia, Africa and transition countries in Eastern Europe. Other information was obtained from country reports and published literature.

The initial biotechnology application data obtained was classified on a country/regional/continental basis, by species, trait analysed or technique used, and by whether the application was in the research or field testing phases or was already commercially released.

FAO-BioDeC currently contains 677 entries related to the use of molecular marker techniques, 489 of which are associated with crop plants and 188 with forest trees. Table 1 suggests that early generation DNA-based molecular markers such as randomly amplified polymorphic DNAs (RAPDs) are more widely used than the more recently developed markers, e.g. amplified fragment length polymorphisms (AFLPs), while isozymes are still largely used in the forestry sector.

Only in five cases have the research initiatives reported reached the final stage of development, giving rise to commercialized products (Table 2). These are one variety of an unspecified ornamental plant released in Brazil; one variety of rice commercialized in Indonesia; one strain of *Rhizobium etli*, the soil bacterium inducing the formation of nitrogen-fixing nodules on the roots of a common bean obtained in Mexico; one rice variety containing pyramided genes for bacterial leaf blight resistance obtained in the Netherlands Antilles; and one variety of an unspecified forest tree in Burundi. In 115 cases (107 in the crop sector and eight

TABLE 3
Number of research initiatives utilizing genetic markers in the crop and forestry sectors according to geographical origin

Region	Crop	Forestry	Total
Africa	52	17	69
Asia and Pacific	98	103	201
Europe (transition countries)	42	13	55
Latin America and Caribbean	249	55	304
Near East and North Africa	48	0	48
Total	**489**	**188**	**677**

for forest trees), the research initiatives have reached the field test stage, while in 523 cases (344 of which are related to the crop sector), they are at earlier stages.

The use of molecular markers is widespread in Latin America and the Caribbean with molecular research being reported from ten countries. Special emphasis is on the crop sector and includes applications on Andean local roots and tubers, sugar cane, rice, cocoa, banana, bean and maize (Table 3). In the Asia and Pacific region, research activities with molecular markers focus on forest trees, sugar cane, rice, jute, banana, coconut and wheat. The FAO-BioDeC database shows that, while research involving molecular markers in Africa is under way in only a few countries including Ethiopia, Nigeria, South Africa and Zimbabwe, the crops under study range from traditional commodities to tropical fruits. Molecular research in the Near East and North Africa is reported for only six countries and focuses on date palm, durum and bread wheat, rice, barley and olive trees. In transition countries of Eastern Europe, molecular markers target several crop plants including wheat, maize, pulses, vegetables and tobacco across seven countries.

Table 4 shows that most attention focuses on cereals, especially durum and bread wheat, barley, maize and rice. Other important cereal or pseudo-cereal species such as sorghum, amaranthus and buckwheat receive less attention and no research initiatives are reported for teff or millets. Among the pulses, molecular research projects are reported for beans (18), chickpea (5), cowpea (9) and soybean (7) and little or no attention is dedicated to lentil, pigeon pea, faba bean and other locally important leguminous plants such as bambara groundnut. Among root and tuber crops, potato, sweet potato and cassava attract the most research effort involving molecular markers, but some research is also undertaken on Andean roots and tubers. Few or no records are available for root and tuber species important for food security in many developing countries such as yam, taro (or dasheen), cocoyam and other aroids. Research on fruit trees involving molecular markers includes tropical fruit trees such as banana, cocoa, coconut and papaya, as well as plants more typical of temperate climates such as strawberry and apple, while less

TABLE 4
Number of research initiatives utilizing genetic markers according to the crop of application

Crop group	Number of projects
Cereals and pseudo-cereals	134
Pulses	54
Root and tubers	51
Fruit trees	53
Vegetables	29
Industrial crops	74
Fodder crops	16
Aromatics	5
Other or not specified	73
Total	**489**

research was reported for citrus, mango, pineapple and many other fruit trees largely cultivated in developing countries. Several research initiatives are applying molecular markers to industrial crop species, such as sugar cane, cotton, rubber, jute, coffee, flax and oil palm.

FAO PLANT BREEDING AND RELATED BIOTECHNOLOGY CAPACITY ASSESSMENT

In 2002, a draft questionnaire was designed to gather country information on resource allocation trends in plant breeding and biotechnology related activities. Later in the same year, a group of experts including representatives from CGIAR centres, the public and private sectors and non-governmental organizations (NGOs), met at FAO headquarters to discuss the nature of the information to be collected and the procedure for its collection. This resulted in a questionnaire being developed and sent to all public and private applied plant breeding programmes as well as to biotechnology laboratories in developing countries and countries in transition. Among other issues, the survey gathered information on the number of full-time equivalent plant breeders and biotechnologists available during each five-year period beginning from 1985. The questionnaire also requested information concerning trends of resources allocated to biotechnology as well as to germplasm improvement (pre-breeding), line development and line evaluation. One of the objectives of the survey was to assess the gap between biotechnology tools and their successful deployment in applied breeding programmes (Guimarães, Kueneman and Carena, 2006). The survey therefore also concentrated on priorities for breeding, potential international support to strengthen national breeding programmes, the number of varieties released and the factors that are most likely to limit the success of applied plant breeding programmes, including the current status of biotechnology. The work of gathering the information and preparing a technical report on the current status of national plant breeding and related biotechnology was assigned to a well-known and respected national plant breeding scientist. This has been the key to identifying gaps in order to develop strategies for strengthening efforts directed at the sustainable use of plant genetic resources for food and agriculture (PGRFA) in national programmes.

For the purposes of this chapter, biotechnology data were gathered from 25 countries to complement the preliminary assessments based on FAO-BioDeC on the use of molecular markers in developing countries (Table 5). The data gathered indicate that tissue culture is the most common biotechnology technique as it was used in 88 percent of all cases, followed by MAS (44 percent), the double-haploid technique (32 percent), interspecific crosses (28 percent), molecular characterization (24 percent) and genetic engineering (12 percent).

Applications of molecular markers include a number of categories within biotechnology such as MAS, molecular characterization, facilitating genetic engineering and tracking desirable chromosome segments when making wide crosses (e.g. interspecific crosses). The results in Table 5 suggest that molecular markers might be an integral part of developing country agricultural efforts. MAS seems to be the second most utilized biotechnology tool applied after tissue culture, implying that emphasis should be given to the development of molecular markers to make selection more efficient. However, rapid and efficient

TABLE 5
Biotechnology applications in plant genetic resources for food and agriculture in use in 25 developing countries

Country	TC	MAS	IC	DH	MC	GE
Algeria	X[1]	X	X	X	N[2]	N
Angola	X	N	N	N	N	N
Armenia	X	N	X	X	X	N
Cameroon	X	X	N	N	N	N
Costa Rica	X	N	X	N	X	X
Dominican Republic	N	N	N	N	N	N
Ethiopia	X	X	N	X	N	N
Georgia	X	X	X	N	X	N
Ghana	X	X	N	N	N	N
Mali	X	N	N	N	N	N
Kenya	X	X	X	X	N	X
Malawi	X	N	N	N	N	N
Moldova	X	N	N	N	N	N
Mozambique	N	N	N	N	N	N
Nicaragua	X	X	N	N	X	N
Niger	X	X	N	N	N	N
Nigeria	X	X	X	X	X	N
Senegal	X	N	X	X	N	N
Sierra Leone	X	N	N	N	N	N
Sri Lanka	N	N	N	N	N	N
Sudan	X	N	N	N	N	N
Tunisia	X	X	N	X	N	N
Uzbekistan	X	N	N	X	X	N
Zambia	X	N	N	N	N	N
Zimbabwe	X	X	N	N	N	X

[1] One or more institutions in the country are using the tool. However, this does not measure its impact.
[2] Not in use.
TC = tissue culture; MAS = marker-assisted selection; IC = interspecific crosses; DH = double-haploid technology; MC = molecular characterization; GE = genetic engineering

advancement of plant breeding efforts might not be achieved through MAS because of the complexity encountered in multitrait and multistage selection for economically important traits. Consequently, today in the developed world, molecular markers do not have a prominent role in breeding programmes (Hallauer, 1999).

USE OF MOLECULAR TECHNIQUES IN AnGR MANAGEMENT

FAO invited 188 countries to participate in the preparation of the First Report on the SoW-AnGR. One hundred and sixty-nine country reports (CR) on AnGR were submitted (available at www.fao.org/dad-is/).

The countries were offered guidelines for the preparation of the country reports, one section of which was to be devoted to reviewing the state of national capacities and assessing future capacity building requirements (FAO, 2001). Countries were assigned to seven regions on the basis of the regional classification established by FAO for the purpose of preparing the SoW-AnGR. This analysis considered 148 country reports available by July 2005, of which 42 were from Africa, 25 from Asia, 39 from Europe and the Caucasus, 22 from Latin America and the Caribbean, 7 from the Near and Middle East, 2 from North America and 11 from the Southwest Pacific (Pilling *et al.*, 2007).

TABLE 6
Use of molecular markers reported in country reports on AnGR management

Region	Number providing information	Reporting use of molecular markers %	Number with information on species	Reporting use of molecular markers	
				In cattle %	In other species %
Europe	29	83	18	89	100
Africa	29	14	3	100	33
Asia	16	50	7	86	100
Latin America and the Caribbean	15	73	9	78	89
Southwest Pacific	9	11	0	-	-
North America	2	100	1	100	100
Near and Middle East	5	40	2	0	100

Not surprisingly, the information provided by the country reports indicates that there is a large gap between developed and developing countries in terms of capacity to utilize molecular markers for the study and management of AnGR (Table 6). Compared with other developing regions, a higher percentage of countries from Asia and Latin America and the Caribbean reported their use. In Africa, the Southwest Pacific (excluding Australia), the Near and Middle East, and Eastern Europe and the Caucasus, very few countries report the use of these technologies, the prominent exception in the last case being Ukraine which has carried out molecular characterization and genetic distance studies on a number of livestock species (CR Ukraine, 2004).

In Africa, only four countries describe the existence of characterization or genetic distance studies based on the use of molecular markers and in all cases the studies relate to local breeds. One country report indicates that local breeds of goat, pig and chicken are the subject of molecular characterization carried out abroad. In no case is the use of MAS reported from this region.

Excluding Japan, seven Asian countries (out of 15 providing information on whether or not the technologies are used) report molecular marker studies, of which five specify genetic distance studies and one mentions research into MAS (CR Malaysia, 2003). A range of species are the subject of molecular characterization, the most common being cattle, chickens, sheep, goats and pigs; however, some studies involving buffaloes, ducks, horses, camels or deer are also reported. Systematic studies of Asian breeds are being conducted by the Society for Research on Native Livestock in Japan, including analysis of genetic relationships based on mitochondrial DNA polymorphisms and other DNA markers (CR Japan, 2003).

In Latin America and the Caribbean, 11 countries out of the 15 that provided information indicate some use of molecular markers. Among nine countries providing information on the species involved in molecular characterization studies, seven mentioned cattle while smaller numbers mention sheep, pigs, chickens, horses, goats, buffaloes, llamas, alpacas, vicuñas or guanacos. Several countries indicate the inclusion of locally adapted breeds in such studies, but there was little indication that molecular markers have been incorporated within breeding programmes. However, the report from Colombia (2003) noted the potential significance of MAS programmes for utilizing the genes of the Blanco Orejinegro cattle breed, which is reported to show resistance to brucellosis and which has been the subject of molecular characterization.

Apart from Australia, no countries in the Southwest Pacific region report the use of molecular markers.

In the Near and Middle East one report (CR Jordan, 2003) refers to molecular characterization and genetic distance studies in indigenous goats, while another (CR Egypt, 2003) notes that molecular studies of buffalo, sheep and goats had recently been initiated with the aid of regional and international organizations.

SURVEY ON THE USE OF MOLECULAR MARKERS IN GENETIC DISTANCE STUDIES IN LIVESTOCK

More specific and detailed information on the use of molecular markers in AnGR research was obtained from a questionnaire study launched in 2003. One hundred and thirty-two questionnaires were sent out via e-mail to research teams that had been involved in genetic distance studies during the past ten years. The researchers were identified through a literature search and enquiry via several Internet discussion groups. The points covered in the survey were: number of breeds and sample sizes; number and type of markers used; additional breed information such as phenotypic traits or geographic spread; and the mathematical and statistical methods chosen for measuring genetic distance. The study also aimed to verify the degree of familiarity and acceptance of measurement of domestic animal diversity (MoDAD) recommendations, which had been proposed as standards for genetic diversity studies by the International Society for Animal Genetics (ISAG) and FAO about ten years earlier (FAO, 1998a; b). Compliance with the recommendations was seen as important as it would enable the compilation of results from different genetic distance studies.

TABLE 7
Number of countries where samples were collected for AnGR genetic distance studies

FAO region	Number of countries
Africa	13
Asia and the Pacific	19
Europe	37
Latin America and the Caribbean	10
Near East	9
North America	2
Total	**93**

Information on 87 genetic distance studies was obtained from 57 researchers. The studies covered breeds from 13 mammalian and avian species and investigated samples from 93 countries; the largest number of countries was in Europe, followed by those in Asia and the Pacific (Table 7). Most of the studies focused on ruminants. The size of the projects varied between one and 120 breeds originating from up to 33 countries. However, a large number of national projects focused on breeds within a specific country or region. There were also a few large international projects involving cattle and goats (Table 8). A smaller number of pig and chicken projects were implemented. No feedback was received regarding breeds of llamas, ducks, turkeys or geese.

With regard to compliance with the recommendations of the FAO/ISAG advisory group, 95 percent of all projects aimed to fulfil the minimum requirement of sampling 25 animals per breed. Although microsatellite markers were used in 90 percent of the studies, in only 23 percent were all markers taken from the recommended marker list. In about 57 percent of studies some recommended microsatellites were used. The degree of acceptance of the recommendations was highest in pigs and lowest in chickens. More detailed information on the results is given by Baumung, Simianer and Hoffmann (2004) and FAO (2004).

TABLE 8
Number of projects and countries in which samples were collected according to animal species and FAO regions

Species	Number of projects	Number of countries	FAO region
Buffalo	3	9	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Cattle	24	40	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Goat	11	28	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Sheep	19	56	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean
Pig	6	19	Africa, Asia and the Pacific, Europe
Ass	1	1	Europe
Horse	5	25	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean, North America
Bactrian camel	1	2	Asia and the Pacific
Dromedary	2	7	Africa, Near East
Alpaca	3	2	Near East, Latin America and the Caribbean
Rabbit	1	19	Africa, Asia and the Pacific, Europe
Chicken	8	34	Africa, Asia and the Pacific, Europe, Latin America and the Caribbean, Near East
Yak	2	8	Asia and the Pacific, Europe, Near East

CONCLUSIONS

Even if still largely incomplete, the current data allow some general conclusions to be drawn regarding the use of molecular markers in agricultural research and development in developing countries.

Molecular markers are widely utilized in the plant production sector of the developing world even if the present uptake of molecular marker technologies does not reflect their actual potential. It might therefore be speculated that a significant increase in their utilization might be expected in the near future. However, it is recommended that each technique is carefully assessed for its actual potential for improving the efficiency of plant breeding and germplasm characterization. Until this is demonstrated, the use of molecular markers would be a costly investment with limited returns. Publishing all marker research that has not been successful is also strongly encouraged in order to avoid potential failures and/or importing inappropriate technologies from developed countries.

Major differences exist between regions (and within regions) regarding the application of molecular marker techniques in plant breeding and genetics. While some countries have developed quite extensive research programmes, vast geographical areas, particularly in Africa, remain excluded from these technological advancements or can count only on minimal activities. This can be explained by the relatively high investments in infrastructure and human resources necessary to undertake research in this field. High costs can also be indicated as a cause of the low technological level of genetic marker research in many countries, which focus on isozymes or on restriction fragment length polymorphisms (RFLPs) and have not yet adopted the more advanced polymerase chain reaction (PCR)-based markers. However, the life span of PCR-based markers is very short and it might be better to wait until improved markers such as single nucleotide polymorphisms (SNPs) become available. The spectrum of application of molecular markers to crop plants in developing countries is quite wide and covers many plant species that are relevant for the enhancement of food security or for the improvement of farmers' incomes

in tropical areas. However, other important plant species are still neglected by the ongoing research initiatives.

According to the data reported in FAO-BioDeC, only five products obtained through the use of molecular markers have been commercially released to date in developing countries. Even if more commercial products have been released but are missing from the database, such as those reported by Toenniessen, O'Toole and DeVries (2003) or others obtained by the international agricultural research centres or the private sector, the totality of practical results obtained from using molecular markers is disappointingly modest compared with the declared potential of the approach. The reasons for the poor results to date are multiple and include: the low level of investments in both biotechnology research and applied plant breeding; the limited coordination between biotechnology laboratories and plant breeding programmes; managerial and political frailties leading to instable, unfocused or ill-addressed research projects; legal, infrastructural or technical weaknesses of the seed production and commercialization systems; and the lack of linkages between research and practical application of research products by farmers.

Applied plant breeding should continue to be the foundation for the application of molecular markers. Focusing useful molecular techniques on the right traits will build a strong linkage between genomics and plant breeding in order to produce new and better cultivars. Therefore, more than ever, there is the need for better communication and cooperation among scientists in plant breeding and biotechnology. Public plant breeding and biotechnology programmes in developing countries are being seriously eroded through lack of funding. This loss of public support affects breeding continuity and objectivity and, equally importantly, the training of future plant breeders and biotechnologists and the utilization and improvement of plant genetic resources currently available. The fact that poor farmers rely on public and private breeding institutions for solving long-term challenges should influence policy-makers to reverse the trend of reduced funding. Cooperation between industry and public institutions is a promising approach to follow. Ensuring strong applied breeding programmes incorporating the application of molecular markers will be essential in ensuring the sustainable use and enhancement of plant genetic resources.

AnGR management shows a similar pattern to the use of MAS in plant breeding management in terms of the differences that exist among regions in the use of molecular marker techniques. Within several regions there are also differences between more and less developed countries. The reasons are similar to those mentioned above, namely a lack of financial, human and technical resources. In particular, human capacities in animal genetics and breeding are much smaller than those existing in the crop sector. Consequently, the use of molecular techniques to evaluate genetic resources, to plan conservation efforts, or to facilitate the achievement of desired breeding objectives is limited or absent in most developing countries.

Nevertheless, country reports expressed a strong desire to develop greater capacity to carry out molecular studies of national AnGR, and the responses to the FAO questionnaire also indicated a high level of interest in doing so. For the near future, microsatellite loci will remain the most useful type of genetic marker for genetic distance studies and for genetic improvement

programmes but SNPs were singled out as promising markers for the future. With partnerships between developed and developing countries within or across regions, genetic diversity studies may be a means of realizing the potential of molecular marker techniques to improve decision-making on breed development and the prioritization of breeds for conservation programmes.

The successful application of MAS in animal breeding necessitates a high level of expenditure in terms of establishment and maintenance costs and requires skilled human resources, equipment, laboratories and supportive infrastructure. As such, the cost-effectiveness of these strategies has to be carefully evaluated before promoting them in resource-poor environments.

REFERENCES

Baumung, R., Simianer, H. & Hoffmann, I. 2004. Genetic diversity studies in farm animals – a survey. *J. Anim. Breed. Genet*. 121: 361-373.

FAO. Country reports on the state of animal genetic resources (available at www.fao.org/dad-is/).

FAO. 1998a. *Secondary guidelines for development of National Farm Animal Genetic Resources Management Plans. Measurement of domestic animal diversity (MoDAD): original Working Group report*. Rome.

FAO. 1998b. *Secondary guidelines for development of National Farm Animal Genetic Resources Management Plans. Measurement of domestic animal diversity (MoDAD): recommended microsatellite markers*. Rome.

FAO. 2001. *Preparation of the first report on the state of the world's animal genetic resources*. Guidelines for the preparation of CRs. Rome.

FAO. 2004. *Measurement of domestic animal diversity – a review of recent diversity studies.* Document prepared for the third session of the Intergovernmental Working Group on Animal Genetic Resources. Rome (available at www.fao.org/ag/againfo/programmes/en/genetics/documents/ITWG3_Inf3.pdf).

FAO. 2005. *Status of research and application of crop biotechnologies in developing countries: preliminary assessment,* by Z. Dhlamini, C. Spillane, J.P. Moss, J. Ruane, N. Urquia & A. Sonnino. Rome (available at www.fao.org/docrep/008/y5800e/y5800e00.htm).

Guimarães, E.P., Kueneman, E. & Carena, M.J. 2006. Assessment of national plant breeding and biotechnology capacity in Africa and recommendations for future capacity building. *Hort. Sci.* 41: 50–52.

Hallauer, A.R. 1999. Temperate maize and heterosis. *In* J.G. Coors & S. Pandey, eds. *The genetics and exploitation of heterosis in crops,* pp. 353-361. Proc. Internat. Symp. of Heterosis in Crops. Mexico City, 18–22 August 1997. ASA, CSSA and SSSA, Madison, WI, USA.

Pilling, D., Cardellino, R., Zjalic, M., Rischkowsky, B., Tempelman, K.A. & Hoffmann, I. 2007. The use of reproductive and molecular biotechnology in animal genetic resources management – a global overview. *Anim. Genet. Resources Information*. FAO, Rome (in press).

Toenniessen G.P., O'Toole, J.C. & DeVries, J. 2003. Advances in plant biotechnology and its adoption in developing countries. *Curr. Opin. in Plant Biol*. 6: 191–198.

SECTION II

Marker-assisted selection in crops – case studies

CHAPTER 3

Molecular markers for use in plant molecular breeding and germplasm evaluation

Jeremy D. Edwards and Susan R. McCouch

SUMMARY

A number of molecular marker technologies exist, each with different advantages and disadvantages. When available, genome sequence allows for the development of greater numbers and higher quality molecular markers. When genome sequence is limited in the organism of interest, related species may serve as sources of molecular markers. Some molecular marker technologies combine the discovery and assay of DNA sequence variations, and therefore can be used in species without the need for prior sequence information and up-front investment in marker development. As a prerequisite for marker-assisted selection (MAS), there must be a known association between genetic markers and genes affecting the phenotype to be modified. Comparative databases can facilitate the transfer of knowledge of genetic marker-phenotype association across species so that discoveries in one species may be applied to many others. Further genomics research and reductions in the costs associated with molecular markers will continue to provide new opportunities to employ MAS.

INTRODUCTION

Molecular markers are valuable tools for the classification of germplasm and in MAS. The purpose of this chapter is to provide guidance in selecting appropriate molecular marker systems based on the availability of technological resources in various species and to provide some examples of MAS applications. One of the many benefits of the increasing amount of DNA sequence information in many organisms is the expanding opportunity for the development of new molecular markers. As the full genome sequence will not be available for most species of interest in the near future, it is important to find strategies for developing and using molecular markers when sequence resources are limited. This chapter describes several technologies that exist for developing molecular markers without DNA sequence information. It also draws on some examples from rice (*Oryza sativa* L.) to illustrate how molecular marker development was influenced by the addition of each layer of sequence information, culminating in the present status of rice as the first crop with nearly complete genome sequence information.

MOLECULAR MARKER TECHNOLOGIES

Restriction fragment length polymorphisms

Restriction fragment length polymorphisms (RFLPs) were the first DNA-based molecular markers. An application of Southern analysis (Southern, 1975), RFLPs exploit the ability of single stranded DNA to bind (hybridize) to DNA with a complementary sequence. RFLP markers detect variation in DNA sequences at the same loci in different individuals or accessions. Technically, RFLP technology involves the hybridization of cloned DNA to restriction fragments of differing molecular weights from restriction enzyme-digested genomic DNA. The digested DNA fragments are size-separated on agarose gels by electrophoresis and transferred as denatured (single stranded) arrays of fragments to filters through capillary action. The filters are then incubated with specific labelled probes (genes or anonymous fragments of single stranded DNA), washed and exposed to x-ray film. To identify polymorphisms between individuals or accessions, the genomic DNA extracted from each individual is digested with a series of restriction enzymes to find enzymes that produce fragments (bands) that differ in molecular weight between accessions and can be distinguished by hybridization with a given probe. To ensure that probes hybridize to single fragments on a gel, the DNA used as a probe should be from a single or low copy (non-repetitive) region of the genome. Probes may represent genes (i.e. derived from complementary DNA [cDNA]) or they may represent anonymous sequences derived from genomic DNA. Genomic probes are generated by shearing or digesting DNA and cloning the fragments into a plasmid vector that allows for amplification of the cloned fragment in a suitable host. To increase the frequency of low copy clones in a genomic library, the DNA may be digested with a methylation-sensitive enzyme, such as *Pst*I. The repetitive regions of a genome are typically heavily methylated and thus produce fragments >25 kb when digested with a methylation-sensitive enzyme. As a result, these fragments do not clone efficiently into plasmid vectors and consequently are effectively filtered out of the analysis. Thus, use of methylation-sensitive enzymes increases the representation of unmethylated and typically low copy gene sequences in RFLP analysis. Sharing of anonymous,

unsequenced RFLP markers among researchers requires an infrastructure for the maintenance and distribution of cloned probes for use by multiple researchers. However, if end-sequence or full-clone sequence information is available, the probes can be amplified readily from genomic DNA via the polymerase chain reaction (PCR), and the cumbersome aspects of clone maintenance and distribution are avoided. The polymorphisms detected by RFLPs may result from single base changes causing a loss of restriction sites or a gain of new restriction sites, or from insertions and deletions (indels) between restriction sites (McCouch *et al.*, 1988; Edwards, Lee and McCouch, 2004).

PCR-based markers

Many advances in molecular marker technology have come through applications of the PCR method (Mullis *et al.*, 1986). In PCR, a thermo-stable DNA polymerase enzyme makes copies of a target sequence beginning from two small pieces of synthetically produced DNA (primers) that are complementary to sequences bracketing the target. Through iterations of the process with heating to separate the double stranded DNA molecules and cooling to allow the primers to re-anneal, the target sequence is exponentially amplified. Polymerase chain reaction-based markers require much less DNA per assay than RFLPs and are more compatible with automated high-throughput genotyping (i.e. the ability to process large numbers of samples quickly and efficiently).

Randomly amplified polymorphic DNA markers

Randomly amplified polymorphic DNA markers (RAPDs) use PCR to amplify stretches of DNA between single primers of arbitrary sequence (Williams *et al.*, 1990; Welsh and McClelland, 1990). Amplification occurs only where sequences complementary to the primers are in close enough proximity for successful PCR. The typical oligonucleotide used for RAPDs is ten bases long and will amplify many loci simultaneously, allowing multiple markers to be assayed in a single PCR reaction and a single lane on an agarose gel. As the primers are arbitrary, RAPD technology can be applied directly to any species with no prior sequence knowledge. This technology is particularly useful when there is a need to assay loci across the entire genome. The polymorphisms are detected only as the presence or absence of a band of a particular molecular weight, and it is not possible to differentiate between homozygous and heterozygous markers. RAPDs are notoriously unreliable because, aside from sequence differences, the amplification or failure of amplification of any band may be sensitive to any number of factors, including DNA template quality, PCR conditions, reagents and equipment.

Amplified fragment length polymorphisms

Amplified fragment length polymorphisms (AFLPs) are molecular markers derived from the selective amplification of restriction fragments (Vos *et al.*, 1995). Genomic DNA is digested with a pair of restriction enzymes and oligonucleotide adaptors are ligated to the ends of each restriction fragment. The fragments are amplified using primers that anneal to the adaptor sequence and extend into the restriction fragment. Only a portion of restriction fragments will be within the range of sizes than can be amplified by PCR and visualized on polyacylamide gels (between 50 and 350 bp). For large genomes, additional selective bases

can be added to the primers to reduce the number of co-amplified bands. AFLPs have many of the advantages of RAPDs, but have much better reproducibility. AFLP technology requires greater technical skill than RAPDs and, because AFLPs run on polyacrylamide gels instead of agarose, they also require a larger investment in equipment than RAPDs. Using manual gels, AFLP bands are detectable using silver stain, or by labelling of the primers with a radioactive isotope. Alternatively, for higher throughput, AFLPs can be detected with an automated DNA sequencer by using fluorescently labelled primers.

Diversity array technology (DArT) is a modification of the AFLP procedure using a microarray platform (Jaccoud *et al*., 2001) that greatly increases throughput. In DArT, DNA fragments from one sample are arrayed and used to detect polymorphisms for the fragments in other samples by differential hybridization (Wenzl *et al.*, 2004).

DEVELOPING MOLECULAR MARKERS WITH DNA SEQUENCE INFORMATION

When the DNA sequence is available, it is possible to design primers to amplify across a specific locus. However not all loci will be polymorphic. Targeting highly variable sequence features increases the likelihood of detecting polymorphism. These highly variable features include tandem repeats such as microsatellites, and dispersed complex repeats such as transposable elements.

Microsatellites

Simple sequence length polymorphisms (SSLPs), also known as simple sequence repeats (SSRs), or microsatellites, consist of tandemly repeated di-, tri- or tetra-nucleotide motifs and are a common feature of most eukaryotic genomes. The number of repeats is highly variable because slipped strand mis-pairing causes frequent gain or loss of repeat units. With their high level of allelic diversity, microsatellites are valuable as molecular markers, particularly for studies of closely related individuals.

PCR-based markers are designed to amplify fragments that contain a microsatellite using primers complementary to unique sequences surrounding the repeat motif (Weber and May, 1989). Differences in the number of tandem repeats are readily assayed by measuring the molecular weight of the resulting PCR fragments. As the differences may be as small as two base pairs, the fragments are separated by electrophoresis on polyacrylamide gels or using capillary DNA sequencers that provide sufficient resolution.

Without prior sequence knowledge, microsatellites can be discovered by screening libraries of clones. Clones containing the repeat motif must be sequenced to find unique sites for primer design flanking the repeats. Microsatellite marker development from pre-existing sequence is far more direct. Good reviews of microsatellite marker development include those of McCouch *et al*. (1997) and Zane, Bargelloni and Atarnello (2002). Microsatellites discovered in non-coding sequence often have a higher rate of polymorphism than microsatellites discovered in genes. However, in some species such as spruce (*Picea* spp.) with highly repetitive genomes, SSR markers developed from gene sequences have fewer instances of null alleles, i.e. failure of PCR amplification (Rungus *et al.*, 2004).

Microsatellite markers have several advantages. They are co-dominant; the heterozygous state can be discerned from the homozygous state. The markers are easily automated using florescent primers on an automated sequencer and it is possible

to multiplex (combine) several markers with non-overlapping size ranges on a single electrophoresis run. The results are highly reproducible, and the markers are easily shared among researchers simply by distributing primer sequences. Although SSRs are abundant in most eukaryotic genomes, their genomic distribution may vary. Uneven distributions of microsatellites limit their usefulness in some species.

Inter-SSRs (ISSRs) are another type of molecular marker that makes use of microsatellite sequences. ISSRs use PCR primers anchored in the termini of the repeats extending into the flanking sequence by several nucleotides (Zietkietkiewicz, Rafalski and Labuda, 1994). PCR products are produced for each pair of microsatellites that are in sufficient proximity for PCR to occur, or may be generated by anchoring one primer in the SSR motif and using a second "universal" primer corresponding to a sequence that has been ligated onto the ends of restriction fragments (as in the AFLP technique described above, where genomic DNA is first digested with a restriction enzyme and oligonucleotide adaptors are ligated to the ends of each restriction fragment, except that one primer resides in an SSR motif that is bracketed by the restriction sites) (Gupta *et al.*, 1994; Goodwin, Aitken and Smith, 1997). Markers at multiple loci are assayed as the presence or absence of bands of particular sizes. ISSRs can be visualized on agarose gels, on silver stained polyacrylamide gels or fluorescently labelled for detection with an automated DNA sequencer.

Transposable element-based markers

Transposable elements (TEs) are another rapidly changing feature of the genome that can be exploited as a source of variability for molecular markers. Discovery of TE sequences is a prerequisite for their use as markers. While TEs may be discovered as mutations in alleles of genes conferring mutant phenotypes, they have also been discovered directly in genomic sequence (reviewed by Feschotte, Jiang and Wessler, 2002). Transposon display is a modified AFLP procedure that differs only in that one of the two primers is designed within the consensus sequence of a TE family so that amplification depends on the presence of a TE insertion within a restriction fragment (Casa *et al.*, 2000). Using this approach, the presence or absence of a TE can be assayed simultaneously at many loci throughout the genome. To assay for a TE insertion at a specific locus, single copy "anchor markers" can be designed with primers located in unique sequences flanking the region of interest. A size polymorphism indicates the presence or absence of the TE in that particular location. Anchor markers are advantageous because they are co-dominant, can be run on a simple agarose gel system and are biologically informative in that they provide evidence of both complete, or incomplete, insertion or excision events. This methodology can also be applied to any known indel feature regardless of whether or not it is derived from a TE.

Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are an abundant source of sequence variants that can be targeted for molecular marker development. Of all the molecular marker technologies available today, SNPs provide the greatest marker density. SNPs are often the only option for finding markers very near or within a gene of interest, and can even be used to detect a known functional nucleotide polymorphism (FNP). Discovery of SNPs

TABLE 1
SNP technologies

Allele discrimination	Detection methods
• Hybridization	• Gel separation
• Primer extension	• Arrays
• Ligation	• Mass spectrometry
• Invasive cleavage	• Plate readers

requires obtaining an initial DNA sequence in a reference individual followed by some form of re-sequencing in other varieties to find variable base pairs. In addition to direct sequencing, SNPs can be discovered through ecotilling with the CEL I enzyme (Comai *et al.*, 2004) or by denaturing high pressure liquid chromatography (DHPLC) to measure small conformational differences when PCR amplified sequences are hybridized to a reference sequence (Kwok, 2001). In addition to SNP discovery, both DHPLC and ecotilling are viable technologies for SNP detection. There is a myriad of other SNP assay technologies in development and to date no single method stands out as superior to the others. Table 1 lists some examples of SNP allele discrimination methods and detection systems that can be combined in various ways (see reviews by Kwok, 2001 and Gut, 2001). The benefits of SNP assays include increased speed of genotyping, lower cost and the parallel assay of multiple SNP.

Single feature polymorphisms and microarray-based genotyping

Indel polymorphisms, also known as single feature polymorphisms (SFPs), are particularly amenable to microarray-based genotyping. These assays are done by labelling genomic DNA (target) and hybridizing to arrayed oligonucleotide probes that are complementary to indel loci. Each SFP is scored by the presence or absence of a hybridization signal with its corresponding oligonucleotide probe on the array. Both spotted oligonucleotides (Barrett *et al.*, 2004) and Affymetrix-type arrays (Borevitz *et al.*, 2003) have been used in these assays. The SFPs can be discovered through sequence alignments or by hybridization of genomic DNA with whole genome microarrays. The advantage of microarray platforms for genotyping is that they are highly parallel, and they are well suited for applications such as quantitative trait loci (QTL) analysis, where whole genome coverage with many markers is desirable.

SPECIAL CONSIDERATIONS FOR DIVERSITY STUDIES AND GERMPLASM EVALUATION

The interpretation of molecular marker data for germplasm classification and diversity can be confounded by uncertainty about the underlying sources of the polymorphisms and by homoplasy (false homology). For RFLPs in rice, indels can account for as much or more of the polymorphism as changes in the restriction sites themselves (Edwards, Lee and McCouch, 2004). AFLPs and RAPDs can also be sensitive to both indels and base changes. The ratio of indels to base changes is important for diversity studies because, when molecular markers are used to estimate nucleotide divergence, the divergence will be overestimated if indel-derived polymorphisms are common (Upholt, 1977; Nei and Miller, 1990; Innan *et al.*, 1999). The greatest certainty of the underlying polymorphism comes from SNP technologies that directly assay for single base changes.

For SSR markers among closely related individuals, most polymorphism should be caused by expansion or contraction of the number of repeat units. However, as genetic distance between the varieties increases, there is an increasing chance that indel events will cause additional size

polymorphism (Chen, Cho and McCouch, 2002). Thus, the use of stepwise SSR mutation models would be inappropriate for highly diverged populations. Homoplasy is also a problem in SSR markers because the hyper-variability leads to some shared allele sizes through parallelism, convergence and reversion (Doyle *et al.,* 1998). Homoplasy from reversions can affect transposon-based markers or any markers with polymorphisms potentially derived from Class II DNA transposable elements. This class of TEs has a cut and paste mechanism of transposition, so a TE may insert onto a locus and later excise.

In RAPDs, ISSRs and AFLPs, homoplasy can occur when two or more loci produce PCR fragments of similar molecular weight. Although it is desirable to have high numbers of bands to maximize the amount of information per lane, this must be balanced against the increasing risk of homoplasy as more loci are represented.

SPECIAL CONSIDERATIONS FOR MARKER-ASSISTED SELECTION

Quality markers for use in MAS should be reliable and easily shared among researchers. Co-dominant markers are preferred to avoid the need for progeny testing. Sometimes less desirable markers for MAS such as RAPDs, ISSRs and AFLPs are useful for finding markers linked to the desired allele. Once such a marker is found, it is possible to extract and sequence the corresponding band. This sequence can be used to develop co-dominant markers such as cleaved amplified polymorphic sequences (CAPS) (Konieczny and Ausubel, 1993) or to sequence characterized polymorphic regions (SCARs) (Paran and Michelmore, 1993). SCAR and CAPS markers are co-dominant and simplify the screening of large numbers of individuals. When a genetic map exists, markers can be positioned on the map and other linked markers can be substituted. The additional markers are useful for higher resolution mapping to find markers more closely linked to the desired allele or ultimately for positional cloning of the underlying gene.

Reproducibility of molecular marker data

For orphan species, clearly there is a huge value to the anonymous primer approaches (AFLP, DArTs, ISSRs and RAPDs) that do not require sequence information or much up-front investment. However, the data can be difficult to score, and reproducibility requires a lot of technical skill. Technologies that depend on the presence or absence of PCR amplified bands are susceptible to changes in PCR conditions and the quality of sample DNA, and the data from separate experiments may differ. Further, in any method that depends on accurate measurement of molecular weight differences between bands (e.g. SSRs), the exact mole-cular weights assigned to each allele may be different in each analysis because of differences in labelling of PCR products, rounding of allele molecular weight estimates and binning of alleles. Without controls for each allele encountered, it is difficult or impossible to merge separate sets of data. Despite discrepancies in the exact data derived from molecular markers, the results and conclusions should be consistent within independent experiments. For reliability in making inferences across independent data-sets, SNP markers are preferred. SNP data-sets can be easily integrated based on sequence, and SNPs have properties (such as a low mutation rate) that are particularly valuable for evolutionary inference (Nielsen, 2000).

TABLE 2
Key features of common molecular marker technologies

Marker type	PCR-based	Uses restriction enzymes	Poly-morphism	Abun-dance	Co-dominant	Auto-mation	Loci per assay	Specialized equipment
RFLP	no	yes	moderate	moderate	yes	no	1 to few	Radioactive isotope
RAPD	yes	no	moderate	moderate	no	yes	many	Agarose gels
AFLP	yes	no	moderate	moderate	no	yes	many	Polyacrylamide gels/capillary
ISSR	yes	no	moderate	moderate	no	yes	many	Agarose/polyacrylamide gels
DArT	yes	yes	moderate	moderate	no	yes	many	Microarray
CAPS	yes	yes	variable	moderate	yes	yes	single	Agarose gels
SCAR	yes	no	low	moderate	yes	yes	single	Agarose gels
SSR	yes	no	low	moderate	yes	yes	1 to about 20	Polyacrylamide gels/capillary
TE-Anchor	yes	no	variable	variable	yes	yes	single	Agarose gels
SNP	yes	no	variable	highest	yes	yes	1 to thousands	Variable

CHOOSING A MOLECULAR MARKER TECHNOLOGY

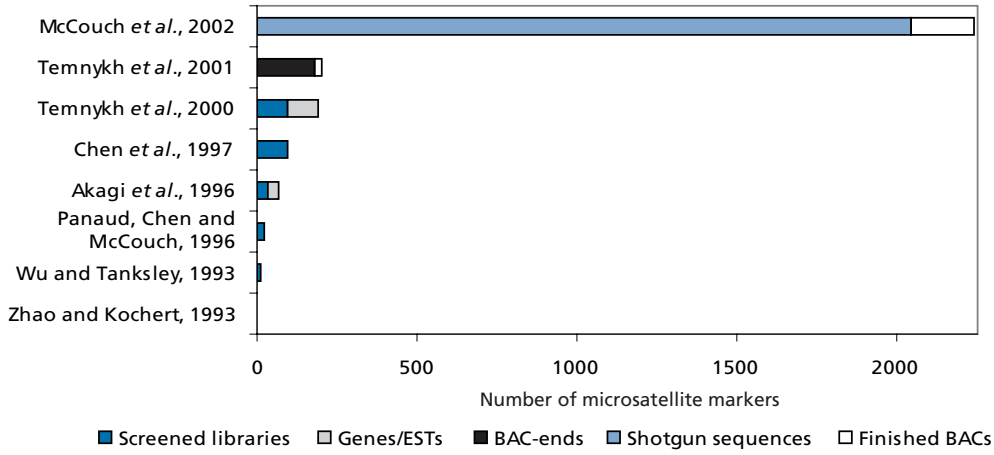
Clearly there is no single best choice of molecular marker for all situations. Factors influencing the decision may include the objectives of the study, availability of organism specific sequences, equipment and technical resources, and biological features of the species. Several important advantages/disadvantages for each type of molecular marker discussed are summarized in Table 2 (see review by Powell *et al*., 1996).

If available, microsatellite or SNP markers are often the best choice. The rate of adoption of SSR markers can be facilitated, and the costs reduced, by preparing "kits" of selected SSR markers for certain species to provide a reliable set of markers with good amplification, reasonable polymorphism and good genome coverage. This was done in the early days of the rice SSR effort and SSR kits were distributed at very low cost through Research Genetics (called Rice-Pairs; McCouch *et al.,* 1997). Similarly, for SNPs, there is a need to develop useful sets of markers that are widely available and can be mass-produced (at reduced cost) for distribution to the international community. SNP kits would also have a clear benefit for databasing and analysing datasets obtained from multiple laboratories. In addition to kits of markers, there is a need to distribute sets of "control genotypes" as samples, particularly to address the problem surrounding the difficulties in integrating SSR datasets. When SNPs or SSRs are not available, it is sometimes possible to transfer molecular markers from closely related species (Gupta *et al.,* 2003; La Rota *et al.,* 2005; Zhang *et al.,* 2005). When financial resources are restricted, RAPDs, AFLPs and ISSRs can provide large numbers of markers with a limited investment. AFLPs, SSRs and ISSRs can provide high throughput using an automated sequencer, while RAPDs and ISSRs can be run on agarose gels with minimal investment in equipment. The effectiveness of each method may vary by species and by application. Therefore, it is reasonable to try to use more than one method, particularly at the early stages of research.

IMPACT OF THE RICE GENOME SEQUENCE: A CASE HISTORY

DNA sequence information greatly accelerates the development of molecular markers. This is evident in the history of rice microsatellite marker proliferation coinciding with the release of data from rice genome sequencing projects. Figure 1

FIGURE 1
Progress of microsatellite marker development in rice



tracks the publication of rice microsatellite markers derived from screening libraries of clones and from the various categories of sequences deposited in public databases. The earliest method of developing microsatellite markers in rice was by using microsatellite sequences as probes to isolate clones from genomic libraries (Zhao and Kochert, 1993; Wu and Tanksley, 1993; Panaud, Chen and McCouch, 1996; Akagi *et al.*, 1996; Chen *et al.*, 1997; Temnykh *et al.*, 2000). In 1996, Akagi *et al.* used microsatellite repeats found in rice sequences from database searches to develop 35 new markers and in 2000, Temnykh *et al.* published 91 new microsatellite markers developed from expressed sequence tag (EST) sequences. Temnykh *et al.* (2001) developed 200 new markers, mostly from end sequences of rice bacterial artificial chromosomes (BACs). However, the most dramatic increase in microsatellite markers (2 240 new markers in 2002 and 25 000 in 2004) was made possible primarily though the use of whole genome shotgun sequences (McCouch *et al.*, 2002; G. Wilson, personal communication). Complete genome sequence provides an additional advantage in electronically determining the position of new markers on genetic and physical maps. However, full genomic sequence is not a requirement for microsatellite marker development, and there are a number of microsatellite markers that have been developed for a wide array of crop species (Table 3) without the benefit of full genomic sequence.

MARKER-ASSISTED SELECTION STRATEGIES AND EXAMPLES

MAS in a breeding context involves scoring indirectly for the presence or absence of a desired phenotype or phenotypic component based on the sequences or banding patterns of molecular markers located in or near the genes controlling the phenotype. The sequence polymorphism or banding pattern of the molecular marker is indicative of the presence or absence of a specific gene or chromosomal segment that is known to carry a desired allele.

DNA markers can increase screening efficiency in breeding programmes in a

TABLE 3
Examples of SSR markers available across different plant species

Common name	Species	Number of SSRs	Reference
Rice	*Oryza sativa*	2240	McCouch *et al.*, 2002
Maize	*Zea mays*	1669	MapPairs (mp.invitrogen.com)
Soybean	*Glycine max*	597	MapPairs (mp.invitrogen.com)
Cassava	*Manihot esculenta*	318	MapPairs (mp.invitrogen.com)
Arabidopsis	*Arabidopsis thaliana*	290	MapPairs (mp.invitrogen.com)
Cotton	*Gossypium* spp.	217	MapPairs (mp.invitrogen.com)
Sugar cane	*Saccharum* spp.	200	www.intl-pag.org/pag/9/abstracts/W30_04.html
Wheat	*Triticum aestivum*	193	MapPairs (mp.invitrogen.com)
Grape	*Vitis vinifera*	152	no
Groundnut	*Arachis hypogaea*	110	Ferguson *et al.*, 2004
Cucumber	*Cucumis sativus*	110	Fazio, Staub and Chung, 2002
Peach	*Prunus persica*	109	Aranzana *et al.*, 2004
Kiwifruit	*Actinidia* spp.	105	Testolin *et al.*, 2001
Barley	*Hordeum vulgare*	44	MapPairs (mp.invitrogen.com)
Potato	*Solanum tuberosum*	31	Ghislain *et al.*, 2004
Pine trees	*Pinus* spp.	28	MapPairs (mp.invitrogen.com)
Banana	*Musa* spp.	28	MapPairs (mp.invitrogen.com)
Sweet potato	*Ipomoea batatas*	26	MapPairs (mp.invitrogen.com)
Sugar beet	*Beta vulgaris*	25	www.intl-pag.org/pag/10/abstracts/PAGX_W306.html
Eggplant	*Solanum melongena*	23	www.intl-pag.org/pag/11/abstracts/P3b_P181_XI.html

From: Thomson, Septiningsih and Sutrisno, 2003 (reprinted with permission of author)

number of ways. For example, they provide:

- the ability to screen in the juvenile stage for traits that are expressed late in the life of the organism (i.e. grain or fruit quality, male sterility, photoperiod sensitivity);
- the ability to screen for traits that are extremely difficult, expensive or time consuming to score phenotypically (i.e. quantitatively inherited or environmentally sensitive traits such as root morphology, resistance to quarantined pests or to specific races or biotypes of diseases or insects, tolerance to certain abiotic stresses such as drought, salt and mineral deficiencies or toxicities);
- the ability to distinguish the homozygous from the heterozygous condition of many loci in a single generation without the need for progeny testing (as molecular markers are co-dominant);
- the ability to perform simultaneous MAS for several characters at one time (or to combine MAS with phenotypic or biochemical evaluation).

This section provides examples of how molecular markers are being used in breeding and germplasm evaluation. While these examples are drawn mostly from rice, they illustrate applications of MAS techniques that are used in other species.

Before molecular markers can be used for selection purposes, their association with genes or traits of interest must be firmly established. While the number of economically important genetic loci that have been cloned or tagged via linkage to molecular markers is still limited in most species, work towards this end is accelerating rapidly. This is particularly true in rice, due to the availability of complete genome sequence information.

Nonetheless, a great deal of time and effort is required to identify the genetic loci and specific allelic variants that are responsible for the tremendous array of

characters that breeders are concerned about in population or variety improvement programmes. Given the complexity of quantitative traits, many different lines or crosses must be carefully analysed over different years and environments to unravel important components of gene interaction. In a breeding context, understanding the genetic basis of genotype by genotype interaction (G x G) and genotype by environment interaction (G x E) is critical as the basis for predicting how QTL are likely to behave. Information from a large number of studies addressing each of these points must then be assembled into a database that offers easy access to users and allows many different kinds of data to be integrated with a simple query.

The Gramene database represents a beginning in the quest to serve this user community. Gramene is a comparative genome database for grasses and currently offers a complete inventory of all published QTL that have been identified in rice (www.gramene.org/qtl/index.html), allowing users to find information about where along the chromosome a QTL is located, what phenotype is associated with the QTL, how it was measured, what germplasm was used, what molecular markers reside nearby, what the corresponding position is on a comparative map of another grass species and with what statistical significance the QTL was detected. The database also provides a link to the published article so that users can readily find more information on the subject. Similar inventories and databases are being assembled for other families of plants and are critical to the implementation of effective molecular breeding strategies.

Comparative genome methods take advantage of the fact that some species have more developed genetic systems than others. Examples of well studied "model" organisms with available genomic sequence include species such as *Arabidopsis* and rice for plants, *Populus* (Taylor, 2002) and *Eucalyptus* (Poke *et al.*, 2005) specifically for forestry, and *Fugu* (Aparicio *et al.*, 2002) and zebrafish (Guryev *et al.*, 2006) for fisheries. Relying heavily on the use of comparative maps and comparative sequence analysis, genome databases allow researchers to make predictions about the location and phenotypic consequences of homologous genes in related species. Thus, understanding how a gene or QTL behaves in one species can potentially shortcut the process of identifying a related gene or QTL in the genetic system of another species. This approach underscores the search for QTL associated with abiotic stress tolerance in cereals. A global effort to identify loci associated with drought tolerance has recently been initiated under the umbrella of the Generation Challenge Programme (www.generationcp.org).

Markers associated with tolerance for a variety of environmental stresses rank as important targets for molecular MAS in cereal breeding because these complex traits are often prohibitively difficult to screen using classical selection techniques. Efforts to identify QTL associated with tolerance to drought, salt and mineral deficiencies or toxicities (Champoux *et al.*, 1995; Flowers *et al.*, 2000; Nguyen *et al.*, 2002; Kamoshita *et al.*, 2002; Price *et al.*, 2002; Gregorio, 2002) in a number of genetic backgrounds represent an important first step towards achieving this goal. Additional studies have specifically addressed the problems associated with G x G and G x E (Zheng *et al.*, 2000; Li *et al.*, 2003; Hittalmani *et al.*, 2003).

In the area of biotic stress, several genes have been cloned and characterized for resistance to major diseases such as bacterial blight and blast (Song *et al.*, 1995;

Yoshimura *et al.,* 1998; Wang *et al.,* 1999; Bryan *et al.,* 2000; Sun *et al.,* 2004) and many other genes for disease resistance have been tagged with linked markers. This opens the door for targeted approaches to MAS (Valent *et al.,* 2001). While the disease resistance literature is too vast to summarize here, it is important to note that advances in this area are having an impact on varietal improvement programmes (www.syix.com/rrb/98rpt/MarkerAssist.htm). Pyramiding of resistance genes into a single variety and the construction of multiline varieties, each with one or more R genes (resistance genes) that can be used in various combinations, are under way to develop more durable forms of disease and insect resistance (Yoshimura *et al.,* 1992; Yoshimura *et al.,* 1995; Hittalmani *et al.,* 1995; Blair and McCouch, 1997; Ndjiondjop *et al.,* 1999; Davierwala *et al.,* 2001; Su *et al.,* 2002; Conaway-Bormans *et al.,* 2003; Lorieux *et al.,* 2003; Hayashi *et al.,* 2004).

Marker-based selection is also helpful in attempts to transfer genes from exotic germplasm into cultivated lines. In rice, several workers have used RFLP and SSR markers to monitor introgression of brown planthopper resistance from *O. officinalis* (Kochert, Jena and Zhao, 1990), bacterial blight resistance from *O. longistaminata (*Ronald *et al.,* 1992), aluminum tolerance or yield and quality-related traits from *O. rufipogon* (Nguyen *et al.,* 2002; Thomson *et al.,* 2003; Septiningsih *et al.,* 2003a, b) or from other wild species such as *O. glumaepatula* (Brondani *et al.,* 2002) or *O. glaberrima* (Jones *et al.,* 1997; Lorieux *et al.,* 2003) into cultivated *O. sativa* backgrounds. Marker-assisted introgression strategies have also been used in a number of livestock breeding programmes but, because of longer generation intervals and lower reproductive rates, this is generally feasible for genes of large effect (Dekkers, 2004; Chapter 10). Identifying the recombinants with the least amount of donor DNA flanking the genes of interest is enhanced by the use of molecular markers (Monna *et al.,* 2002; Takeuchi *et al.,* 2003; Blair, Panaud and McCouch, 2003). In these examples, MAS offers a powerful strategy for making efficient use of the wealth of useful genetic variation that exists in the early landraces and wild species of cultivated food crops (Tanksley and McCouch, 1997).

As this kind of information accumulates, MAS permits rapid identification of individuals that may contain only one genetic component of a complex trait. Once identified, such an individual can be crossed with another individual in a breeding programme so that multiple, complementary genes are combined to optimize a quantitatively inherited trait. Individuals containing only one gene of interest often defy accurate phenotypic identification where polygenic traits are concerned because various types of epistasis, or gene interaction, may be required to generate the phenotype of interest (Yamamoto *et al.,* 2000; Zheng *et al.,* 2000).

Linkage disequilibrium (LD) mapping is another marker-assisted approach that provides important information that is immediately relevant to breeding programmes (Remington, Ungerer and Purugganan, 2001; Flint-Garcia, Thornsberry and Buckler, 2003). Using collections of distantly related germplasm accessions rather than populations derived from bi-parental crosses allows researchers to explore the relationship between phenotype and genotype in materials that have been amply tested over years and environments, often as part of an applied breeding programme. This provides

critical information about how specific combinations of genes and alleles interact in relevant varietal backgrounds and allows breeders to compare the phenotypic effect of genes or chromosomal segments that have been inherited from a common ancestor and selected in multiple-cross combinations.

In addition to the use of MAS in traditional crossing and selection programmes, breeders also have opportunities to adjust particular traits or phenotypes via the introduction of genes using a transgenic approach (Ye *et al.,* 2000; James, 2003; Nuffield Council on Bioethics, 2004). Once introduced into the gene pool, a transgene can be tracked with the aid of molecular markers (designed to tag the transgene sequence itself) through subsequent crosses, just as would be done for any other gene of interest in a breeding programme.

Another use of molecular markers in variety improvement involves marker-assisted germplasm evaluation (Xu, Ishii and McCouch, 2003). Population structure analysis offers insight about how diversity is partitioned within a species and can help define clusters, or subpopulations, of germplasm that are likely to contain high frequencies of particular alleles (Garris, McCouch and Kresovich, 2003). This type of analysis can also guide allele mining efforts aimed at identifying valuable accessions in a germplasm collection for use as parents in a breeding programme. Such approaches have the potential to make parental selection more efficient, to expand the gene pool of modern cultivars and ultimately to speed up the development of productive new varieties. As information is generated about which genes and alleles are associated with phenotypic characters of agronomic importance, and as the complex interactions among genes are enumerated in the context of specific gene pools and the environments to which they are adapted, breeders are increasingly empowered to make predictions about how to combine diverse alleles productively.

To exploit molecular breeding strategies fully, information resources must be developed so that the overwhelming amount of information about genes, alleles and natural genetic variation can be funnelled into a useful tool for breeding applications. This will involve a very different approach to information resources than currently employed by the large genome databases, which are oriented towards genomics researchers and molecular biologists rather than the breeding community. Nonetheless, a few examples offer beacons of inspiration in this area, including the emerging International Rice Information System (IRIS) database (Bruskiewich *et al.,* 2003; www.icis.cgiar.org/), the GeneFlow database (www.geneflow.com), the marker-assisted selection wheat (MASwheat) database (http://maswheat.ucdavis.edu/) and software such as Real Time QTL (http://zamir.sgn.cornell.edu/Qtl/Html/home.htm).

In conclusion, genomics research is generating information about the location and phenotypic consequences of specific genes and alleles in a wide range of species. This information can be translated into tools for breeders. Molecular marker technology can benefit breeding objectives by increasing the efficiency and reliability of selection and by providing essential insights into how genes behave in different environments and in different genetic backgrounds. Once genes and QTL are identified, markers allow interesting alleles to be traced through the pedigrees of breeding programmes or mined out of germplasm collections to serve as the basis for future varietal improvement. Using markers in combination with both QTL and association approaches, the

effect of specific alleles on a phenotype can be monitored with relative precision. As all this information is assembled and organized in databases that provide easy querying capabilities for plant breeders, breeders will take advantage of the power that comes from the application of genome based strategies for plant improvement.

REFERENCES

Akagi, H., Yokozeki, Y., Inagaki, A. & Fujimura, T.Y. 1996. Microsatellite DNA markers for rice chromosomes. *Theor. Appl. Genet*. 93: 1071–1077.

Aparicio, S., Chapman, J., Stupka, E., Putman, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M.D.S., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J.K., Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D. & Brenner, S. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297: 1301–1310.

Aranzana, M.J., Pineda, A., Cosson, P., Dirlewanger, E., Ascasibar, J., Cipriani, G., Ryder, C.D., Testolin, R., Abbott, A., King, G.J., Iezzoni, A.F. & Arus, P. 2003. A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. *Theor. Appl. Genet.* 106: 819–825.

Barrett, M.T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P.S., Yakhini, Z., Bruhn, L. & Laderman, S. 2004. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc. Nat. Acad. Sci. USA* 101: 17765–17770.

Blair, M.W. & McCouch, S.R. 1997. Microsatellite and sequence-tagged site markers diagnostic for the rice bacterial leaf blight resistance gene *xa-5*. *Theor. Appl. Genet.* 95: 174-184.

Blair, M.W., Panaud O. & McCouch, S.R. 1999. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 98: 780–792.

Borevitz, J. O., Liang, D., Plouffe, D., Chang, H.S., Zhu, T., Weigel, D., Berry, C.C., Winzeler, E. & Chory, J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res*. 13: 513–523.

Brondani, C., Rangel, N., Brondani V. & Ferreira, E. 2002. QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. *Theor. Appl. Genet.* 104: 1192–1203.

Bruskiewich, R.M., Cosico, A.B., Eusebio, W., Portugal, A.M., Ramos, L.M., Reyes, M.T., Sallan, M.A., Ulat, V.J., Wang, X., McNally, K.L., Sackville Hamilton, R. & McLaren, C.G. 2003. Linking genotype to phenotype: the International Rice Information System (IRIS). *Bioinformatics 19 Suppl 1*: i63–i65.

Bryan, G.T., Wu, K.S., Farrall, L., Jia, Y., Hershey, H.P., McAdams, S.A., Faulk, K.N., Donaldson, G.K., Tarchini, R. & Valent, B. 2000. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene Pi-ta. *The Plant Cell* 12: 2033–2046.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang Q., Kresovich, S. & Wessler, S.R. 2000. Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. *Proc. Nat. Acad. Sci. USA* 97: 10083–10089.

Champoux, M.C., Sarkarung, G.W.S., Mackill, D.J., O'Toole, J.C., Huang, N. & McCouch, S.R. 1995. Locating genes associated with root morphology and drought avoidance via linkage to molecular markers. *Theor. Appl. Genet.* 90: 969–981.

Chen, X., Cho Y.G. & McCouch, S.R. 2002. Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Mol. Genet. Genomics* 268: 331–343.

Chen, X., Temnykh, S., Xu, Y., Cho, X.G. & McCouch, S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor. Appl. Genet*. 95: 553–567.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R. & Henikoff, S. 2004. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* 37: 778–786.

Conaway-Bormans, C.A., Marchetti, M.A., Johnson, C.W., McClung, A.M. & Park, W.D. 2003. Molecular markers linked to the blast resistance gene Pi-z in rice for use in marker-assisted selection. *Theor. Appl. Genet*. 107: 1014–1020.

Davierwala, A.P., Reddy, A.P., Lagu, M.D., Ranjekar, P.K. & Gupta, V.S. 2001. Marker assisted selection of bacterial blight resistance genes in rice. *Biochem. Genet*. 39: 261–278.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82 E-Suppl: E313–E328.

Doyle, J.J., Morgante, M., Tingey S.V. & Powell, W. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Mol. Biol. Evol.* 15: 215–218.

Edwards, J.D., Lee, V.M. & McCouch, S.R. 2004. Sources and predictors of resolvable indel polymorphism assessed using rice as a model. *Mol. Genet. Genomics* 271: 298–307.

Fazio, G., Staub, J.E. & Chung, S.M. 2002. Development and characterization of PCR markers in cucumber. *J. Am. Soc. Hortic. Sci.* 127: 545–557.

Ferguson, M.E., Burow, M.D., Schulze, S.R., Bramel, P.J., Paterson, A.H., Kresovich, S. & Mitchell, S. 2004. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). *Theor. Appl. Genet.* 108: 1064–1070.

Feschotte, C., Jiang N. & Wessler, S.R. 2002. Plant transposable elements: where genetics meets genomics. *Nature Revs. Genet.* 3: 329–341.

Flint-Garcia, S.A., Thornsberry, J.M. & Buckler, E.S.T. 2003. Structure of linkage disequilibrium in plants. *Ann. Rev. Plant Biol.* 54: 357–374.

Flowers, T.J., Koyama, M.L., Flowers, S.A., Sudhakar, C., Singh, K.P. & Yeo, A.R. 2000. QTL: their place in engineering tolerance of rice to salinity. *J. Exp. Bot.* 51: 99–106.

Garris A., McCouch, S. & Kresovich, S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics* 165: 759–769.

Ghislain, M., Spooner, D.M., Rodriguez, F., Villamon, F., Nunez, J., Vasquez, C., Waugh, R. & Bonierbale, M. 2004. Selection of highly informative and user friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theor. Appl. Genet.* 108: 881–890.

Goodwin, I.D., Aitken, E.A.B. & Smith, L.W. 1997. Application of inter-simple sequence repeat (ISSR) markers to plant genetics. *Electrophoresis* 18: 1524–1528.

Gregorio, G.B. 2002. Progress in breeding for trace minerals in staple crops. *J. Nutr*. 132: 500S–502S.

Gupta, M., Chyi, Y.S., Romero-Severson, J. & Owen, J.L. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.* 89: 998–1006.

Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N. & Balyan, H.S. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics* 270: 315–323.

Guryev, V., Koudijs, M.J., Berezikov, E., Johnson, S.L., Plasterk, R.H., van Eeden, F.J. & Cuppen, E. 2006. Genetic variation in the zebrafish. *Genome Res.* **16:** 491–497.

Gut, I.G. 2001. Automation in genotyping of single nucleotide polymorphisms. *Hum. Mutat.* 17: 475–492.

Hayashi, K., Hashimoto, N., Daigen, M. & Ashikawa, I. 2004. Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theor. Appl. Genet*. 108: 1212–1220.

Hittalmani, S., Foolad, M.R., Mew, T., Rodriguez, R.L. & Huang, N. 1995. Development of a PCR-based marker to identify rice blast resistance gene, Pi-2(T), in a segregating population. *Theor. Appl. Genet.* 91: 9–14.

Hittalmani, S., Huang, N., Courtois, B., Venuprasad, R., Shashidhar, H.E., Zhuang, J.Y., Zheng, K.L., Liu, G.F., Wang, G.C., Sidhu, J.S., Srivantaneeyakul, S., Singh, V.P., Bagali, P.G., Prasanna, H.C., McLaren, G. & Khush, G.S. 2003. Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. *Theor. Appl. Genet*. 107: 679–690.

Innan, H., Terauchi, R., Kahl, G. & Tajima, F. 1999. A method for estimating nucleotide diversity from AFLP data. *Genetics* 151: 1157-1164.

Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucl. Acids Res*. 29: E25.

James, C. 2003. Global review of commercialized transgenic crops: 2002 Feature: *Bt* Maize. *ISAAA Briefs.*

Jones, M.P., Dingkuhn, M. Aluko, G.K. & Semon, M. 1997. Interspecific *O. sativa* L.x *O. glaberrima* Steud. Progenies in upland rice improvement. *Euphytica* 92: 237–246.

Kamoshita, A., Wade, J., Ali, L., Pathan, S., Zhang, J., Sarkarung, S. & Nguyen, T. 2002. Mapping QTLs for root morphology of a rice population adapted to rain-fed lowland conditions. *Theor. Appl. Genet.* 104: 880–893.

Kochert, G., Jena, K. & Zhao, X.P. 1990. Genome characterization and introgression. In *4th Ann. Mtg. Rockefeller Foundation's Internat. Program on Rice Biotech.* Los Baños, Philippines, IRRI.

Konieczny, A. & Ausubel, F.M. 1993. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J*. 4: 403–410.

Kwok, P.Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Ann. Rev. Genomics Hum. Genet*. 2: 235–258.

La Rota, M., Kantety, R.V., Yu, J.K. & Sorrells, M.E. 2005. Non-random distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6: 23.

Li, Z.K., Yu, S.B., Lafitte, H.R., Huang, N., Courtois, B., Hittalmani, S., Vijayakumar, C.H., Liu, G.F., Wang, G.C., Shashidhar, H.E., Zhuang, J.Y., Zheng, K.L., Singh, V.P., Sidhu, J.S., Srivantaneeyakul, S. & Khush, G.S. 2003. QTL x environment interactions in rice. I. Heading date and plant height. *Theor. Appl. Genet.* 108: 141–153.

Lorieux, M., Reversat, G., Garcia Diaz, S.X., Denance, C., Jouvenet, N., Orieux, Y., Bourger, N., Pando-Bahuon, A. & Ghesquiere, A. 2003. Linkage mapping of Hsa-1(Og), a resistance gene of African rice to the cyst nematode, *Heterodera sacchari*. *Theor. Appl. Genet*. 107: 691–696.

McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Khush, G.S., Coffman, W.R. & Tanksley, S.D. 1988. Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* 76: 815–829.

McCouch, S.R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y.G., Huang, N., Ishii, T. & Blair, M. 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol. Biol.* 35**:** 89–99.

McCouch, S.R., Teytelman, L., Xu, Y., Lobos, K.B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y., Zhang, Q., Kono, I., Yano, M., Fjellstrom, R., DeClerck, G., Schneider, D., Cartinhour, S., Ware, D. & Stein, L. 2002. Development and mapping of 2 240 new SSR markers for rice (*Oryza sativa* L.) (supplement). *DNA Res*. 9: 257–279.

Monna, L., Lin, X., Kojima, S., Sasaki, T. & Yano, M. 2002. Genetic dissection of a genomic region for a quantitative trait locus, Hd3, into two loci, Hd3a and Hd3b, controlling heading date in rice. *Theor. Appl. Genet.* 104**:** 772–778.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1**:** 263–273.

Ndjiondjop, M.N., Albar, L., Fargette, D., Fauquet, C. & Ghesquiere, A. 1999. The genetic basis of high resistance to rice yellow mottle virus (RYMV) in cultivars of two cultivated rice species. *Plant Disease* 83: 931–935.

Nei, M. & Miller, J.C. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* 125: 873–879.

Nguyen, B.D., Brar, D.S., Bui, B.C., Nguyen, T.V., Pham, L.N. & Nguyen, H.T. 2002. Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 106: 583–593.

Nielsen, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154: 931–942.

Nuffield Council on Bioethics. 2004. *The use of genetically modified crops in developing countries*. London, UK.

Panaud, O., Chen, X. & McCouch, S.R. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol. Gen. Genet*. 252: 597–607.

Paran, I. & Michelmore, R.W. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Genetics* 85: 985–993.

Poke, F.S., Vaillancourt, R.E., Potts, B.M. & Reid, J.B. 2005. Genomic research in *Eucalyptus*. *Genetica* 125: 79–101.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. & Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225–238.

Price, A.H., Townend, J., Jones, M.P., Audebert, A. & Courtois, B. 2002. Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant Mol. Biol*. 48: 683–695.

Remington, D.L., Ungerer, M.C. & Purugganan, M.D. 2001. Map-based cloning of quantitative trait loci: progress and prospects. *Genet. Res.* 78: 213–218.

Ronald, P.C., Salmeron, J.M., Carland, F.M. & Staskawicz, B.J. 1992. The cloned avirulence gene avrPto induces disease resistance in tomato cultivars containing the Pto resistance gene. *J. Bacteriol.* 174: 1604–1611.

Rungus, D., Berube, Y., Zhang, J., Ralph, S., Ritland, C.E., Ellis, B.E., Douglas, C., Bohlmann, J. & Ritland, K. 2004. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor. Appl. Genet*. 109: 1283–1294.

Septiningsih, E.M., Prasetiyono, J., Lubis, E., Tai, T.H., Tjubaryat, T., Moeljopawiro, S. & McCouch, S.R. 2003a. Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor. Appl. Genet*. 107**:** 1419–1432.

Septiningsih, E.M., Trijatmiko, K.R., Moeljopawiro, S. & McCouch, S.R. 2003b. Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor. Appl. Genet.* 107: 1433–1441.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C. & Ronald, P. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270: 1804–1806.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol*. 98: 503–510.

Su, C.C., Cheng, X.N., Zhai, H.Q. & Wan, J.M. 2002. Detection and analysis of QTL for resistance to the brown planthopper, *Nilaparvata lugens* (Stal), in rice (*Oryza sativa* L.), using backcross inbred lines]. *Yi Chuan Xue Bao* 29: 332–338.

Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S. & Zhang, Q. 2004. Xa26, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. *Plant J.* 37: 517–527

Takeuchi, Y., Lin, S.Y., Sasaki, T. & Yano, M. 2003. Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. *Theor. Appl. Genet.* 107: 1174–1180.

Tanksley, S.D. & McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063–1066.

Taylor, G., 2002. *Populus*: *arabidopsis* for forestry. Do we need a model tree? *Ann. Bot. (London)* 90: 681–689.

Temnykh, S.P., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho,Y-G., Ishii, T. & McCouch, S.R. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa L.*). *Theor. Appl. Genet*. 100: 697–712.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. & McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11: 1441–1452.

Testolin, R., Huang, W.G., Lain, O., Messina, R., Vecchione, A. & Cipriani, G. 2001. A kiwifruit (*Actinidia* spp.) linkage map based on microsatellites and integrated with AFLP markers. *Theor. Appl. Genet.* 103**:** 30–36.

Thomson M., Septiningsih E. and Sutrisno. 2003. Workshop on IPR in relation to Plant Breeding, Cipanas, Indonesia, 10–18 November.

Thomson, M.J., Tai, T.H., McClung, A.M., Lai, X.H., Hinga, M.E., Lobos, K.B., Xu, Y., Martinez, C.P. & McCouch, S.R. 2003. Mapping quantitative trait loci for yield, yield components and

morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor. Appl. Genet.* 107: 479–493.

Upholt, W.B. 1977. Estimation of DNA sequence divergence from comparison of restriction endonuclease digests. *Nucl. Acids Res*. 4: 1257–1265.

Valent, B., Bryan, G.T., Jia, Y., Farrall, L., McAdams, S.A., Faulk, K.N. & Levy, M. 2001. Enhancing deployment of rice blast resistance genes: opportunities from cloning a resistance gene/avirulence gene pair. *In* G.J. Khush, D.S. Brar & B. Hardy, eds. *Rice genetics IV,* pp. 309-322. Enfield, NH, USA, Science Publishers.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.* 23: 4407–4414.

Wang Z.X., Yano, M., Yamanouchi, U., Iwamoto, M.I., Monna, L., Hayasaka, H., Katayose, Y. & Sasaki, T. 1999. The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J.* 19: 55–64.

Weber, J.L. & May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet*. 44: 388–396.

Welsh, J. & McClelland, M. 1990. Finger-printing genomes using PCR with arbitrary primers. *Nucl. Acids Res*. 18: 7213–7218.

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A. & Kilian, A. 2004. Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* 101: 9915–9920.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski J.A. & Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18: 6531–6535.

Wu, K.S. & Tanksley, S.D. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet.* 241: 225–235.

Xu, Y., Ishii, T. & McCouch, S.R. 2003. Marker-assisted evaluation of germplasm resources for plant breeding, *In* T.W. Mew, D.S. Brar, S. Peng, D. Dawe & B. Hardy, eds. *Rice science: innovations and impact for livelihoods. Proc. Internat. Rice Res. Conf.,* pp. 213-230. Beijing, China, IRRI, Chinese Academy of Engineering and Chinese Academy of Agricultural Sciences.

Yamamoto, T., Lin, H., Sasaki, T. & Yano, M. 2000. Identification of heading date quantitative trait locus Hd6 and characterization of its epistatic interactions with Hd2 in rice using advanced backcross progeny. *Genetics* 154: 885–891.

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P. & Potrykus, I. 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287: 303–305.

Yoshimura, S., Yoshimura, A., Saito, A., Kishimoto, N., Kawase, M., Yano, M., Nakagahara, M., Ogawa, T. & Iwata, N. 1992. RFLP analysis of introgressed chromosomal segments in three near-isogenic lines of rice for bacterial blight resistance genes, Xa-1, Xa-3 and Xa-4. *Jpn. J. Genet*. 67: 29–37.

Yoshimura. S., Iwata, N.A.Y., McCouch, S.R., Abenes, M.L., Baraoidan, M.R., Mew, T.W. & Nelson, R.J. 1995. Tagging and combining bacterial blight resistance genes using RAPD and RFLP markers. *Mol. Breed.* 1: 375–387.

Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M., Iwata, N. & Sasaki, T. 1998. Expression of Xa1, a bacterial blight-resistance in rice, is induced by bacterial inoculation. *Proc. Nat. Acad. Sci. USA* 95: 1663–1668.

Zane, L., Bargelloni, L. & Atarnello, T.P. 2002. Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11: 1–16.

Zhao, X. & Kochert, G. 1993. Phylogenetic distribution and genetic mapping of a (GGC)n microsatellite from rice (*Oryza sativa* L.). *Plant Mol. Biol.* 21: 607–614.

Zhang, L.Y., Bernard, M., Leroy, P., Feuillet, C. & Sourdille, P. 2005. High transferability of bread wheat EST-derived SSRs to other cereals. *Theor. Appl. Genet.* 111: 677–687.

Zheng, H.G., Babu, R.C., Pathan, M.S., Ali, L., Huang, N., Courtois, B. & Nguyen, H.T. 2000. Quantitative trait loci for root-penetration ability and root thickness in rice: comparison of genetic backgrounds. *Genome* 43: 53–61.

Zietkietkiewicz, E.A.R., Rafalski, A. & Labuda, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176–183.

CHAPTER 4

Marker-assisted selection in wheat: evolution, not revolution

Robert Koebner and Richard Summers

SUMMARY

This chapter reviews the uptake of marker-assisted selection (MAS) in wheat in a European context. Although less intense than the scale of its application in maize, reflecting the fact that maize varieties are predominantly F_1 hybrids, the use of MAS in wheat has grown over the last few years. This growth has been encouraged by an increase in the number of amenable target traits, but more significantly by a combination of technological improvements, particularly in the areas of DNA acquisition, laboratory management systems and integration into the breeding cycle, which together have served to reduce the per unit cost of each data point. Microsatellites (simple sequence repeats [SSRs]) are, and will likely remain for some time, the marker of choice because of their flexibility and the knowledge base associated with them. Some current examples are provided of the use of MAS in a major United Kingdom commercial breeding programme.

INTRODUCTION

Wheat is a very important world staple crop. The 2005 United States Department of Agriculture (USDA) estimates for the global production of wheat (both bread and durum) and maize are, respectively, 627 million tonnes and 708 million tonnes. In Europe, bread wheat is without doubt the most important broad-acre crop, with a production in the extended European Union of 25 states of 115 million tonnes (maize 48 million tonnes). The largest production and highest productivity of bread wheat are achieved in northwest Europe. Historically, wheat has been bred largely by government-sponsored national and regional programmes, but the introduction of plant variety rights into Europe in the 1960s encouraged participation by the private sector. Currently, wheat breeding in northwest Europe is almost exclusively carried out by private companies, with some research underpinning by the public sector. Breeders continue to be successful in the production of high-yielding, disease-resistant, high-quality varieties and, in the United Kingdom at least, genetic advances for yield have been running at between 0.5 to 1 percent per annum for many years.

Wheat is a naturally inbreeding species, and although a level of heterosis can be demonstrated, difficulties in enforcing cross-pollination in a reliable and cost-effective way have hindered the development of any significant contribution of F_1 hybrids to the variety pool. Most varietal development programmes are therefore based on versions of the long-established pedigree breeding system, where large F_2 populations are generated and conventional phenotypic selection is carried out in early generations for highly heritable, qualitative traits (such as disease resistance) and in later ones for quantitative traits (primarily yield and quality). Thus, most varieties are bred and grown as inbred, pure breeding lines. As a result, the unit value of seed and economic margins for breeders are low. By contrast, maize is a naturally out-crossing species that shows highly significant levels of heterosis. This has resulted in the majority of maize breeding being geared to the production of F_1 hybrids. In industrialized countries, maize hybrid breeding has for some time been dominated by a small number of large private sector companies that are able to sustain profitability through their control over the genotype of their varieties. No revenue is lost as a result of the use of farm-saved seed, and the inbred components of a successful hybrid are not available to competitors to use as parental material for their own varietal improvement programmes. This has far-reaching implications on the feasibility of MAS in maize, and largely explains the lead that maize enjoys over wheat in the deployment of MAS technology.

The continuing development of molecular marker technology over the last decade has been a happy by-product of "big biology" genomics research. As recently as 1996, the definition of 5 000 SSR loci in the human genome merited a major publication in *Nature* (Dib *et al.*, 1996), but the number of known human single nucleotide polymorphisms (SNPs) now runs into millions. Thus, although marker availability, potentially at least, is no longer limiting in crops, and the clear potential benefits of marker deployment to plant breeding are undisputed, only relatively recently has it begun to make more than a marginal impact on breeding methodology. Even in maize, where the level of DNA marker polymorphism is high, large-scale deployment of MAS did not gather any significant momentum until more than 15 years after the publication of the first restriction frag-

ment length polymorphism (RFLP)-based maize genetic map. In the less genetically variable cereals, prominently wheat, the level of polymorphism is not now in practice likely to represent the major constraint to MAS uptake, although in the past it was argued that this was the case. What has changed in recent times is that current marker technology, and systems of DNA acquisition, laboratory management and integration into the breeding cycle, have all developed to the extent where the benefits of MAS can be increasingly realized in actual practice. As many of these improvements are incremental rather than sudden, we argue that the trends in MAS application in wheat are characteristically evolutionary rather than revolutionary.

TARGET TRAITS FOR MAS IN NORTHWEST EUROPEAN WINTER WHEAT BREEDING

The use of MAS to date has a history of about 20 years, and until recently involved the exploitation of just two non-DNA-based assays. The first, which has been retained with only slight modifications since its inception, exploits a correlation between bread-making quality and allelic status at the *Glu-1* (endosperm storage protein subunit) loci. It uses electrophoretic profiles obtained by the straightforward, robust and cheap procedure sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) from crude seed protein extracts, which have been shown to be partially predictive of end-use quality. The second is predictive for the presence of the gene *Pch1*, which confers a high level of resistance to eyespot, a stem base disease that is difficult to screen using conventional pathology methods. Both these targets have in the meanwhile become assayable by polymerase chain reaction (PCR)-based assays, although SDS-PAGE remains in routine use thanks to its flexibility and cost effectiveness. In recent years, the number of loci for which DNA-based assays have been generated has increased dramatically, the majority using PCR as a technology platform. Over 50 of these are described (specifically in a United States of America context) at http://maswheat.ucdavis.edu/, which reports the output of an ongoing United States Department of Agriculture (USDA)-funded programme. The focus is heavily on disease and pest resistance, reflecting the generally simple inheritance of genes conferring these traits.

Some of the above traits are of sufficient relevance to the United Kingdom context that identical or equivalent assays have been incorporated in a number of breeding programmes, where they are used as guides to parental selection and/or in early generation selection. Prominent among these are markers for the genes *Rht-1* (responsible for the "Green Revolution" semi-dwarfism), *Pinb* (grain texture), *Pch1*, *Lr37/Yr17* (a gene complex conferring resistance to two of the most important leaf fungal pathogens) and the wheat/rye translocation 1B/1R (which is associated with high levels of yield). Emerging MAS targets are necessarily programme-dependent, but the broad focus is on quantitative trait locus (QTL) targets that could have a major impact on breeding efficiency. In the United Kingdom, as elsewhere worldwide, current focus is on resistance to the diseases *Fusarium* head blight (FHB), *Septoria tritici* blotch (STB) and barley yellow dwarf virus (BYDV), and on durable resistance to yellow rust. Other current targets, more specific to the United Kingdom and northwest European context, but in routine use, are resistance to the insect pest orange blossom midge (OBM), and soil-borne mosaic virus (SBMV).

FHB

The importance of FHB is less in its effect on yield reduction, but rather on the potentially damaging reduction in grain quality associated with infected grain, which can be heavily contaminated by the fungal tricothecin toxins. An important source of FHB resistance originates from the Chinese variety Sumai 3, and a major component of this resistance (up to 50 percent) has been associated with a single QTL (Waldron *et al.,* 1999; Anderson *et al.,* 2001; Buerstmayr *et al*., 2002). While this QTL is largely effective in preventing the spread of the pathogen following infection, a further QTL that gives a significant degree of protection against initial infection has been mapped to a different chromosome (Buerstmayr *et al.,* 2003). Selection for FHB resistance by conventional means is complicated both by the quantitative nature of the Sumai 3 resistance and by difficulties in ensuring even and reliable artificial infections in breeding nurseries. However, SSR-based MAS protocols have been developed for both QTL (see http://maswheat.ucdavis.edu/ and Buerstmayr *et al.,* 2003), and the urgency of breeding for resistance has ensured that increasing use is being made of such assays. Both these QTL in concert do not explain all the genetic resistance of Sumai 3 to FHB, but the remainder appears to be determined by QTL of minor effects and/or pleiotropic effects associated with an ear morphology, which is inconsistent with a northwest European winter wheat ideotype.

STB

STB of wheat is caused by the fungus *Mycosphaerella graminicola* (syn. *Septoria tritici*), and in recent years has become the major leaf disease of wheat in many regions of the world. In past years, good levels of control were achieved by the application of strobilurin fungicides, but their heavy use has led to the emergence of pathogen strains that cannot be so easily controlled by chemical means. A number both of major genes giving near-complete resistance to specific races of the pathogen and of quantitative race non-specific resistances with polygenic inheritance have been defined, and one of the former, *Stb6*, which maps close to the SSR locus *Xgwm369* on chromosome 3A (Chartrain, Brading and Brown, 2004), is common in many gene pools. This ensures that the gene has been retained in elite materials, and its known map position has made it relatively straightforward to use a marker assay to track its presence in breeding populations.

BYDV

Significant grain yield losses are attributable to natural infections of BYDV, and no major source of resistance has been identified to date in wheat. Control is achieved in the absence of genetic resistance by insecticidal spray, which is associated with both an economic and an environmental cost. However, a potent resistance is present in the related species *Thinopyrum intermedium*. It is possible to generate sexual hybrids between wheat and this grass, but the F_1 plants are self-sterile and either have to be rescued by chromosome doubling or back-crossed to wheat. By this route, a distal segment of the grass chromosome that carries the BYDV resistance gene *Bdv2* has been introduced into wheat. As this introgression comprises a significant length of non-wheat chromosome, it has been relatively straightforward to generate markers suitable for MAS use (Ayala *et al.,* 2001a; Zhang *et al.,* 2004). A MAS approach for screening is attractive because artificial inoculation involves the propagation of virus-bearing aphids, while natural infections

are unreliable. Interestingly, unlike the experience with many alien introgression segments, no obvious negative effects of its presence on agronomic performance have yet been detected either in International Maize and Wheat Improvement Center (CIMMYT) trials (Ayala *et al.,* 2001b) or at RAGT Seeds (Cambridge, UK).

Durable resistance to yellow rust

Yellow rust is historically the most damaging of the leaf fungal pathogens in temperate Europe. Control has been achieved in the past largely by a combination of fungicide application and of combinations of major seedling resistance genes, of which a significant number have been described in the literature. However, like most race-specific resistances, most of these major genes have lost their effectiveness, and this has led to a renewed effort in the definition of partial or adult plant resistances to this disease. The French variety Cappelle-Desprez dominated the wheat crop across France and the United Kingdom during the 1960s and 1970s, and maintained its level of adult resistance to yellow rust over the whole of this period. A major part of the genetic basis for this durable resistance was located to a translocated wheat chromosome (Law and Worland, 1997), and this has been confirmed by a rigorous QTL analysis (Mallard *et al.,* 2005), which has provided a number of informative SSR markers for this effect. Other independent sources of adult resistance have been identified in French and Eastern European germplasm at RAGT Seeds, and the major QTL responsible have been defined and marked.

OBM

OBM larvae feed on developing grain and heavy infestations result in a significant reduction in grain quality and some loss in yield. As for many sporadic pests, phenotypic screening is unreliable and an indirect means of selection would be valuable. The gene *Sm1* confers resistance to OBM (*Sitodiplosis mosellana*) by the expression of an antibiotic that kills or slows the development of larvae. Thomas *et al.* (2005) defined the map position of *Sm1* and proposed a close linkage with an SSR locus *Xbarc35*. This linkage remains to be validated in United Kingdom breeding populations, as it remains unclear whether the antibiotic effect shown by a few United Kingdom wheat varieties is conferred by *Sm1.*

SBMV

SBMV is one of two known viral pathogens transmitted by the soil fungus *Polymyxa graminis* (another one being yellow mosaic virus [YMV]), and can be an important agent of yield loss in some areas. Chemical control is not feasible, and once soil is infected by the virus-bearing host, the only solutions possible are to abandon wheat culture or to use resistant varieties. Phenotyping is particularly difficult as plant infection is environmentally sensitive, and the detection of infection is laborious and prone to error. A proprietary assay for resistance to SBMV originating from European germplasm has been in routine use at RAGT Seeds since 2000 with a very high level of marker/phenotype association. More recently, a bulk segregant analysis along with a QTL approach has allowed the definition of a resistance locus to YMV from Chinese germplasm, and a number of linked SSR markers have been identified (Liu *et al.,* 2005).

HIGH-THROUGHPUT INFRASTRUCTURES

Technical considerations of DNA acquisition, laboratory information management

system (LIMS), laboratory automation and data capture and analysis are generic for any MAS set-up, and these are well covered elsewhere in this volume. The limitations affecting MAS deployment in wheat flow from the restricted revenue generated by breeding a self-pollinated, homozygous, non-hybrid product. As a result, the volume of capital investment affordable in maize is not available to a wheat MAS programme. Financial constraints also affect the development of marker platforms. It is well known that the predictive ability of a linked marker will be disrupted by recombination, and therefore that "perfect" markers are more desirable than linked ones. However, the development of genome-wide gene-based markers, pre-eminently SNPs, which are particularly suited to high-throughput genotyping on automated platforms, is still some way off. At present, an insufficient number of such assays has been established (grain hardness, semi-dwarfness and grain texture) to consider adjusting the present major genotyping methodology, which is founded on SSRs. Doubts have been raised that SNP frequency in exon sequence will be high enough to generate informative assays for many critical genes, but early experience suggests that sequence polymorphism is more than adequate in introns and other untranslated regions of wheat genes. At present, the consensus is that there is plenty of mileage left in SSR technology, and wheat maps continue to be refined by the addition of new SSR loci.

CONCLUSION

In 1999, Young set out his "cautiously optimistic vision" for MAS. Seven years on, the situation continues to crystallize. The technology itself is no longer limiting. With respect to marker availability, SSRs remain useful and SSR-based genetic maps are becoming increasingly densely populated, while SNPs may eventually represent a source of plentiful perfect markers for genes of defined function. The "big biology" spawned by the genomics revolution has brought miniaturization and automation to biological assays so that levels of throughput relevant to the wheat breeding process are becoming attainable. The issue that remains unresolved is the affordability of large-scale MAS. As wheat is a broad-acre commodity product, its value is low, and this impedes the ability of the industry to invest in MAS infrastructure to the extent that is possible for crops such as maize where the generation of F_1 hybrid seed is a viable proposition. However, as economies of scale and improvements in technology continue to drive down assay price, the penetration of MAS into commercial wheat breeding will surely grow. This growth should progressively allow a widening in the range of possible MAS targets, in particular extending to critical ones such as QTL for yield and its components (mean kernel size, kernel number per ear and number of fertile tillers per unit area). These are already widely exploited in maize breeding and their definition and validation in wheat represent a significant research theme in both the public and private sectors. In the meantime, much MAS use will be directed towards specific purposes such as accelerated selection of a few traits that are difficult to manage by conventional phenotyping, for the maintenance of recessive alleles in backcrossing programmes, for the pyramiding of disease resistance genes and for guiding the choice of parents to be used in crossing programmes.

REFERENCES

Anderson, J.A., Stack, RW., Liu, S., Waldron B.L., Fjeld, A.D., Coyne, C., Moreno-Sevilla, B., Mitchell Fetch, J., Song, Q.J., Cregan, P.B. & Frohberg, R.C. 2001. DNA markers for *Fusarium* head blight resistance QTLs in two wheat populations. *Theor. Appl. Genet.* 102: 1164–1168.

Ayala, L., Henry, M., Gonzalez-de-Leon, D., van Ginkel, M., Mujeeb-Kazi, A., Keller, B. & Khairallah, M.A. 2001a. Diagnostic molecular marker allowing the study of *Th. intermedium*-derived resistance to BYDV in bread wheat segregating populations. *Theor. Appl. Genet.* 102: 942–949.

Ayala, L., van Ginkel, M., Khairallah, M., Keller, B. & Henry, M. 2001b. Expression of *Thinopyrum intermedium*-derived barley yellow dwarf virus resistance in elite bread wheat backgrounds. *Phytopath.* 91: 55–62.

Buerstmayr, H., Lemmens, M., Hartl, L., Doldi, L., Steiner, B., Stierschneider, M. & Ruckenbauer, P. 2002. Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). *Theor. Appl. Genet*. 104: 84–91.

Buerstmayr, H., Steiner, B., Hartl, L., Griesser, M., Angerer, N., Lengauer, D., Miedaner, T., Schneider, B. & Lemmens, M. 2003. Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. *Theor. Appl. Genet*. 107: 503–508.

Chartrain, L., Brading, P.A. & Brown, J.K.M. 2004. Presence of the *Stb6* gene for resistance to *Septoria tritici* blotch (*Mycosphaerella graminicola*) in cultivars used in wheat-breeding programmes worldwide. *Plant Pathol*. 54: 134–143.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. & Weissenbach, J. 1996. A comprehensive genetic map of the human genome based on 5 264 microsatellites. *Nature* 380: 152–154.

Law, C.N. & Worland, A.J. 1997. The control of adult plant resistance to yellow rust by the translocated chromosome 5BS-7BS of bread wheat. *Plant Breed*. 116: 59–63.

Liu, W.H., Nie, H., Wang, S.B., Li, X., He, Z.T., Han, C.G., Wang, J.R., Chen, X.L., Li, L.H. & Yu, J.L. 2005. Mapping a resistance gene in wheat cultivar Yangfu 9311 to yellow mosaic virus, using microsatellite markers. *Theor. Appl. Genet.* 111: 651–657.

Mallard, S., Gaudet, D., Aldeia, A., Abelard, C., Besnard, A.L., Sourdille, P. & Dedryver, F. 2005. Genetic analysis of durable resistance to yellow rust in bread wheat. *Theor. Appl. Genet.* 110: 1401–1409.

Thomas, J., Fineberg, N., Penner, G., McCartney, C., Aung, T., Wise, I. & McCallum, B. 2005. Chromosome location and markers of *Sm1*: a gene of wheat that conditions antibiotic resistance to orange wheat blossom midge. *Mol. Breed.* 15: 183–192.

Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W. & Frohberg, R.C. 1999. RFLP mapping of QTLs for *Fusarium* head blight resistance in wheat. *Crop Sci.* 39: 805–811.

Young, N.D. 1999. A cautiously optimistic vision for marker-assisted selection. *Mol. Breed.* 5: 505–510.

Zhang, Z.Y., Xu, J.S., Xu, Q.J., Larkin, P. & Xin, Z.Y. 2004. Development of novel PCR markers linked to the BYDV resistance gene *Bdv2* useful in wheat for marker-assisted selection. *Theor. Appl. Genet.* 109: 433–439.

CHAPTER 5

Marker-assisted selection for improving quantitative traits of forage crops

Oene Dolstra, Christel Denneboom,
Ab L.F. de Vos and E.N. van Loo

SUMMARY

This chapter provides an example of using marker-assisted selection (MAS) for breeding perennial ryegrass (*Lolium perenne*), a pasture species. A mapping study had shown the presence of quantitative trait loci (QTL) for seven component traits of nitrogen use efficiency (NUE). The NUE-related QTL clustered in five chromosomal regions. These QTL were validated through divergent marker selection in an F_2 population. The criterion used for plant selection was a summation index based on the number of positive QTL alleles. The evaluation studies showed a strong indirect response of marker selection on NUE. Marker selection using a summation index such as applied here proved to be very effective for difficult and complex quantitative traits such as NUE. The strategy is easily applicable in outbreeding crops to raise the frequency of several desirable alleles simultaneously.

INTRODUCTION

Most agronomical characteristics of forage crops have a quantitative, polygenic and mostly complex nature. For these reasons, genetic improvement of such traits is laborious and time consuming. Improving nitrogen use efficiency (NUE) in perennial ryegrass (*Lolium perenne, 2n = 14*), the major grass species in northern Europe, is in this respect a good example. The high input of nitrogen needed to attain high forage yields for animal husbandry has caused severe water pollution (van Loo *et al.,* 2003), and therefore lowering nitrogen inputs through improving nitrogen use by breeding is of utmost importance.

Selection for NUE, however, is not easily implemented in conventional grass breeding based on field evaluations. Adequate testing requires separate and long-term trials with good control of the N stress, and such experiments tend to be rather inaccurate. To circumvent the disadvantages of field testing, a hydroponics system was used in this study in which the crop situation is simulated with growth-dependent N application (van Loo *et al.,* 1992), the aim being to grow plants having an equal suboptimal N content. The set-up has a capacity to test about 1 600 plants in parallel and enables all plants to experience more or less the same N strain. Criteria used to measure NUE are several plant growth characteristics, such as tillering, and shoot and root growth. Each test usually requires four to five cuts. The trait is vigour-related and complex, and is extremely important in relation to regrowth after cutting. Together, all these aspects make NUE a very attractive trait for MAS.

ANALYSIS OF GENETIC VARIATION

The genetic variation for NUE present in an F_1 plant originating from a cross between two contrasting genotypes for NUE was first analysed by crossing the F_1 with a doubled haploid. The resulting test cross progeny was then used to produce a molecular marker map and analyse the variation. This approach was chosen to avoid inbreeding effects and to be able to use dominant molecular markers. The performance of the mapping population for NUE-related traits was studied on hydroponics with the system set at a moderately low nitrogen deficiency (3.6 percent N of leaf dry weight). The outcome of the mapping study was a genetic map with seven linkage groups.

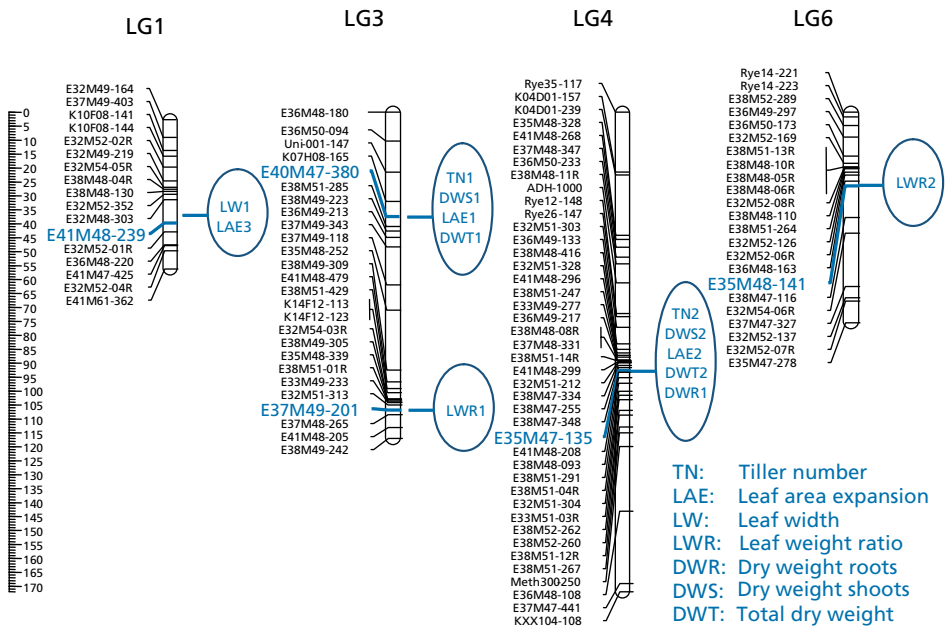
Putative genes (quantitative trait loci [QTL]) for the components of NUE were found on four linkage groups. The location of the selection markers for QTL is depicted in Figure 1. The map shows five genomic sites with 1-5 QTL. In total, 13 QTL for seven NUE related traits were found. Three sites contain more than one QTL.

The findings of the current study are typical for genetic analyses of quantitative traits in forage crops and also indicative of the problems associated with exploitation of QTL information through marker-assisted breeding. These included uncertainties with respect to effect and location of QTL, the fairly large number of QTL often found in genetic analyses, the co-segregation of QTL and the weighing of the different component traits of NUE and NUE-QTL. Below is a description of how these breeding problems were solved or circumvented in a divergent marker selection study to validate the QTL found in the mapping study.

DIVERGENT MARKER SELECTION

The plant materials used in the validation study were an F_2 generation obtained by selfing of the heterozygous F_1 genotype

FIGURE 1
Map position of the selection markers (in blue) for QTL for seven components of nitrogen use efficiency



used to generate the mapping population mentioned above (van Loo *et al.*, 2003). In total, about 200 genotypes were genotyped for five amplified fragment length polymorphism (AFLP) selection markers using the fluorescent AFLP technique developed by Applied BioSystems (Figure 1). The markers were co-dominantly scored using the heights of the fluorescence peaks relative to those of homozygous fragments as a criterion.

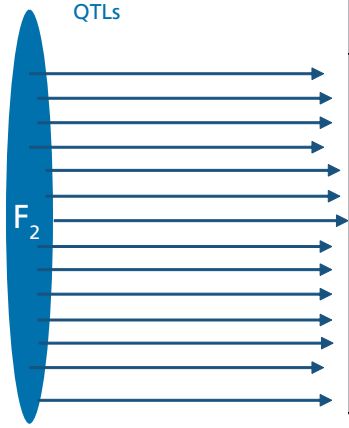
The genotyping data were used subsequently as a basis for a divergent mass selection programme. The selection strategy is outlined in Figure 2. The selection criterion was a genotype-specific selection index, being the summation of all positive QTL alleles (or chromosome segments) over the five QTL sites considered (Figures 1 and 2).

APPLICATION OF MARKER SELECTION

The AFLP technique is usually not the marker technology of choice for selection purposes because of its dominant nature and high costs per selection marker. However, co-dominant scoring of the five selection markers was quite adequate. The trimodal frequency distributions allowed proper classification of plants, although some misclassification cannot be fully excluded. The advantages of co-dominant AFLP scoring from a selection point of view are so large that a small number of genotyping errors are acceptable.

The decision to use a summation index as the criterion for selection was made primarily because of the difficulty of weighting the individual NUE related traits and the co-localization of QTL. The designation of the positive QTL alleles (chromosome

FIGURE 2
Strategy applied for marker selection



Genotype	Number of plus alleles					
	QTL1	QTL2	QLT3	QTL4	QTL5	Sum
1	2	1	0	0	2	5
2	1	1	0	0	2	4
3	0	2	2	1	1	6
4	2	2	2	1	1	8
5	2	2	1	1	1	7
.						
.						
.						
195	1	1	2	2	2	8
196	1	1	2	0	2	6
197	0	2	0	1	2	5
198	0	2	0	1	1	4
199	2	2	1	1	1	7
200	2	1	2	2	1	8

Selection criterion

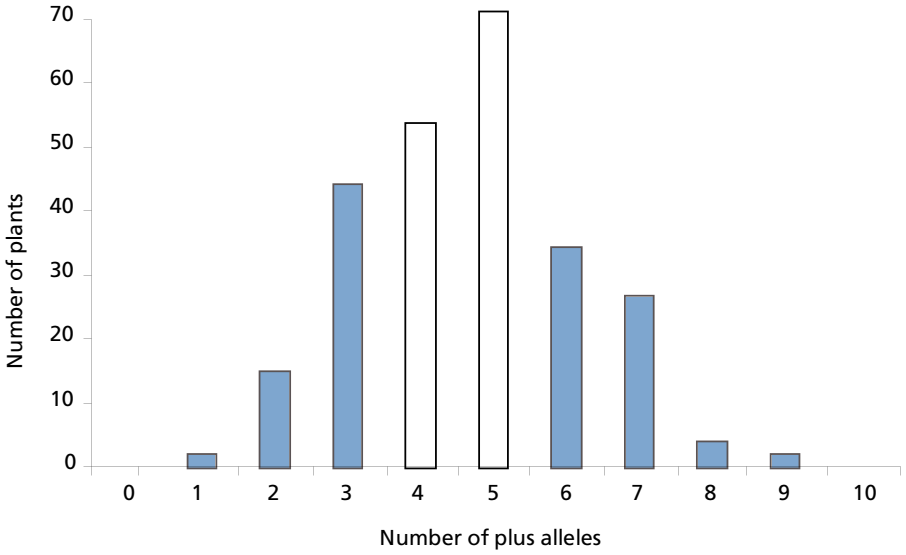
fragments) turned out to be straightforward. Figure 3 shows the F_2 frequency distribution for the number of "plus alleles". The population mean is somewhat below the expected number of five owing to the fact that the AFLP marker on LG1 showed a skewed segregation. This is likely due to gametophytic selection in favour of the negative QTL allele, perhaps due to linkage with an incompatibility locus.

The intensities of selection were set at about 25 percent, representing about 50 genotypes per selection (Figure 3). The selection pressure was kept fairly low because of the need to have sufficient seeds for measuring selection responses. In this way, the influence of genetic drift accompanying marker selection was minimized. The cut-off point for the top selection was six positive alleles and three for the opposite selection (Figure 3). The frequency of the plus alleles was on average 0.66 and 0.27, respectively. Selection showed a positive response for all NUE loci. However, the between-selection difference in allele frequency of the loci ranged from 0.18 to 0.77, showing that index selection did not affect all NUE loci to the same degree. The differences were probably mainly due to chance.

INDIRECT RESPONSE TO MARKER SELECTION

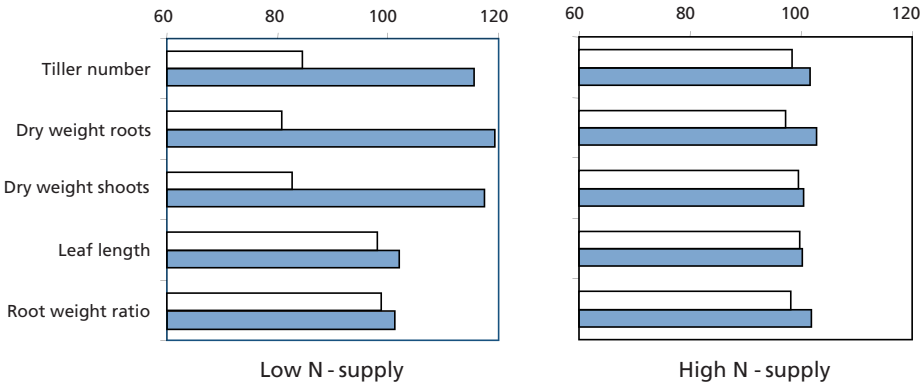
The selections were then multiplied using a polycross scheme (after vegetative propagation) to obtain sufficient seeds for evaluation on hydroponics and under various field conditions. The marker selections were evaluated for NUE in a replicated trial with two cuts on hydroponics at two N levels, being 2.5 and 5 percent N in leaves (van Loo *et al.*, 2003). The same set of plant characteristics as in the original mapping studies was monitored after each cut. Leaf area expansion rate, leaf length and width, as well as tiller number, were determined one week after cutting. The determination of shoot and root dry weight followed three weeks later. The indirect responses to marker selection are summarized in Figure 4. At low N supply, the NUEplus

FIGURE 3
Frequency distribution for number of plus alleles found for five QTLs



The two contrasting marker selections, i.e. NUEplus (right) and NUEmin (left), are marked in blue.

FIGURE 4
Performance of the marker selections, NUEmin (white) and NUEplus (blue) on hydroponics at low N-supply (2.5%N in leaves) and high N-supply (5%N in leaves). The performance is expressed as a percentage of the mean of the two selections



selection showed a remarkable 40 percent higher tillering rate and dry matter production than the NUEmin selection. The 40 percent higher tillering rate is associated with a 40 percent higher leaf area increase after defoliation (data not shown). Relative root growth (expressed as the ratio of root to total growth) and leaf length were hardly

changed through marker selection. At high N supply, the performances of NUEplus and NUEmin were fairly similar (Figure 4). The selections also showed striking differences in field trials in Germany, England and the Netherlands. At suboptimal N, the NUEplus selection significantly outperformed its counterpart in yields of dry matter and water soluble carbohydrates, while total N uptake was slightly lower.

CONCLUSIONS

Divergent mass selection has shown that marker selection using a summation index can be very effective for difficult and complex quantitative traits such as NUE. A collateral advantage of such an approach is that it offers a true validation of the putative genes (QTL) for the traits of interest. The associated response to marker selection distinctively indicates the presence of true genes affecting NUE, particularly in the vicinity of markers, which were strongly affected by the selection imposed. The results also indicate that recurrent mass selection to increase the number of positive alleles is worthwhile. The strategy is easily applicable in outbreeding crops.

ACKNOWLEDGEMENTS

The research work was carried out within the framework of the EU-FAIR project NIMGRASS (CT98-4063).

REFERENCES

van Loo, E.N., Schapendonk, A.H.C.M. & de Vos, A.L.F. 1992. Effects of nitrogen supply on tillering dynamics and regrowth of perennial ryegrass populations. *Netherlands J. Agric. Sci.* 40: 401–419.

van Loo, E.N., Dolstra, O., Humphreys, M.O., Wolters, L., Luessink, W., de Riek, W. & Bark N. 2003. Lower nitrogen losses through marker assisted selection for nitrogen use efficiency and feeding value (NIMGRASS). *Vorträge Pflanzenzüchtung* 59: 270–279.

CHAPTER 6

Targeted introgression of cotton fibre quality quantitative trait loci using molecular markers

Jean-Marc Lacape, Trung-Bieu Nguyen, Bernad Hau and Marc Giband

SUMMARY

Within the framework of a cotton breeding programme, molecular markers are used to improve the efficiency of the introgression of fibre quality traits of *Gossypium barbadense* into *G. hirsutum*. A saturated genetic map was developed based on genotyping data obtained from the BC_1 (75 plants) and BC_2 (200 plants) generations. Phenotypic measurements conducted over three generations (BC_1, BC_2 and BC_2S_1) allowed 80 quantitative trait loci (QTL) to be detected for fibre length, uniformity, strength, elongation, fineness and colour. Positive QTL, i.e. those for which favourable alleles came from the *G. barbadense* parent, were harboured by 19 QTL-rich regions on 15 "carrier" chromosomes. In subsequent generations (BC_3 and BC_4), markers framing the QTL-rich regions were used to select about 10 percent of over 400 plants analysed in each generation. Although BC plants selected through the marker-assisted selection (MAS) process show promising fibre quality, only their full field evaluation will allow validation of the procedure.

INTRODUCTION

Among the four species of *Gossypium* that produce seeds with spinnable fibres called cotton, *Gossypium hirsutum* dominates the world's cotton fibre production, accounting for approximately 90 percent of total world production. The second most cultivated species, *G. barbadense*, includes superior extra long, strong and fine cottons. However, compared with *G. hirsutum*, the marketing advantage of "high quality" *G. barbadense* cottons is offset by their lower productivity and a narrower adaptability to harsh environments. Breeding approaches within these two species have essentially relied on hybridization and selection methods (subsequent to simple or complex crosses, a pedigree system, sometimes combined with recurrent selection, is applied). Although *G. hirsutum* and *G. barbadense* display complementary characteristics, attempts to utilize deliberate interspecific *G. hirsutum*/*G. barbadense* recombinations through conventional breeding have had limited impact on cultivar development.

In the past 10–15 years, DNA markers for analyses of QTL and MAS have received considerable attention by plant and animal breeders (Dekkers and Hospital, 2002). However, following an initial keen interest and promises for molecular-based breeding approaches, the successful application of this technology has been shown to depend on the reliability and accuracy of the QTL analyses, which in turn are strongly affected by both population size and environmental factors (Schön *et al.,* 2004). Examples of applied MAS in breeding programmes are still scarce, particularly when complex traits (yield components, product quality) are under consideration.

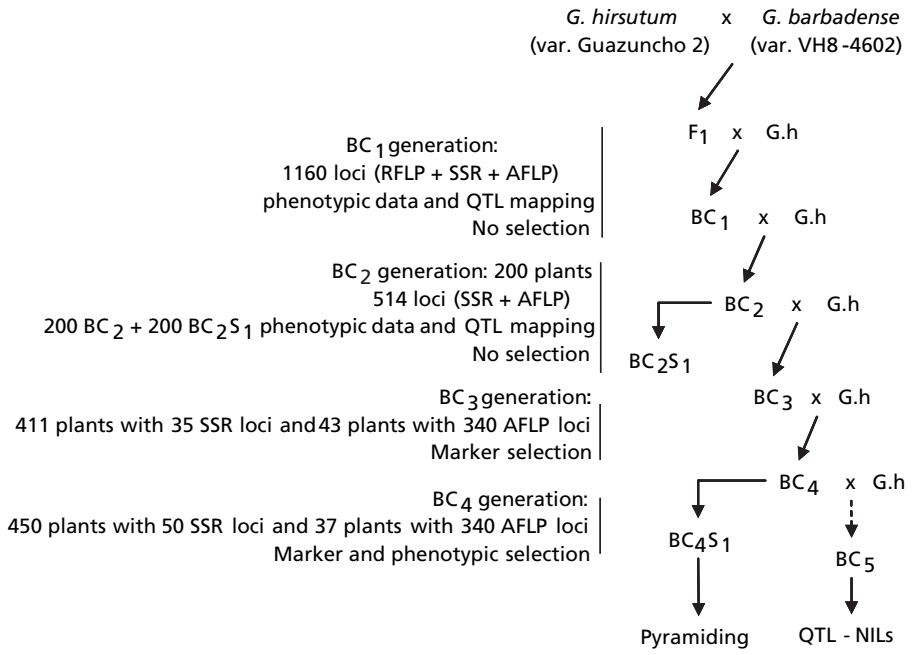
In the case of cotton, it is only recently that the results of efforts to gain a better understanding of the genome and the molecular basis of fibre quality have been published. Most of the earlier efforts in cotton molecular breeding concentrated on interspecific hybridization, due to the fact that, intraspecifically, the major species *G. hirsutum* displayed a very low level of molecular variability (Brubaker and Wendel, 2001). Based on studies of interspecific *G. hirsutum* x *G. barbadense* populations, published reports relate (i) to the construction of high-resolution genetic maps (Lacape *et al.,* 2003; Rong *et al.,* 2004); and (ii) to the identification of fibre quality-related QTL (Jiang *et al.*, 1998; Kohel *et al.*, 2001; Paterson *et al.,* 2003; Lacape *et al.,* 2005). In parallel, data have accumulated describing the cotton fibre transcriptome (reviewed by Wilkins and Arpat, 2005). These studies confirmed that key fibre quality properties, such as length, fineness and strength, are controlled quantitatively, thus complicating conventional breeding for fibre improvement.

Within the framework of a marker-assisted backcross introgression scheme aimed at transferring fibre quality traits from a low-productivity line of *G. barbadense* (donor) into a productive line of *G. hirsutum* (recipient), a saturated genetic map of tetraploid cotton was first developed (Lacape *et al.,* 2003). This chapter describes how molecular markers were used in the early BC_1 and BC_2 generations to identify QTL-rich regions involved in determining fibre quality, as recently reported by Lacape *et al.* (2005), and how MAS was actually implemented in the later BC_3 and BC_4 generations.

METHODOLOGY

The major milestones (Figure 1) in the marker-assisted backcross selection process included the construction of two genetic maps from the BC_1 and BC_2 populations, the detection of fibre quality QTL from

FIGURE 1
Schematic representation of the marker-assisted backcross selection scheme



three phenotyping data sets (BC_1, BC_2 and BC_2S_1) and the actual marker-based selection in the BC_3 and BC_4 generations, followed by the analysis of marker-trait associations in the BC_3 and BC_4 generations.

Plant material

The initial interspecific cross involved the *G. hirsutum* variety Guazuncho 2 and the *G. barbadense* variety VH8-4602. Guazuncho 2 is a modern pure line *G. hirsutum* variety created in Argentina and was chosen as a recipient in the backcross generations for its good overall agronomic performance. VH8-4602, a *G. barbadense* variety of the Sea Island type, was the donor parent for superior fibre quality, in particular for length (+9 to +12 mm as compared with Guazuncho 2), strength (+12 to +16 g/tex) and fineness (-30 to -50 millitex); conversely its fibre colour indices (reflectance and yellowness) are of lower value.[1]

The plant material used in the multi-generation QTL analyses included three populations: BC_1, BC_2 and BC_2S_1 (Lacape *et al.*, 2005). The first backcross generation (BC_1), consisted of 75 plants grown in a greenhouse in Montpellier (France) during the summer of 1999; these served as female parents for the second backcross to Guazuncho 2. Two hundred individual field-grown BC_2 plants that had shown a satisfactory production of BC_3 seeds and originating from 53 different BC_1 plants were used in 2000. Open pollinated seeds harvested from BC_2 plants were grown as 200 BC_2S_1 progenies in 2001 under field conditions in Brazil. Each BC_2S_1 line was

[1] 1 tex = 1 gram/kilometre

planted in two replications, each plot (one row) measuring 5 m. The next BC_3 and BC_4 generations were grown under field (411 BC_3 in 2002) or greenhouse (450 BC_4 in 2003) conditions in Montpellier. Every plant in each $BC_{1\text{-}4}$ generation was used for DNA extraction from young fresh leaves using different methods described elsewhere (Lacape *et al.,* 2003; Nguyen *et al.,* 2004). In each BC_3 and BC_4 generation, an early genotyping was conducted (before flowering of BC_3 plants and at the seedling stage for BC_4 plants), to reduce the number of plants to be manipulated and raised to flowering for selfing and backcrossing.

From each generation (75 BC_1, 200 BC_2, 400 BC_2S_1, 43 selected BC_3 and 37 selected BC_4), the cotton seed harvest was ginned (separation of the fibre from the seed) on a laboratory roller gin and the fibre was sampled for analyses at the Fibre Technology Laboratory of the French Agricultural Research Centre for International Development (CIRAD).

Fibre analyses

All fibre quality measurements (11 traits) were conducted at CIRAD, Montpellier, on a high volume instrument line (Zellweger Uster 900, Uster Technologies, Switzerland). These included length, uniformity, strength, elongation and colour. A FMT3 maturimeter (Shirley Dev Ltd., UK) was used to determine micronaire value, maturity and fineness.

Molecular analyses

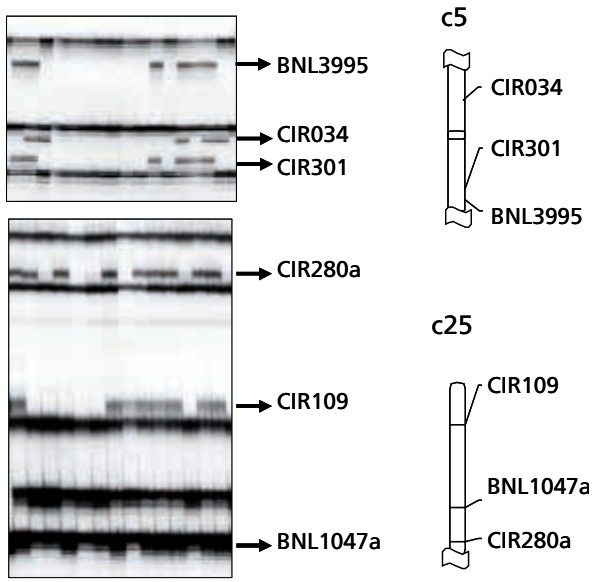
The different types of markers displaying polymorphism between *G. hirsutum* and *G. barbadense* included restriction fragment length polymorphisms (RFLPs) (used only in the BC_1 generation), simple sequence repeats (SSRs) and amplified fragment length polymorphisms (AFLPs). Details of the markers and protocols used are provided in Lacape *et al.* (2003) and Nguyen *et al.* (2004). The AFLP markers were all derived from combinations of EcoRI/MseI primer pairs (64 pairs in the BC_1, 45 in the BC_2 and 30 in the BC_3 and BC_4 generations). The cotton microsatellites were derived essentially from two public libraries, Brookhaven National Laboratory (BNL) and CIRAD (CIR). In the BC_1 generation, the microsatellites used included 188 polymorphic BNL markers out of the 216 available (Lacape *et al.,* 2003) and 204 CIR markers out of 392 developed (Nguyen *et al.,* 2004). From the results of the combined QTL analyses of the BC_1/BC_2/BC_2S_1 generations (Lacape *et al.,* 2005), QTL-rich regions were identified on "carrier" chromosomes, and SSR loci present within or in the vicinity of these regions were assembled for constituting groups of three SSRs (one group per region) to be tested as multiplexes, taking into account both annealing temperature and compatibility of sizes of amplified fragments. A subset of 60 SSR (20 region-specific triplexes) was used for early genotyping of all 411 BC_3 and 450 BC_4 plants (see examples in Figure 2). The individual plants selected from BC_3 and BC_4 (43 and 37 plants respectively) were further analysed using known AFLPs to provide broad genome coverage. In the context of our marker-assisted introgression programme, the SSR markers target the QTL-rich regions, i.e. those loci of the "foreground genome" expected to have been introgressed, while the AFLP markers essentially serve to cover the rest of the genome, i.e. the "background genome", aimed at returning to the recipient genome composition.

Construction of genetic map

The BC_1 (75 individuals) and BC_2 (200 individuals) maps were constructed separately using the MapMaker 3.0 software

FIGURE 2

Examples of autoradiograms showing the segregation of 2 SSR triplexes observed among BC_4 plants and used for the targeted genotyping of specific regions along chromosomes c5 and c25 (a subset of around 15 BC_4 plants from the 450 plants analysed is shown)



(Lander *et al.*, 1987). The MapMaker "group" (using a logarithm of the odds ratio [LOD] of 5.0 and 30 as a maximal recombination frequency), "order" and "sequence" commands were used in each case. After aligning the BC_1 and BC_2 maps using common loci, a consensus framework BC_1/BC_2 map was constructed by simple extrapolation of the positions of the additional BC_2 loci on the BC_1 map used as a backbone map. The allelic constitution throughout the 26 chromosomes of all $BC_{1\text{-}4}$ individuals was displayed graphically using Graphical GenoTyping software (R. van Berloo, Laboratory of Plant Breeding, Wageningen, Netherlands) and represented along the consensus BC_1 map data.

QTL analyses

The combined marker and phenotypic data then served for three (BC_1, BC_2 and BC_2S_1) separate QTL analyses of fibre quality components. The association between phenotype and marker genotype was investigated through simple marker analysis (SMA), interval mapping (IM) and composite interval mapping (CIM) using the computer software QTL Cartographer 1.13 (Basten, Weir and Beng, 1999) as described in Lacape *et al.* (2005). In each data set (trait, generation), permutation-based thresholds were considered at a 5 percent risk at the genome level. Interval methods relied on the positions of the loci on the consensus BC_1 map. Molecular data of further generations (BC_3 and BC_4) were also combined with phenotypic measurements for conducting the SMA option of QTL Cartographer. Cotton fibre properties were considered from a product transformation perspective, meaning that decreases the in fibre fineness and yellow-

TABLE 1
Range of parental (*G. hirsutum [Gh]* and *G. barbadense [Gb]*) values over the five sets of data

	Gh	*Gb*	BC_1 N=75	BC_2 N=200	BC_2S_1 N=200	BC_3 N=43	BC_4 N=37
Length (mm)*	27.5–31.8	39.2–43.7	33.8 (27.8–38.2)	28.6 (22.9–35.3)	30.7 (26.9–36.1)	30.0 (25.0–33.9)	31.7 (28.1–37.1)
Length uniformity	81.3–85.5	83.9–87.1	85.0 (82.0–88.2)	81.3 (73.3–86.7)	83.3 (80.6–85.1)	81.9 (77.6–86.4)	86.1 (82.5–89.5)
Strength (g/tex)	26.5–32.5	41.4–46.7	35.7 (29.7–41.6)	28.3 (17.8–43.7)	29.0 (23.7–34.5)	24.5 (16.8–32.8)	33.8 (29.7–39.1)
Elongation	5.1–6.4	5.5–6.0	6.3 (5.7–7.4)	5.5 (3.9–7.6)	6.3 (5.4–7.4)	5.7 (4.5–7.0)	6.3 (4.9–7.4)
Fineness (mtex)**	207–243	178–191	218 (177–308)	224 (165–379)	225 (176–285)	128*** (117–148)	243 (192–283)
Colour reflectance	71.2–77.7	74.6–75.6	74.3 (65.9–81.1)	72.2 (56.8–81.3)	74.1 (69.8–77.5)	71.5 (64.7–76.6)	75.1 (67.1–82.0)

* Length is upper half mean length (UHML), ** standard fineness, *** low fibre fineness values in BC_3 generation because of poor maturities
Note: Mean values and range (in brackets) observed in each BC_{1-4} generation (number of plants, N, indicated) of fibre technological parameters.

ness index, for example, were positively considered.

Details of the plant material used and the types of analyses undertaken during the different steps of the MAS process are given in Figure 1.

RESULTS

Phenotypic variation

The two parents were characterized by their contrasting fibre properties (Table 1) with significant advantages for the *G. barbadense* parent in terms of length (+9.7 mm on average over all data sets), strength (+15.9 g/tex) and fineness (-38 mtex). By contrast, the *G. hirsutum* parent displayed better yellowness index/colour reflectance. For each BC population, it was observed that the data fitted normal distributions, that transgressive segregants were regularly in the lower range of phenotypic values and that, although progeny values rarely reached those of *G. barbadense*, high phenotypic values were observed, including within the most advanced BC_4 generation (Table 1).

Genetic mapping

The first step in the programme involved the construction of two genetic maps of tetraploid cotton by combining RFLP, SSR and AFLP markers generated separately from the first two backcross generations (BC_1 and BC_2). The initial BC_1 map comprising 888 loci grouped in 37 linkage groups and spanning 4 400 cM (Lacape *et al.*, 2003), benefited from the development and integration of new additional microsatellite markers (Nguyen *et al.*, 2004). This updated saturated BC1 map spans 5 500 cM and comprises a total of 1 160 loci ordered along 26 chromosomes or linkage groups (Nguyen *et al.*, 2004). On the other hand, the BC_2 map constructed using AFLP and SSR markers had 514 loci in total. The two maps agreed perfectly for loci order. They had 373 loci in common (between seven and 26 per chromosome throughout the 26 chromosomes), thus allowing their merger into a combined consensus map. The consensus framework map comprises 1 306 loci and spans 5 597 cM, with an average marker interval of 4.3 cM.

QTL detection

The QTL analyses, conducted through composite interval mapping, used two molecular data sets (BC_1 and BC_2) and three sets of fibre measurements (per plant

TABLE 2
Number of QTL for each trait and range of observed phenotypic effects conferred by the *G. barbadense* alleles (either positive, "*Gb* +", or negative, "*Gb* –") detected over the three populations (BC_1, BC_2 and BC_2S_1)

	QTL *Gb* +	Range phenotypic effects	QTL *Gb* –	Range phenotypic effects
Length (mm)*	12	+0.7 to +2.1	3	–1.6 to –1.8
Length uniformity	3	+0.5 to +1.5	3	–1.1 to –3.3
Strength (g/tex)	8	+0.8 to +2.8	4	–0.9 to –3.4
Elongation	6	+0.2 to +0.5	4	–0.3 to –0.6
Fineness (mtex)**	13	–10 to –20	8	+9 to +40
Colour reflectance	3	+1.8 to +2.5	13	–0.9 to –3.5
Total	**45**		**35**	

* Length is upper half mean length (UHML), ** standard fineness.

basis for BC_1 and BC_2 and per-line basis with two replicates for BC_2S_1). The generations BC_1 and BC_2 were conducted with no selection, except for choosing those plants that produced backcrossed seeds. The fibre measurements, which initially included eleven traits, were reduced to six groups after considering the strong correlations that existed between some traits. The fibre characteristics that were retained for measurement included length, length uniformity, strength, elongation, fineness or maturity, and colour.

For the six fibre quality components studied, 50 QTL were identified that met permutation-based LOD thresholds (ranging between 3.2 and 4.0 for most of the traits). Thirty additional suggestive QTL (having a LOD value below threshold but above 2.5) were also taken into consideration after comparing the results between the three populations or between the present results and those reported in the literature (Jiang *et al*., 1998; Kohel *et al.*, 2001; Paterson *et al.,* 2003; Mei *et al*., 2004). Table 2 summarizes the data generated from the QTL analyses for the six traits of interest and the phenotypic effects of the detected QTL. In general, the contribution of each QTL, measured as a percentage of explained variation of a given trait, was variable and in most cases fairly low. For example, for traits of economic importance, individual contributions varied from 4.8 to 14.8 percent in the case of fibre length, 4.4 to 21.3 percent for fibre strength and 4.6 to 29.1 percent for colour reflectance.

Overall, it was observed that these 80 QTL partitioned as expected from the phenotypic values of the *G. hirsutum* and *G. barbadense* parents: a majority of positive alleles for length (12 of the 15 QTL), strength (8 of the 12 QTL) and fineness (13 of the 21 QTL) derived from the *G. barbadense* parent, while a majority of positive alleles for fibre colour (13 of the 16 QTL) derived from the *G. hirsutum* parent (Table 2). Furthermore, the QTL detected for the various traits often co-localized within QTL-rich regions (Lacape *et al.,* 2005). In some cases, QTL detection and mapping were in agreement between generations (BC_1 and BC_2) and, very interestingly, in 26 cases (33 percent of the 80 QTL) they confirmed the results reported in the literature, both for the position of a QTL and for the sign of its phenotypic effect. The most prominent cases of QTL consistently detected in this study as well as in those of Paterson *et al*. (2003) and Kohel *et al.* (2001), i.e. in different crosses/populations, were found

TABLE 3
Identification of the 19 targeted regions mapped on 15 different chromosomes and comprising one or several co–localized fibre quality QTL from *G. barbadense* for introgression into a *G. hirsutum* genetic background

Carrier chromosome	Chromosome length (cM)	Target interval (cM)	Target size (cM)	Trait
c14	197	28–57	29	Length
c3	153	32–67	35	Length, fineness
		90–138	48	Length, strength, fineness
c4	190	102–118	16	Fineness
c22	139	112–139	27	Fineness
c5	360	78–101	23	Strength
c6	296	137–144	7	Length, fineness
c25	183	44–73	29	Length, strength
c16	168	65–117	52	Strength, fineness, colour
c23	173	45–66	21	Strength (elongation –, colour –)
		113–135	22	Length, strength
c10	192	0–21	21	Fineness
		78–120	42	Length, fineness, colour
c20	268	88–161	73	Elongation, fineness
c26	195	67–143	76	Length (colour –)
A01	233	16–54	38	Length
		171–209	38	Strength
c18	158	32–46	14	Fineness
A03	271	209–234	25	Strength, uniformity
Total	**3176**	**Total**	**636**	

Note: All targeted QTL show a positive contribution from the ***G. barbadense*** allele, except for a few negative cases indicated in brackets. The target region is defined as situated between the two loci flanking the QTL peak LOD value at a one LOD confidence interval.

along chromosome 3 for QTL for fibre strength and fineness, and chromosome 23 for QTL for fibre strength and length.

The chromosome regions carrying co-localized QTL (corresponding to a single or to several traits measured on a single or on several populations) whose positive alleles derived from the *G. barbadense* donor genome, were reduced to 19 QTL-rich regions that were carried by 15 different "carrier" chromosomes (Table 3). Altogether, the confidence intervals (one LOD) of the involved QTL-rich regions delimited a total length of 636 cM (20 percent of the carrier genome), or 11.5 percent of the total genome (Table 3). Eleven non-carrier chromosomes were devoid of positive QTL, or harboured negative (positive alleles derived from the *G. hirsutum* alleles) QTL.

MAS in the BC_3 and BC_4 generations and allelic transmission throughout generations

The early selection of BC_3 and BC_4 plants using SSR markers that framed the 19 targeted regions of interest made it possible to choose those plants that showed an allelic constitution with as many introgressed loci within the targeted regions as possible. In total, 43 BC_3 plants out of 411 (11.4 percent) and 37 BC_4 plants out of 450 (8.2 percent) were retained based upon the information provided by the markers, i.e. without any phenotypic selection at this stage. These plants were backcrossed to the recurrent parent (and self-pollinated in the case of the BC_4 plants).

The allelic transmission observed in the four groups of BC_4 derived from four different BC_1 plants is given in Table 4.

TABLE 4
Percentage of introgressed loci (I%), at the heterozygous state, of 37 BC_4 plants that were derived from four different BC_1 plants (Nos. 3, 11, 16, and 27)

BC_1			BC_2			BC_3			BC_4			
Plant number	Number of loci	I%	Number of plants	Number of loci	I%	Number of plants	Number of loci	I%	Number of plants	Number of loci	I% global	I% target/non-target
No. 3	646	55	1	479	14	1	467	8	9	456	5	10/4
No. 11	654	64	1	446	28	1	403	13	1	408	10	29/6
No. 16	681	67	2	464	31	3	420	15	21	428	9	25/6
No. 27	668	63	1	471	26	1	433	15	6	439	10	25/6
Mean		62			26			14			8	21/5

Note: The number of plants and of loci analysed at each generation are given. At the BC_4 generation, the percentage of introgression is also differentiated between target and non-target (as defined in Table 3) regions.

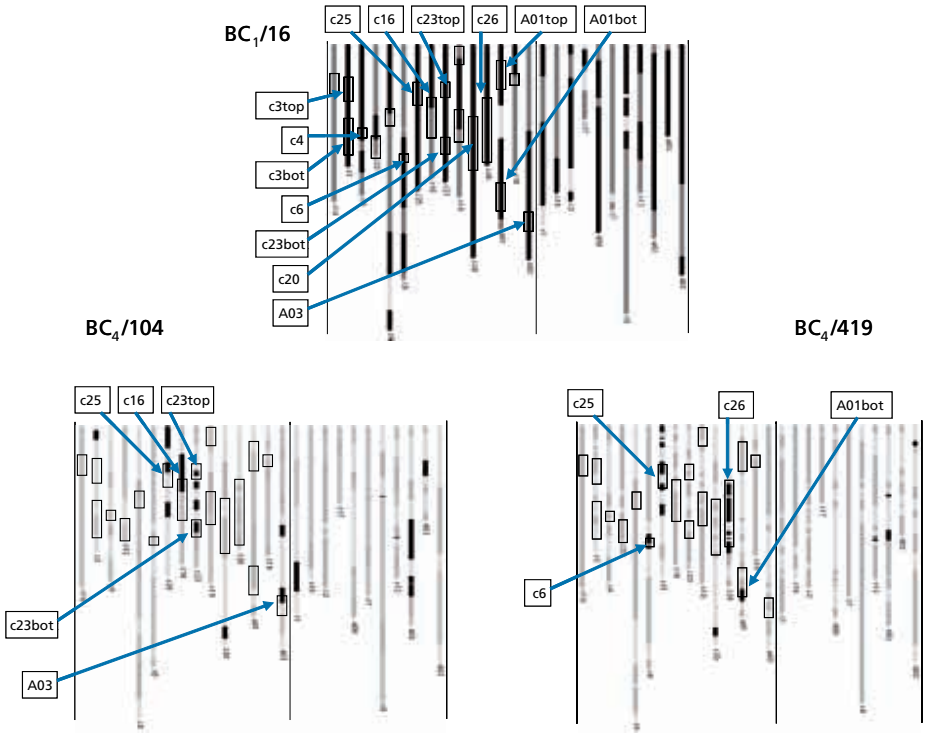
Moderate deviations were observed from theoretical transmission values (62, 26, 14 and 8 percent compared with 50, 25, 12.5 and 6.25 percent at the BC_1, BC_2, BC_3 and BC_4 stages, respectively), with a bias in favour of a higher rate of *G. barbadense* allele transmission. This bias was probably due to the selection pressure imposed at least in the BC_3 and BC_4 generations. Throughout the BC_1 and BC_2 generations that have undergone no deliberate selection, the introgression of *G. barbadense* alleles (at the heterozygous state) covered the complete genome fairly well, i.e. introgressed segments were found on all of the 26 chromosomes (not shown). This result contradicts the findings of Jiang *et al.* (2000) who detected important deficiencies in donor (*G. barbadense*) allele transmission in a population of 3 662 BC_3 plants originating from 21 BC_1 plants.

After combining the SSR and AFLP marker data, it was observed that the introgression rate differed between target and non-target regions. When averaged over the 37 BC_4 plants, the percentage of introgressed loci (8 percent genome-wide) was much lower in the non-target regions (5 percent) than that reached within target regions (21 percent) (Table 4). The different BC_4 plants introgressed between three and six QTL-rich target regions in different combinations. As an illustration of the selection pressure applied through the use of molecular markers, Figure 3 shows the graphical genotype of two BC_4 individuals as well as that of the BC_1 plant (No. 16) from which these individuals were derived. The two BC_4 plants had a common BC_1 ancestor but originated from two different BC_2 plants. In this particular example, starting from a common BC_1 plant (No. 16) which harboured 13 out of 19 possible QTL-rich regions, the two BC_4 plants (Nos. 104 and 419) derived from it partly or completely retained respectively five (c16, c23top, c23bot, c25 and A03) and four (c6, c25, c26 and A01bot) genomic regions carrying favourable alleles. The other regions carrying QTL on c3, c4, c23, c20, A01 and A03, which had been introgressed and were heterozygous in the BC_1 plant, had returned to the homozygous *G. hirsutum*/*G. hirsutum* state. The percentages of introgressed loci in target and non-target regions in these two examples were 29 and 10 percent, and of 29 and 5 percent in the two BC_4 plants (Nos. 104 and 419) respectively.

This example shows that, at least in some cases, the process used was efficient in selecting for chromosomal regions of interest (foreground selection), while allowing the rest of the genome to return towards that of the recurrent parent.

FIGURE 3

Graphical genotypes (26 chromosomes) of a BC_1 plant (No. BC1/16) (upper panel) and of two selected BC_4 plant (Nos. BC_4/104 and BC_4/419) (lower panel) derived from it



The two possible allelic forms, homozygous *Gh*/*Gh* and heterozygous *Gh*/*Gb,* are denoted in dark grey and black respectively. Regions in black are introgressed with *G. barbadense* alleles. Light grey areas indicate portions of unknown allelic composition. Boxed areas represent the localization of QTL-rich regions localized on 15 carrier chromosomes shown to the left (11 non-carrier chromosomes are shown to the right). Arrows indicate the regions totally or partially introgressed.

Fibre characteristics of BC_3 and BC_4 generation plants

Owing to the limited number of individuals and the unbalanced frequencies of genotypic classes in the BC_3 and BC_4 material, significant marker-trait associations were less frequent than observed from the BC_1 and BC_2 data. For example, markers mapped along five, nine and six chromosome regions contributed (P=0.01), respectively, to length, strength or fineness variation using BC_4 marker-trait data, as compared with 15, 12 and 21 from the BC_1 and BC_2 data (Table 2). However, the majority of significant associations, particularly those determined in the BC_4 generation, were observed within previously detected regions (not shown). Using fibre strength as an example, out of the eight strength QTL-harbouring regions on chromosomes c3bot, c5, c16, c23sup, c23bot, c25, A01 and A03 identified from the combined BC_1 and BC_2 data (Table 3), the BC_4 data confirmed significant marker-trait associations in five of these regions, i.e. for markers mapped on chromosomes c3bot,

c16, c23bot, c25 and A03. Furthermore, it is worth noting that the BC_4 plant No. 104 presented in Figure 3, which had introgressed all these five regions, also displayed the highest fibre strength value of its generation (39.1 g/tex, compared with 33.1 g/tex for the Guazuncho 2 parent). The concomitant introgression of *G. barbadense* alleles displaying positive marker-trait associations for other fibre properties such as length or fineness was also observed. This translated into the development of different highly valuable BC progenies. These preliminary results suggest that the improvement of *G. hirsutum* fibre properties through the introgression of *G. barbadense* fibre QTL appears feasible.

DISCUSSION

In an attempt to overcome the limitations of conventional breeding for improving cotton fibre quality through the use of interspecific hybridization, molecular markers were used in a MAS scheme to improve the efficiency of introgressing fibre quality traits. The advanced backcross-QTL (AB-QTL) strategy (Tanksley and Nelson, 1996) was used as this allowed concomitant development of a genetic map of the cotton genome and analysis of fibre quality QTL, and attempts to introgress favourable alleles in an adequate recipient genetic background (Figure 1).

In contrast to monogenic characteristics such as disease and insect resistance, many important traits including yield and quality show continuous phenotypic variation and are governed by a number of QTL. Cotton fibre quality is a complex concept that involves a number of traits or characteristics. Each of these is under the influence of numerous QTL, indicating a complex genetic determinism. Indeed, from the present results, at least six QTL govern fibre uniformity and up to 21 QTL influence fibre fineness. When considering six traits that can account for fibre quality, a total of 80 QTL were detected (Table 2). This figure falls within the same range as that found by Paterson *et al.* (2003). As some of these QTL co-localized within the same chromosome region, by choosing those QTL whose positive allele derived from the donor parent and had the strongest effect on economically important fibre characteristics, the number of target regions to be introgressed was reduced to 19 (Table 3). Nevertheless, this number of QTL remains too high to identify a single plant that would carry them all. Indeed, in the authors' experience, at the BC_3 stage, single plants carried a maximum of five regions of interest (eight if considering regions only partially introgressed), while at the BC_4 stage, this number was reduced to four (seven if considering regions only partially introgressed).

At this stage of the MAS process, two routes are under way (Figure 1). The first involves identifying the best BC_4 plants, i.e. those showing the highest amount of favourable QTL introgression, and then fixing the favourable allele by self-pollination. Such BC_4S_1 plants have been crossed with other BC_4S_1 plants of different ascent in order to pyramid as many QTL as possible (each contributing to different traits) within the same genome. Similarly, BC_4S_1 plants were used to pyramid various QTL responsible for a given trait ("selective pyramiding"). This latter strategy could especially apply to traits of commercial importance, such as fibre strength or fineness. The second avenue involves repeating the backcrossing process until near isogenic lines differing only at a given QTL (QTL-NILs) are developed. Such plant material could prove useful not only to study the

effect of a single given QTL on the phenotypic value of a plant harbouring it, but also in case the introgressed QTL is proven to contribute significantly to the improvement of a given trait (Bernacchi *et al.*, 1998). Also, QTL-NILs could be used as donor material for QTL pyramiding (Peleman and van der Voort, 2003). Finally, an introgression library, i.e. a collection of NILs, will typically serve as primary plant material for QTL fine mapping and eventual QTL cloning (Salvi and Tuberosa, 2005).

However successful marker-aided introgression of genomic regions of interest may be, only phenotypic analysis of plant material stemming from the MAS process, including the assessment of its adaptability to any given set of local agronomic and ecological conditions, will allow validation of this procedure.

REFERENCES

Basten, C., Weir, B. & Beng, Z-B. 1999. *QTL cartographer, version 1.13.* Dept. of Statistics, North Carolina State Univ., Raleigh, NC, USA.

Bernacchi, D., Beck-Bunn, T., Emmatty, D., Eshed, Y., Inai, S., Petiard, V., Sayama, H., Uhlig, J., Zamir, D. & Tanksley, S.D. 1998. Advanced backcross QTL analysis in tomato. II. Evaluation of near isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derives from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor. Appl. Genet.* 97: 170–180.

Brubaker, C.L. & Wendel, J.F. 2001. RFLP diversity in cotton. *In* J.N. Jenkins & S. Saha, eds. *Genetic improvement of cotton, emerging technologies,* pp 81-101. Enfield, NH, USA, Science Publisher Inc.

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Genet.* 3: 22–32.

Jiang, C., Chee, P., Draye, X., Morrell, P., Smith, C.W. & Paterson, A.H. 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution* 54: 798–814.

Jiang, C., Wright, R., El-Zik, K.M. & Paterson, A.H. 1998. Polyploid formation created unique venues for response to selection in *Gossypium* (cotton). *Proc. Nat. Acad. Sc. USA* 95: 4419–4424.

Kohel, R.J., Yu, J., Park, Y.-H. & Lazo, G. 2001. Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica* 121: 163–172

Lacape, J.-M., Nguyen, T.-B., Thibivilliers, S., Courtois, B., Bojinov, B.M., Cantrell, R.G., Burr, B. & Hau, B. 2003. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome* 46: 612–626.

Lacape, J.-M., Nguyen, T.-B., Courtois, B., Belot, J.-L., Giband, M., Gourlot, J.-P., Gawryziak, G., Roques, S. & Hau, B. 2005. QTL analysis of cotton fiber quality using multiple *G. hirsutum* x *G. barbadense* backcross generations. *Crop Sci.* 45: 123–140.

Lander, E., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S. & Newburg, L. 1987. MapMaker: an interactive computer package for constructing primary linkage maps of experimental and natural populations. *Genomics* 1:174–181.

Mei, M., Syed, N.H., Gao, W., Thaxton, P.M., Smith, C.W., Stelly, D.M. & Chen, Z.J. 2004. Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *Theor. Appl. Genet.* 108: 280–291.

Nguyen, T.-B., Giband, M., Brottier, P., Risterucci, A.-M. & Lacape, J.-M. 2004. Wide coverage of tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* 109: 167–175.

Paterson, A.H., Saranga, Y., Menz, M., Jiang, C. & Wright, R.J. 2003. QTL analysis of genotype x environment interactions affecting cotton fiber quality. *Theor. Appl. Genet.* 106: 384–396.

Peleman, J.D. & van der Voort, J.R. 2003. Breeding by design. *Trends Plant Sci.* 7: 330–334.

Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C.-H., Pierce, G.J., Rainey, K.M., Rastogi, V.K., Schulze, S.R., Trolinder, N., Wendel, J.F., Wilkins, T.A., Williams-Coplin, T.D., Wing, R.A., Wright, R.J., Zhao, X., Zhu, L. & Paterson, A.H. 2004. A 3347 locus genetic recombination map of sequence tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166: 389–417.

Salvi, S. & Tuberosa, R. 2005. To clone or not to clone plant QTLs: present and future challenges. *Trends in Plant Sci.* 10: 297–304.

Schön, C., Utz, H.F., Groh, S., Truberg, B., Openshaw, S. & Melchinger, A.E. 2004. Quantitative trait locus mapping based on resampling in a vast maize test cross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167: 485–498.

Tanksley, S.D. & Nelson, J.C. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92: 191–203.

Wilkins, T.A. & Arpat, A. 2005. The cotton fiber transcriptome. *Physiologia plantarum* 124: 295–300.

CHAPTER 7

Marker-assisted selection in common beans and cassava

Mathew W. Blair, Martin A. Fregene, Steve E. Beebe and Hernán Ceballos

SUMMARY

Marker-assisted selection (MAS) in common beans (*Phaseolus vulgaris* L.) and cassava (*Manihot esculenta*) is reviewed in relation to the breeding system of each crop and the breeding goals of International Agricultural Research Centres (IARCs) and National Agricultural Research Systems (NARS). The importance of each crop is highlighted and examples of successful use of molecular markers within selection cycles and breeding programmes are given for each. For common beans, examples are given of gene tagging for several traits that are important for bean breeding for tropical environments and aspects considered that contribute to successful application of MAS. Simple traits that are tagged with easy-to-use markers are discussed first as they were the first traits prioritized for breeding at the International Center for Tropical Agriculture (CIAT) and with NARS partners in Central America, Colombia and eastern Africa. The specific genes for MAS selection were the *bgm*-1 gene for bean golden yellow mosaic virus (BGYMV) resistance and the *bc*-3 gene for bean common mosaic virus (BCMV) resistance. MAS was efficient for reducing breeding costs under both circumstances as land and labour savings resulted from eliminating susceptible individuals. The use of markers for other simply inherited traits in marker-assisted backcrossing and introgression across Andean and Mesoamerican gene pools is suggested. The possibility of using MAS for quantitative traits such as low soil phosphorus adaptation is also discussed as are the advantages and disadvantages of MAS in a breeding programme. For cassava, the use of multiple flanking markers for selection of a dominant gene, *CMD2* for cassava mosaic virus (CMV) resistance at CIAT and the International Institute of Tropical Agriculture (IITA) as well as with NARS partners in the United Republic of Tanzania using a participatory plant breeding scheme are reviewed. MAS for the same gene is important during introgression of cassava green mite (CGM) and cassava brown streak (CBS) resistance from a wild relative, *M. esculenta* sub spp. *flabellifolia*. The use of advanced backcrossing with additional wild relatives is proposed as a way to discover genes for high protein content, waxy starch, delayed post-harvest physiological deterioration, and resistance to whiteflies and hornworm. Other potential targets of MAS such as beta carotene and dry matter content as well as lower cyanogenic potential are given. In addition, suggestions are made for the use of molecular markers to estimate average heterozygosity during inbreeding of cassava and for the delineation of heterotic groups within the species. A final section describes the similarities and differences between the MAS schemes presented for the two crops. Differences between the species can be ascribed partially to the breeding and propagation systems of common beans (seed propagated, self-pollinating) and cassava (clonally propagated, cross-pollinating). In addition, differences in growth cycles, breeding methods, availability of genetic markers, access to selection environments and the accompanying opportunities for phenotypic selection influence the decisions in both crops of when and how to apply MAS. Recommendations are made for applying MAS in breeding of both crops including careful prioritization of traits, marker systems, genetic stocks, scaling up, planning of crosses and the balance between MAS and phenotypic selection.

COMMON BEANS: IMPORTANCE AND GENETICS

Common beans (*Phaseolus vulgaris* L.) are the most important grain legume for direct human consumption, especially in Latin America and eastern and southern Africa. They are seed-propagated, true diploids (2n = 22) and have a relatively small genome (650 Mb) (Broughton *et al.*, 2003). Originating in the Neotropics, common beans were domesticated in at least two major centres in Mesoamerica and the Andes (Gepts, 1988) and possibly in a third minor centre in the northern Andes (Islam *et al.*, 2002). Wide DNA polymorphism is expressed between the two major gene pools. Mesoamerican beans typically have small to medium size seeds and can be classed into four races that are distinguished by randomly amplified polymorphic DNA (RAPD) polymorphisms (Beebe *et al.*, 2000). Andean beans usually have medium to large seeds, and landraces have been classed into three races based on plant morphology and agro-ecological adaptation (Singh, Gepts and Debouck, 1991). These can be differentiated by microsatellites (M. Blair, unpublished data) but the genetic distance among Andean races is narrower than that among Mesoamerican races (Beebe *et al.*, 2001). A large number of gene tagging studies have been conducted in common beans, predominantly with RAPD markers, some of which have been converted subsequently to sequence characterized amplified regions (SCARs; reviewed most recently by Miklas *et al.*, 2006).

Beans display a wide range of growth habits (Van Schoonhoven and Pastor-Corrales, 1987), from determinate bush types, to indeterminate upright or viny bush types, to vigorous climbers. Bush types are the most widely grown, and are a relatively short season crop, maturing in as little as 60 days from seeding in a tropical climate and yielding from 700 to 2 000 kg/ha on average. On the other hand, in smallholder agriculture where land is scarce, labour-intensive, high-yielding climbing beans enjoy continuing or even expanding popularity. Climbing beans can mature in 100 to 120 days at mid-elevations, but can delay as long as ten months at higher elevations and can produce the highest yields for the crop, up to 5 000 kg/ha. These features have significant implications for breeding programmes. In bush types it is possible to obtain up to three cycles per year in the field, or even four cycles in greenhouse conditions. Breeding bush beans is thus quite agile with regard to advance of generations, although seed harvest of individual plants is sometimes limited. With climbing beans, on the other hand, at best it is possible to obtain two cycles per year with field grown plants, while managing climbing beans in the greenhouse is logistically difficult. However, while bush beans produce on average 20 to 50 seeds/plant, individual plants of climbing beans often produce enough seeds to plant several rows (100 to 150 seeds).

Beans are self-pollinating and thus breeding methods for autogamous crops are employed. Pedigree selection or some adaptation thereof is most common, and both recurrent (Muñoz *et al.*, 2004) and advanced (or inbred) backcrossing (Sullivan and Bliss, 1983; Buendia *et al.*, 2003; Blair, Iriarte and Beebe, 2003) have been used. Recurrent selection has also been employed (Kelly and Adams, 1987; Beaver *et al.*, 2003) but seldom in a formal sense with a defined population structure. Singh *et al.* (1998) suggested a system that they called gamete selection in which individual F_1 plants of multiple parent crosses give rise

to families. This system takes advantage of the variability among F_1 plants that is created between segregating parental plants. The choice of breeding method and its adaptation to specific circumstances, the growth cycle of the crop in relation to different planting seasons, the access to selection environments and the accompanying opportunities for phenotypic selection and the ease of implementing the specific markers to be used will all influence the decisions about where and how MAS will be most cost effective and used to best advantage.

MAS in bean breeding: experiences of CIAT and NARS

Molecular markers have been sought for both simple and complex traits in beans, with an eye to eventual application in MAS. Tagging of genes and QTL in common bean and their application to MAS have been reviewed previously (Kelly *et al.,* 2003; Miklas *et al.,* 2006). In the present chapter, some of the aspects that contribute to the successful use of MAS are considered in greater detail, referring to examples taken from bean breeding in the tropics at CIAT and within NARS. Simple and complex traits are discussed separately, as they represent two contrasting sorts of experience.

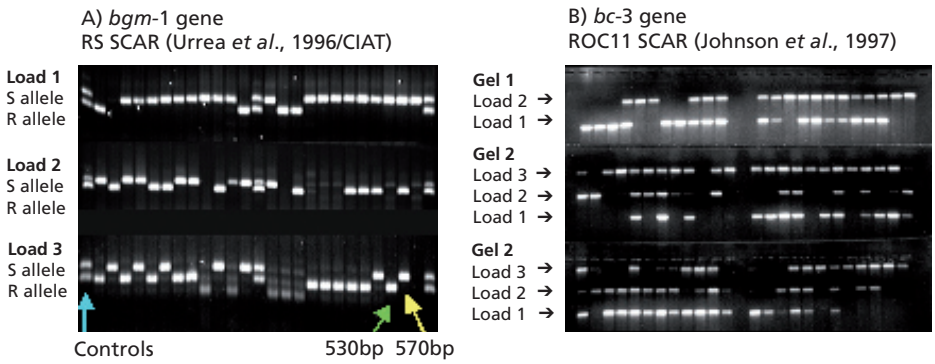
Simple traits

Bean golden yellow mosaic virus resistance

Bean golden yellow mosaic virus (BGYMV) is a white fly-transmitted Gemini virus, and a major production limitation of beans in the mid-to-low altitude areas of Central America, Mexico and the Caribbean. Host resistance to the virus is the most practical means of control, and any new variety in these production areas must carry resistance. Studies on inheritance of resistance revealed a major gene denominated *bgm*-1 in breeding line A429 (Blair and Beaver, 1993) that originates in the Mexican (Durango race) accession "Garrapato" or G2402. Minor genes (Miklas *et al.,* 2000c) as well as additional recessive and dominant resistance genes exist for the virus (Miklas *et al.,* 2006). In most production areas where BGYMV exists, it is necessary to pyramid genes for adequate disease control. Although lines developed in CIAT target these areas, BGYMV does not exist at levels that would permit selection under field conditions in Palmira, Colombia, at CIAT headquarters. Therefore, MAS was desirable to assure recovery of at least the most important resistance genes. MAS has also been employed in the Panamerican School in Zamorano, Honduras, as a complement to field screening, to extend selection to sites and seasons with less disease pressure (J.C. Rosas, personal communication).

A co-dominant RAPD marker was identified for the *bgm*-1 gene (Urrea *et al.,* 1996) that was subsequently converted to a SCAR marker named SR2 (CIAT, 1997). The DNA fragment associated with *bgm*-1 gene has only been observed in one genotype other than G2402 and its derivatives, and thus the polymorphism has been very useful for recognizing the presence of the gene in different genetic backgrounds. This SCAR was evaluated on as many as 7 000 plants in a single sowing (CIAT, 2001; 2003). The uniqueness of the marker's polymorphism and its reliability over laboratories, seasons and genetic backgrounds have facilitated its wide use. More recently, a second SCAR (SW12.700) was developed from the W12.700 RAPD for a QTL located on linkage group b04 (Miklas *et al.,* 2000c), and this has also been incorporated into the breeding programme of CIAT. The combination of *bgm*-1 and the QTL is expected to offer an

FIGURE 1
Examples of gel multiplexing for MAS of A) BGYMV and B) BCMV resistance genes



intermediate level of resistance, while other minor genes must be recovered through conventional phenotypic selection to assure higher resistance.

Scaling up of MAS required the development of simple operational procedures in both the field (tagging, tissue collection) and the laboratory (DNA extraction, marker evaluation). For gamete selection strategies in the field, individual, evenly-spaced plants from segregating populations were marked with numbered tags that were coated with paraffin to protect them until seed harvest. Leaf disks were sampled from young vegetative tissue with a paper hole puncher and placed directly into pre-numbered cells of microtitre 96-well plates stored on ice, ready for grinding and extraction in the laboratory. The implementation of MAS for *bgm*-1 and subsequently for SW12.700 in the laboratory required substantial adaptation of standard protocols to establish high-throughput procedures. Grinding of samples in microtitre plates was accomplished with a block of 96 pegs that fit into each well. Alkaline DNA extraction (Klimyuk *et al*., 1993) was employed with success for both markers, and eventually it was possible to multiplex the markers in both the amplification and gel phases using multiple primer PCR and multiple loading per gel wells (Figure 1A). With experience and improved procedures, efficiency more than doubled over a two-year period. MAS was often carried out before flowering to decide on a plant's status as a carrier of the resistant allele for further use in crossing.

Two small red seeded lines developed in the Panamerican School using MAS have reached the stage of validation in Honduras (J.C. Rosas, personal communication) and shown resistance to the BGYMV strains prevalent there. Resistance to BGYMV of drought tolerant lines selected at CIAT was maintained using MAS for one or more genes, followed by field selection in Central America. Similarly, red mottled lines developed in CIAT with the aid of MAS showed field resistance in the Caribbean and one of these lines from the red mottled advanced line for the Caribbean (RMC) series has been released (Blair *et al.,* 2006). MAS has also been an important element of maintaining BGYMV viral resistance in CIAT's programme as other breeding objectives such as nutritional value have been assumed, necessitating the inclusion of susceptible parents in crosses with resistant

lines. MAS for this trait has also been practised at the University of Puerto Rico and at the Biotechnology Institute of Cuba.

Bean common mosaic virus and bean common mosaic necrotic virus

Bean common mosaic virus (BCMV) and the related necrotic strains (bean common mosaic necrotic virus [BCMNV]) are aphid-transmitted potyviruses that are found worldwide and are seed-borne from season to season. BCMNV resistance is very important in Africa where necrotic strains are prevalent and has become a renewed priority for parts of the Caribbean where necrotic strains have been discovered. BCMV is also endemic in the Andean region where it persists in farmer-saved seed and long-season climbing beans. Climbing beans are grown in both intensive (trellised/staked monoculture) and extensive (intercropping with maize) farming systems. In both systems the need to protect the crop from easily transmitted viral diseases such as BCMV or BCMNV is great; however, very few climbing beans have been bred for resistance to BCMV. A number of BCMV/BCMNV resistance genes have been tagged including the dominant *I* gene (with which the necrotic strains interact to produce necrosis) and the recessive *bc*-3, *bc*-2 and *bc*-1^2 genes (Haley, Afanador and Kelly, 1994; Melotto, Afanador and Kelly, 1996; Johnson *et al.,* 1997; Miklas *et al*., 2000a). The genes can be distinguished by inoculation with different viral isolates, and a range of molecular marker tags are available for each gene (reviewed in Kelly *et al*., 2003; Miklas *et al.,* 2006). The dominant *I* gene was incorporated into a wide range of small seeded bush beans at CIAT, while resistant bush beans of the bush bean resistant to black root (BRB) series carrying recessive genes were developed in the 1990s and have been widely distributed as breeding parents. The need to reselect the recessive genes with confidence from segregating populations makes MAS a priority.

CIAT started a collaborative project with the Colombian national bean programme based at the Colombian Agricultural Research Corporation (CORPOICA) in 2002 to introgress BCMV resistance genes from BRB lines into local landraces and improved genotypes of Andean climbing beans (CIAT, 2002, 2003, 2004; Santana *et al.,* 2004). During the breeding programme for BCMV and over the course of four years, MAS was used extensively based primarily on the SCAR marker ROC11 developed for the *bc*-3 gene (Johnson *et al.,* 1997) and the SCAR marker SW13 for the *I* gene (Melotto, Afanador and Kelly, 1996) along with virus screening to confirm the selection of resistant progeny. The programme was successful in moving *bc*-3 resistance into a background of cream mottled and red mottled seed types for both highland areas (known as Cargamanto commercial class) as well as mid-altitude areas through triple-, double- and backcrosses. Although virus resistance was also screened phenotypically, the frequency of escape, the complex interaction of multiple genes and the recessive nature of most of these made MAS the best option for breeding resistant varieties rapidly. In addition, as climbing bean breeding is a more time-consuming and expensive endeavour than bush bean breeding due to the longer season, wider plant spacing and need for staking material, MAS was also found to be a very effective measure to reduce breeding costs and save on breeding nursery space.

The implementation of MAS for BCMV was based on a combination of the previously developed SCAR markers previously mentioned and techniques developed at

CIAT for the selection of BGYMV resistance as discussed previously. Although most BCMV and BCMNV resistance genes had been tagged with SCAR markers, implementation required efforts to validate and scale up the use of the markers in applied breeding programmes. Genotyping for the ROC11 marker was carried out on advanced lines given that this marker is dominant and in repulsion with the resistance allele. In other words, the absence of a band was indicative of the presence of the recessive *bc*-3 allele and therefore it was more appropriate to evaluate after fixation of the alleles to homozygosity through mass or pedigree selection with single plant selections in the F_4 and F_5 generation when single plant rows were evaluated for the resistance gene marker. To determine whether the advanced line continued to segregate for the gene, alkaline DNA extraction was conducted on leaf discs collected from four leaflets from four individual plants per line using a hole-puncher rather than from a single plant per family or advanced line. The presence or absence of polymerase chain reaction (PCR) products was evaluated for each genotype based on scanned photographs or gel capture imagery of multiplexed gels (Figure 1B) to predict if the genotype contained the resistance or the susceptible allele.

Once optimized for parental genotypes, MAS was conducted on a large number of progeny rows. For example in 2003, more than 4 000 advanced lines were evaluated for the ROC11 marker for genotypes grown at three sites within Colombia (CIAT-Darien, CIAT headquarters and CORPOICA-Rionegro). DNA was collected at all three sites and shipped successfully to the laboratory in 96-well plate format as discussed above. Both the ROC11 and SW13 markers were single copy SCARs that did not produce extra bands and therefore were easy to multiplex. To facilitate the evaluation of markers on a large number of advanced lines, usually within two to three weeks, and increase the efficiency of MAS, several innovations were implemented: loading of agarose gels (first with two and then three loadings), increasing numbers of wells per comb (first 30-well and then 42-well combs were used), use of 384-well PCR plates and multipipetor loading of gels. The resulting savings decreased the time to PCR amplify and load a gel by approximately 50 percent and increased the number of genotypes run per gel by 225 percent.

The rapid increase in efficiency obtained during the application of the ROC11 marker shows the advantages of testing new markers in practical breeding programmes. The use and advantages of these molecular markers has been presented at an Organization of American States-sponsored course in Colombia given in 2002 and a Rockefeller Foundation-sponsored course in Uganda given in 2003. Based on this programme and the training courses, MAS for BCMV genes was initiated as part of a recently approved Association for Strengthening Agricultural Research in Eastern and Central Africa (ASARECA) project for three countries in eastern Africa and training of researchers from the Andean region has allowed more breeding lines from Peru to be screened (CIAT, 2004).

Other examples of MAS for simply inherited traits

Several pathogens, especially fungal pathogens, have co-evolved with the bean host, and present a population structure (Andean/MesoAmerican) that mimics the major gene pools of bean (Pastor-Corrales, Jara and Singh, 1998). This is the case with *Phaeoisariopsis griseola*, the

causal agent of angular leaf spot (ALS), and *Colletotrichum lindemutheanum*, which induces anthracnose. In both cases, pathogen isolates tend to be more virulent on host genotypes of the same gene pool (Andean or Mesoamerican) and less so on host genotypes from the contrary gene pool. Resistance genes of utility to one host gene pool thus tend to originate in the other gene pool and require introgression from one gene pool to the other. MAS has great potential for introgression as DNA polymorphisms are maximized in wide crosses across gene pools, and markers are available for this purpose for both ALS (Carvalho *et al.,* 1998; Sartorato *et al.,* 1999; Nietsche *et al*., 2000; Ferreira *et al.,* 2000; Mahuku *et al.,* 2004) and anthracnose (Young *et al*., 1998; Awale and Kelly 2001; Vallejo and Kelly, 2001).

Other cases of wide crosses in which MAS can be of use include those for the selection of genes for resistance to a storage insect, the Mexican bean weevil (*Zabrotes subfasciatus* [Boheman]) derived from wild bean accessions from Mexico. Selection for resistance has also been achieved by analysis for the active resistance agent, a seed protein called arcelin, by either antibody reaction or electrophoresis, but MAS is simpler and more efficient than either of these analyses that require protein extraction. Even wider crosses of common bean with *Phaseolus acutifolius* have recovered resistance to common bacterial blight (caused by *Xanthomonas axonopodis* pv. *phaseoli*) (Muñoz *et al.,* 2004) and markers have also been developed for these resistance genes (Jung *et al.,* 1997; Miklas *et al.,* 2000b; Park *et al*., 1999; CIAT, unpublished data). In these cases also, the fact of deploying genes from relatively wide crosses favours maintaining a state of DNA polymorphism in relation to the target genotypes.

Complex multigenic traits

In addition to the studies previously discussed, several attempts have been carried out to tag quantitative trait loci (QTL) for abiotic stress tolerance or insect resistance in common bean, although most of these traits might better be described as oligogenic, as results usually suggest that a limited number of loci (from three to six) are involved in their genetic control.

One example is tolerance to low soil phosphorus that was investigated in the landrace G21212. Linkage group b08 proved to be especially important to yield under low phosphorus, with as many as three important and loosely linked QTL (Beebe, Velasco and Pedraza, 1999; Miklas *et al.,* 2006). Interestingly, these same QTL were linked to QTL for resistance to *Thrips palmi* Karny derived from the same source (Frei *et al.,* 2005). This is a promising candidate for applying MAS in the short term for abiotic stress tolerance, although another notable attempt was also made for drought tolerance breeding with MAS through a joint programme between Michigan State University and the National Institute for Forestry, Agriculture and Livestock Research (INIFAP) in Mexico (Schneider, Brothers and Kelly, 1997).

In theory, a breeder would prefer markers for low heritability quantitative traits that are difficult to select through phenotypic selection. However, in general, markers for polygenic or oligogenic traits have not moved into the application phase. The same problems that make phenotypic selection difficult apply in some degree to MAS. Multiple minor genes that are often associated with poor heritability also imply that it is difficult to identify QTL with highly significant effects and that merit the investment of MAS. Furthermore, good genome coverage is usually necessary to

detect the QTL that explain the highest amount of genetic variability, and this has been difficult to achieve in intragene pool crosses in common beans.

However, genetic analysis by markers has been very useful for revealing the inheritance of quantitative traits, especially physiological traits, even when the markers involved did not result in application in MAS. Analysis of QTL was applied to root traits of bean as they relate to absorption of phosphorus from soil (Liao *et al.*, 2004; Yan *et al.*, 2005; Beebe *et al.*, 2006). This permitted associating different physiological traits to P uptake and estimating their importance in nutrient acquisition. Once traits are better understood, then an appropriate selection strategy can be devised, be it phenotypic or MAS. Thus, markers can be useful to a breeding programme by elucidating basic plant mechanisms even if they are not applied directly in selection.

Breeding schemes: adaptation to include MAS

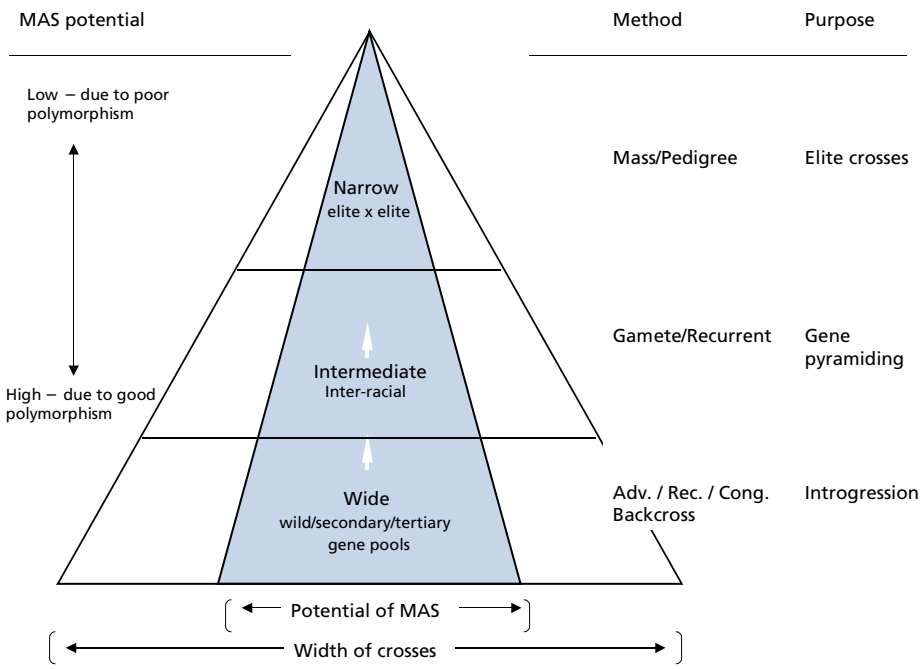
The eventual application of MAS requires careful prioritization of traits and even specific genes for which markers are to be sought, in light of the importance of the trait and genes, and options for phenotypic selection. One should never assume that MAS is necessarily superior to phenotypic selection, which for some traits may be as effective and efficient as the use of molecular markers. However, if a gene is sufficiently important in a breeding programme to demand that advanced lines have such a gene (as in the case of the *bgm*-1 gene for virus resistance in Central America), there is probably some point in the selection process at which MAS would be useful. Also, it is not necessary to select many genes by MAS for it to be of great value. For example, if a single gene is segregating and 50 percent of plants lack the gene in advanced generations, an effective selection would eliminate half the population and increase the subsequent efficiency of the breeding programme by a factor of 2.

Once markers are available, a key issue is determining the range of parental genotypes within which a marker is polymorphic and therefore useful for selection. Markers of genes that originate from wider crosses (e.g. from different races, gene pools or species) will have a progressively greater chance of being polymorphic among a range of parents (Figure 2) and therefore diagnostic for the gene of interest. The example of *bgm*-1 is again a good case in point as the resistance allele and the SR2 marker are both unique to the Durango gene pool and polymorphic in combinations across other Mesoamerican races as well as the Andean gene pool. In contrast, the ROC11 marker for the *bc*-3 gene is only polymorphic across gene pools and therefore not diagnostic for the resistant allele.

If a breeder has several potential parents among which to choose and these are comparable with regard to other traits, it might be preferable to eliminate those that carry a band that would be confused with the linked marker and would result in false positives. Conversely, if more than one marker is available for a given gene, one might focus on those linked markers that maintain polymorphism in the greater number of combinations. In some combinations it might be informative to use both linked markers simultaneously, both to discern recombinants and to confirm markers.

Several possible schemes for the introduction of MAS to different breeding schemes are represented in Figure 3. A breeder must consider at what generation in the breeding programme selection

FIGURE 2
Potential of MAS in crosses of varying genetic diversity



Modified from Kelly, Schneider & Kolkman, 1999 and Singh, 1999

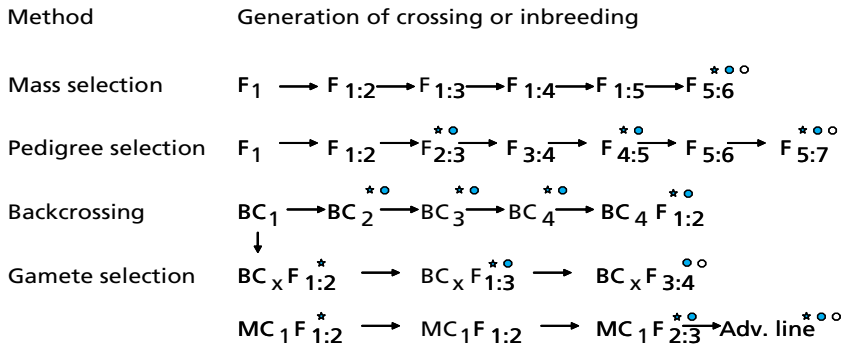
by MAS will give the greatest cost/benefit ratio. This would probably be early in the breeding programme for the pedigree method or for gamete selection while it would be later in the programme for bulk method or mass selection (Figure 3). In the case of early generation selection, elimination of plants without the gene(s) will avoid unproductive investment.

Advantages and disadvantages of MAS

MAS provides real advantages where the conditions are not favourable for phenotypic selection, for example, in the case of BGYMV, which does not exist at epiphytotic levels in CIAT. Indeed, *bgm*-1 behaves as a recessive gene, so phenotypic selection in early generations would be inefficient in recovering the gene in the heterozygous state.

The same principle would apply to the recessive *bc*-3 gene, although the lack of a marker linked in coupling to this gene has been a serious drawback and has limited the effectiveness of MAS to advanced generations when the gene is fixed by inbreeding. In this case, early generation selection with MAS would be limited to negative selection against homozygous dominant and heterozygous plants, and this eliminates potentially useful allele-carrying genotypes. Indeed, MAS is impossible in generations such as the F_1 or BC_1F_1 to the susceptible parent when no homozygous recessive plants exist at all.

FIGURE 3
Application of co-dominant and dominant markers during the breeding cycle in common beans



Note: Asterisks indicate co-dominant marker, while open circles indicate dominant marker in repulsion and closed circles indicate dominant marker in coupling.

In other cases where phenotypic selection methods are available, the advantage of MAS resides in its simplicity. This is the case in the selection of arcelin, which can be achieved through protein extraction followed by antibody detection or electrophoresis, but both of these are laborious while MAS can be applied more rapidly and with much greater throughput. Similarly, markers for common blight resistance and anthracnose have the advantage of obviating the need for field inoculations that are sometimes ineffective if environmental conditions are not favourable. The advantage of MAS is much greater if a single DNA extraction can serve for the evaluation of several markers, as in the multiplexing of *bgm*-1 and SW12.700 markers.

In spite of attempts to apply MAS to complex traits, examples of successful application are still limited to relatively simple traits. This is contrary to some previous expectations that markers would benefit mostly traits of low heritability. However, experience has shown that the ability to manipulate even one important gene with confidence can make a breeding programme more efficient, if that gene is highly desirable and valuable for advanced materials.

Meanwhile the disadvantages of MAS compared with phenotypic selection are based on effectiveness and cost considerations. The effectiveness of MAS is relative to the ease of applying a given marker, its reliability and its level of linkage with the gene of interest. Although molecular markers theoretically have a heritability of 1.0, variability among laboratories or among runs within a laboratory make markers less than 100 percent reliable. This is especially true for RAPD markers for which band amplification is dependent on DNA concentration and quality, annealing temperature and thermocycling conditions, Taq polymerase concentration and the relative proportion of various other ingredients to the PCR cocktail. In comparison, SCAR markers are much more reliable and repeatable and therefore have higher heritability than RAPD markers. Linkage distance between a marker for a

gene of interest and the actual locus itself also affects the reliability of a marker. In turn, the type of cross (wide versus narrow) and parents involved (closely or distantly related) affect the frequency of recombination around introgressed genes as well as the level of polymorphism of the cross and whether the marker will present distinct alleles for the desirable and undesirable character states. In this regard, there is a tradeoff as MAS is most effective when there is high polymorphism in the crosses being evaluated (Figure 2). However, this is precisely the breeding situation in which gene introgression is most difficult, time-consuming and plagued by linkage drag, as is the case for interspecific or interspecific-derived crosses, hybridizations with wild or wild-derived genotypes and crosses between the Andean and Mesoamerican gene pools. This issue is being addressed in beans with the development and mapping of microsatellite markers (Blair *et al.,* 2003) that are much more polymorphic and useful for diagnosing the inheritance of genomic segments in narrow crosses. The first application of microsatellite markers for MAS in common beans was the selection of arcelin based bruchid resistance using gene-derived simple sequence repeats that are diagnostic for the introgression of alleles for resistance from wild beans into cultivated backgrounds (CIAT, 2004), but others should also show promise.

In terms of cost considerations, the relative costs of MAS versus phenotypic selection are relative to each trait and situation. The widely held perception that MAS is expensive is often due to the ingredients and time used to prepare DNA extractions and PCR reactions, although these costs have been reduced by innovations such as the alkaline DNA extraction technique (Klimyuk *et al.,* 1993) that obviates the need for organic solvents or expensive enzymes involved in other mini-preparation techniques (Afanador and Hadley, 1993). While experienced labour was previously required for DNA extraction at CIAT or in NARS breeding programmes, the alkaline extraction method allows most laboratory steps to be carried out even by untrained personnel. Furthermore, MAS costs can be reduced by miniaturization, especially in the PCR reaction (for example, use of 384-well PCR plates and small reaction volumes) and re-use of ingredients (for example plasticware including pipette tips and microtitre plates as well as agarose from used gels). As previously mentioned, multiplexing adds to the efficiency and therefore reduces the datapoint costs of MAS.

Currently, MAS with SCAR markers and alkaline extraction at CIAT cost less than US$0.25 per datapoint. Therefore the expense of MAS is now not as important an issue as previously. In this regard, MAS sometimes has the advantage of being implemented in any generation and under both field or greenhouse conditions, while phenotypic selection often requires a separate planting and specialized labour for inoculation, agronomic management and evaluations or scoring. However, in the final analysis, the most efficient and cost effective breeding programme will probably be one that combines MAS and phenotypic selection in some optimal combination. It is precisely the challenge of the breeder to define that optimal combination.

One last disadvantage of relying on MAS is that it commits a breeder to a unique gene(s) for a given trait. For example, there might be multiple genes or gene combinations for resistance to a disease, or for a physiological trait such as root structure. To the extent that a breeder relies on MAS for selection, this excludes other

possible genes and the use of other potentially useful parents that do not share the DNA polymorphism that is used in MAS. On the other hand, phenotypic selection would permit recognizing different genetic options for a desired phenotype. Thus, MAS is most useful when it is applied to truly unique genes.

CASSAVA: IMPORTANCE AND GENETICS

Cassava is a perennial shrub but it is generally harvested as an annual crop at 10–11 months of age. Basically every part of the plant can be utilized. The starchy roots are a valuable source of energy and can be boiled or processed in different ways for human consumption and different industrial purposes such as starches, animal feed or alcohol (Ceballos *et al*., 2006). Cassava storage roots are not tubers and therefore cannot be used for reproductive purposes; stems are the common planting materials. Cassava foliage is not widely exploited in spite of its high nutritive value (Buitrago, 1990; Babu and Chatterjee, 1999). Foliage consumption by humans is relatively common in certain countries of Africa, Asia and Latin America. The use of foliage for animal feeding is generating increased interest in Asia.

Cassava can be propagated by either stem cuttings or botanical seed. However, the former is the practice most widely used by farmers for multiplication and planting purposes. Propagation from true seed occurs under natural conditions and is common in breeding programmes. Occasionally botanical seed is also used in commercial propagation schemes (Rajendran *et al.,* 2000).

Cassava is monoecious and allogamous, with female flowers opening 10–14 days before the male ones on the same branch. Pollination can be done manually in a controlled way to produce full-sib families or else in polycross nurseries where open pollination takes place and, therefore, half-sib families are produced. Self-pollination is feasible when using male and female flowers on different branches or on different plants of the same genotypes (Jennings and Iglesias, 2002). Some clones flower relatively early at four or five months after planting whereas others only do so at eight to ten months after planting. As a result, the time required for the seed to mature, the growing cycle of the crop and the need to plant with the arrival of the rains take about two years between a given cross being planned and the respective seed becoming available. On average, between one and two seeds (out of the three possible in the trilocular fruit) per pollination are obtained (Kawano, 1980; Jennings and Iglesias, 2002).

Breeding objectives

Productivity plays a major role in industrial uses of cassava, whereas stability of production is fundamental in the many regions where cassava is the main subsistence crop. Industrial uses of cassava require high dry matter content as the main quality trait for the roots, whereas for human consumption the emphasis is on cooking quality, frequently even over productivity, as the determining trait. Stability of production is associated with resistance or tolerance to major biotic and abiotic stresses, with the emphasis varying with the target environment. Genetic resistance to the most important diseases and pests and the prevalent abiotic stresses can be found in cassava germplasm (Hillocks and Wydra, 2002; Bellotti *et al.,* 2002; Belloti, 2002; Ceballos *et al.,* 2004). Although cyanogenic glucosides are found in every tissue except

the cassava seed, most processing methods allow a rapid release and elimination of the cyanide. Depending on the end use, high or low cyanide clones are preferred. Other relevant traits for the roots are dry matter, protein and carotenoid content (Chávez *et al.*, 2005).

Breeding schemes

Genetic improvement of clonally propagated non-inbred crops such as cassava is made possible by the fact that a superior genotype can be fixed at any stage in the breeding scheme, even after a single cross, the equivalent of an F_1 in commercial hybrids such as maize. Therefore, non-additive gene actions including dominance and epitasis become important components of the genetic variance to be manipulated by the breeder (Jaramillo *et al.*, 2005; Calle *et al.*, 2005; Perez *et al*., 2005a). Large effective breeding population sizes are required to retain favourable dominant alleles and epistatic loci combination.

As in most crop breeding activities, cassava genetic improvement starts with the production of new recombinant genotypes derived from selected elite clones. Scientific cassava breeding began only a few decades ago, and the divergence between landraces and improved germplasm is not as wide as in other crops. Therefore, accessions for germplasm bank collections from different research institutions play a more relevant role in cassava than in other crops that have been scientifically bred for longer periods of time. Parental lines are selected based mainly on their performance *per se* and little progress has been made to use general combining ability (Hallauer and Miranda Fo, 1988) as a criterion for parental selection. Sexual seeds obtained by the different crossing schemes are germinated to initiate a new cycle of selection. The multiplication rate of cassava planting material is low as five to ten cuttings can be obtained from one plant. This implies a lengthy selection process, and in fact it takes about six years from the time the botanical seed is germinated until enough planting material is available for multilocation replicated trials.

Table 1 illustrates a typical selection cycle in cassava. It begins with the crossing of elite clones and finishes when the few clones surviving the selection process reach the stage of regional trials across several locations. It should be emphasized that there is some variation among the few cassava-breeding programmes in the world with respect to the number of genotypes

TABLE 1
Typical selection cycle in cassava beginning with the crossing of elite clones to the point when few clones surviving the selection process reach the stage of regional trials across several locations

Year	Activity	Number	Plants per genotype
1-2	Crosses among elite clones planned, nurseries planted and pollinations made	Up to 100 000	
3	F_1: Evaluation of seedlings from botanical seeds. Strong selection for African cassava mosaic virus (ACMV) in Africa.	100 000[a]; 50 0000[b]; 50 000 [c]	1
4	Clonal evaluation trial (CET)	20 000–30 000 [a, b] 700 [c]	6–8 (1 rep, 1 location)
5	Preliminary yield trial (PYT)	100 [a]; 300 [b]; 80 [c]	20–60 (3 reps, 1 location)
6	Advanced yield trial (AYT)	25 [a]; 100 [b]; 20–25 [c]	100–500 (3 reps, 2–3 location)
7-9	Regional trials (RT)	5-30 [a, b, c]	500-4 000 (3 reps, 3–4 locations)

Figures for cassava breeding at [a] IITA (Ibadan, Nigeria); [b] CIAT (Cali, Colombia) and [c] CIAT and Rayong Field Research Station from Department of Agriculture (Thailand).
Source: adapted from Jennings and Iglesias, 2002.

and plants representing them through the different stages. Table 1 also provides an idea of the selection pressures generally applied.

Strong emphasis on highly heritable traits (plant type, branching habits and reaction to diseases, harvest index and dry matter content) is applied during the early phases of selection (F_1 and CET), (Hahn, Howland and Terry, 1980; Hahn, Terry and Leuschner, 1980; Hershey, 1984; Kawano, 2003; Ceballos *et al*., 2004). As the number of plants representing each genotype increases, the weight of selection criteria shifts towards low heritability traits such as root yield. The clones that show outstanding performance in the regional trials are released as new varieties and, eventually, incorporated as parents in the crossing nurseries. With that the selection cycle is finished and a new one begins. The whole process has the following characteristics (Ceballos *et al.*, 2004):

- the process is indeed phenotypic selection because no family data are involved;
- no data are collected in the early stages of selection. Therefore, data regarding general combining ability effects (~ breeding value) are not available for a better selection of parental materials;
- there is no proper separation between general (GCA ~ additive) and specific (SCA ~ heterotic) combining ability effects. The outstanding performance of selected materials is likely to depend on positive heterotic effects that cannot be transferred to the progenies that are sexually derived from them;
- no inbreeding is incorporated purposely in the selection process. Therefore, large genetic loads are likely to remain hidden in cassava populations and useful recessive traits are difficult to detect;
- several stages of selection are based on unreplicated trials. A large proportion of genotypes is eliminated without proper evaluation.

For the above-mentioned reasons, cassava breeding is difficult, expensive and to a certain degree inefficient (Perez *et al.*, 2005a; Cach *et al*., 2005a, b). Kawano *et al.* (1998) mention that, during a 14-year period about 372 000 genotypes derived from 4 130 crosses were evaluated at the CIAT-Rayong Field Crop Research Center. Only three genotypes emerged from the selection process to be released as official varieties. Similar experiences have been observed at the International Institute of Tropical Agriculture (IITA), CIAT-Colombia and Brazil. Therefore, the development and adaptation of molecular tools for cassava genetic improvement offer important advantages to make the process more efficient and effective.

MAS in cassava breeding

Cassava genetic improvement can be made more efficient through the use of easily assayable molecular genetic or DNA markers (MAS) that enable the precise identification of genotype without the confounding effect of the environment, thereby increasing heritability. MAS can also contribute to the efficient reduction of large breeding populations at the seedling stage based upon "minimum selection criteria". This is particularly important given the length of the growing cycle of cassava and the expense involved in the evaluation process. Therefore, a pre-selection at the F_1 phase (see Table 1) could greatly enhance the efficiency of the CET experiments. The selection of progenies based on genetic values derived from molecular marker data substantially increases the rate of genetic gain, especially if the number of cycles of evaluation or generations can be reduced (Meuwissen, Hayes and Goddard, 2001).

Another application of MAS in cassava breeding is reducing the length of time required for the introgression of traits from wild relatives. Wild relatives are important sources of genes for pest and disease resistance in cassava (Hahn, Howland and Terry, 1980; Hahn, Terry and Leuschner, 1980; Chavarriaga *et al.,* 2004), but the need to reduce or eliminate undesirable donor genome content and linkage drag can lengthen the process, making it unrealistic for most breeders. Simulations by Stam and Zeven (1981) indicate that markers could reduce linkage drag and would reduce the number of generations required in the backcross scheme. Hospital, Chevalet and Mulsant (1992) corroborated this in achieving a reduction of two backcross generations with the use of molecular marker selection. Frisch, Bohn and Melchinger (1999), through a simulation study, found that use of molecular markers for the introgression of a single target allele saved two to four backcross generations. They inferred that MAS had the potential to reach the same level of recurrent parent genome in generation BC_3 as reached in BC_7 without molecular markers.

The decision to employ DNA-based markers in cassava breeding is primarily based on the heritability of a trait and the amount of genotypic variance explained by the marker. There are many instances in cassava breeding where h^2 is low or zero. Some examples are:

- plant health traits where the pathogen or pest pressure is absent or low, such as cassava mosaic disease (CMD) in the New World tropics or cassava green mite (CGM) during the wet season;
- variable or erratic pest pressure, e.g. the CGM or diseases such as the cassava frog skin disease (FSD);
- evaluation based upon a single plant;
- variable experimental fields and/or poor management resulting in large experimental errors;
- traits that are affected by the stage of plant growth or the part of the organ used for tissue analysis, e.g. cyanogenic potential.

In the above-mentioned instances, having a marker(s) that explains a large proportion of the genetic variance can accelerate progress in breeding. Even where h^2 is moderate or high, selection by markers can be advantageous:

- where different sources of genes exist for the trait that are indistinguishable by phenotype alone and pyramiding is difficult and time consuming, e.g. for different sources of resistance to a disease or pest;
- where molecular tags that can be used inexpensively and rapidly to identify desirable genotypes early in the breeding cycle exist, thereby eliminating the need to evaluate large numbers of plants phenotypically, and obviating the confounding effects of the environment. Markers may permit the efficient elimination of undesirable genotypes at the seedling stage. For example, the number of genotypes at the seedling stage can be reduced by 50 percent if a trait is controlled by a single gene, or by 87.5 percent if controlled by three genes;
- for the introgression of useful genes from exotic germplasm into adapted gene pools. MAS can be used to identify genotypes that carry minimal amounts of flanking donor parent genome around the gene of interest for faster backcrossing;
- for definition of heterotic pools in a group of germplasm accessions for more directed crosses;

- for definition of average heterozygosity in the selection of partially inbred lines for tolerance to inbreeding;
- for identification of the male parent in elite germplasm derived from polycrosses by fingerprinting. This tool is also useful for checking the identity of different genotypes to eliminate duplication in germplasm collections.

Best results are achieved when MAS is combined with phenotypic data as compared with either approach independently (Hospital, Chevalet and Mulsant, 1992). Phenotypic data would reduce the cost of genotyping especially if phenotypic evaluation is conducted on early generations (Gimelfarb and Lande, 1994). This not only reduces the cost of MAS but also increases its efficiency. Some examples of MAS in cassava breeding conducted at an international centre and national programmes are described below.

Molecular MAS for CMD resistance at an IARC

An ideal target for MAS is breeding for disease resistance in the absence of the pathogen. This is the case of CMD in the Americas, where the disease does not occur. CMD is a viral disease first reported by Warburg in 1894 in eastern Africa (quoted by Storey and Nichols, 1938). Several variants of the disease (East Africa cassava mosaic virus [EACMV], South Africa cassava mosaic virus [SACMV], Indian cassava mosaic virus [ICMV]) have been reported (Swanson and Harrison, 1994) and are endemic in all cassava growing regions of Africa and southern India, where it is the most severe production constraint. The white fly vector of CMD, *Bemisia tabaci* biotype A, does not colonize cassava in the New World but recently a new biotype of *B. tabaci*, biotype B (also referred to as *B. argentifolia*), has become widespread in the Americas and has a wide host range including cassava (Polston and Anderson, 1997), increasing the possibility that CMD, EACMV, SACMV, ICMV or a native American gemini virus will become established on cassava in the neo-tropics. This is a frightening prospect for cassava production in Latin America, considering that most Latin American cassava germplasm is very susceptible to CMD (Okogbenin *et al*., 1998). The susceptibility of neo-tropical germplasm to CMD also limits the utilization of germplasm from the crop's centre of diversity in the neo-tropics for these key cassava production regions. Breeding for resistance to CMD in Latin America, where the disease does not exist and is unlikely to be introduced due to very strict quarantine controls, requires the tools of MAS.

Evaluations at IITA identified an excellent source of resistance to CMD in some Nigerian landraces (A.G.O. Dixon 1989, unpublished data), namely TME3, TME7, TME5, TME8, TME14 and TME28. This resistance is effective against all known strains of the virus, including the virulent Ugandan variant (UgV) (Akano *et al.,* 2002; CIAT, 2001). CIAT, in collaboration with IITA in Ibadan, Nigeria, and with support from the Rockefeller Foundation, devel-oped several molecular markers for this source of CMD resistance, revealed to be controlled by a single dominant gene designated as *CMD2* (Akano *et al.,* 2002). At least five markers tightly associated to *CMD2* have been developed, the closest being RME1 and NS158 at distances of four and seven cM respectively. The dominant nature of *CMD2* and its effectiveness against a wide spectrum of viral strains makes its deployment very appealing for protecting cassava against the actual or potential ravages of CMD

in both Africa and Latin America. CIAT and IITA undertook a project to verify the utility of these markers for MAS in breeding CMD resistance by developing crosses between the sources of TME3 and susceptible varieties. A total of six families, ranging in size from 36–840 genotypes, and a total of 2 490 genotypes were used. The crosses were genotyped with two markers and also evaluated for CMD resistance in a high CMD pressure area in Nigeria. Results of the marker analysis and phenotypic evaluation of CMD resistance in the field revealed that the markers RME1 and NS158 SSR were excellent prediction tools for CMD resistance in some crosses (a prediction accuracy of 70–80 percent). In a few families, however, the markers were not polymorphic between the resistant and susceptible parent and, therefore, were not useful. This highlights the need to develop many markers around a gene of interest in a MAS programme and then to use those markers to evaluate the parents and identify the best markers for the different cross combinations.

Eighteen progenies from TME3 carrying the *CMD2* marker were established from embryo axes and imported to CIAT from IITA.[1] They were crossed extensively to elite parents. Seeds harvested from the crosses were germinated *in vitro* from embryo axes according to standard protocols for cassava (Fregene *et al.,* 1997, CIAT, 2002) to allow sharing the CMD resistant genotypes with collaborators in Africa and India. Each plantlet was multiplied after three to four weeks of growth to obtain three to five plants. After another four weeks, leaves of all plants were removed for molecular analysis and the plants multiplied again to obtain 10–20 plantlets. DNA isolation was by a rapid mini preparation method developed for rice (Nobuyuki *et al*., 2000). The DNA obtained is sufficient for 100 reactions and can be held in the Costar plates for two months at –20 °C without any degradation. PCR amplification, polyacrylamide gel electrophoresis (PAGE) or agarose gel analysis of SSR markers NS158 and RME1 were as described by Mba *et al*. (2001). The versatility of spreadsheets makes them the appropriate software to handle the diverse information generated by MAS. Gel images from the marker analysis were entered directly into a spreadsheet that contains information on the parents, tissue culture and greenhouse records, and subsequent phenotypic evaluation of the progenies. After molecular analysis, genotypes that carry the marker allele associated with *CMD2* were further multiplied to obtain at least 30 plants. Ten plants were sent to the greenhouse for hardening and later transferred to the breeding programme for evaluation. Five plants were kept *in vitro*, while 15 plants were shipped to partners in India and Africa as shown in the flow chart for MAS (Figure 4).

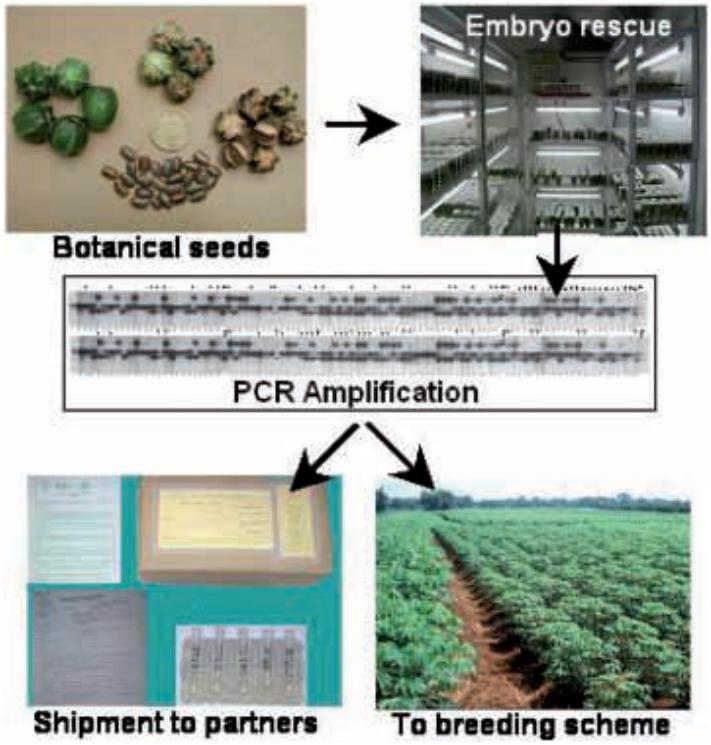
To date, more than 50 000 progeny have been evaluated with CMD linked markers and resistant lines shared with national programmes in India or Africa, and also incorporated into the breeding scheme at CIAT. The cost of a single marker data point is US$0.30 and 32 000 samples can be processed in a year.

MAS for CMD resistance at a NARS

Although evaluation for CMD resistance in sub-Saharan Africa is relatively easy and most areas have sufficient disease pressure to permit moderate to high heritability

[1] Phytosanitary conditions for the exchange of cassava germplasm between Africa and Asia are very stringent, but appropriately indexed *in vitro* cultures of embryo axes are permitted for experimental purposes.

FIGURE 4
Schematic representation of steps employed in breeding for resistance to the CMD in Latin America cassava gene pools

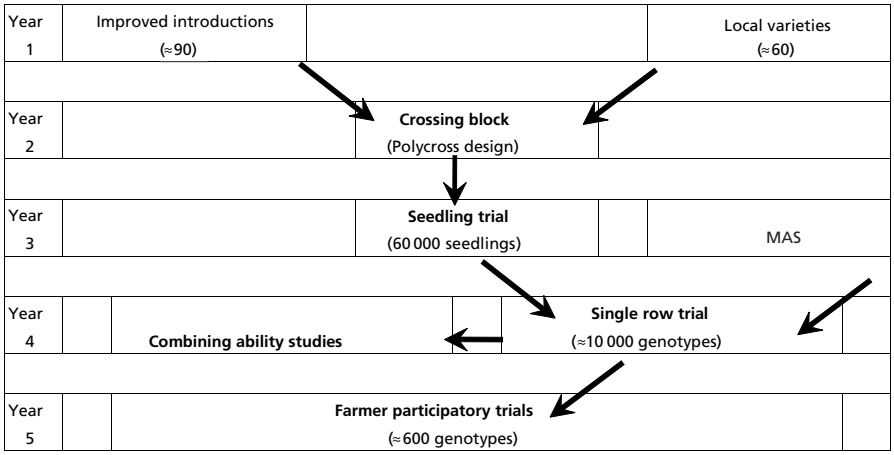


The entire process from sexual seeds to tissue plants for shipment or transfer to the screen house takes approximately three months.

of resistance, overlapping outbreaks of CGM, cassava bacterial blight (CBB), and CMD are common (Legg and Ogwal, 1998) and the need for modest-sized breeding populations make MAS for CMD resistance a powerful tool to accelerate cassava improvement even in Africa. A MAS and participatory plant breeding (PPB) project was initiated in 2003 with funding from the Rockefeller Foundation to improve the resistance of local cassava varieties in the United Republic of Tanzania to CMD and CGM and also to provide proof of concept for the use of MAS to accelerate cassava improvement. The United Republic of Tanzania is the fourth largest producer of cassava in Africa with average yields of about 8 tonnes/ha (FAO, 2001). This is below the continent's average of 10 tonnes/ha, and well below the average yield of 14 tonnes/ha of Africa's (and the world's) largest producer, Nigeria.

The low yield in the United Republic of Tanzania is caused by many factors, including the susceptibility of commonly grown varieties to major diseases and pests such as CMD and the cassava brown streak disease (CBSD). The project crosses farmer-preferred germplasm, by agro-ecology, to improved introductions that are resistant to CMD and to CGM. Markers associated with resistance to CMD are used to reduce

FIGURE 5
MAS scheme to improve local varieties of cassava in the United Republic of Tanzania using improved disease and pest resistant introductions from Latin America



The scheme is now in its second year.

the population size and a small set of genotypes with the "minimum criteria" for successful cassava production are evaluated in a single season in the corresponding agroecology and then evaluated over two cycles in collaboration with end-users (rural communities and cassava processors). Figure 5 describes the scheme of the United Republic of Tanzanian MAS and PPB project. CMD resistant F_1 generated by MAS at CIAT were crossed to BC_1 derivatives of *M. esculenta* sub spp. *flabellifolia*, showing good resistance to CGM, to produce progenies that combine some CMD and CGM resistance (Kullaya *et al.*, 2004). The progenies were established from embryo axes as *in vitro* plants to aid shipment to Africa. Molecular markers associated with resistance to CMD and phenotypic evaluation for CGM resistance were used to screen and select progenies that combine resistance to CMD and CGM. Resistant plants (300 genotypes and ten plants per genotype), were shipped to the United Republic of Tanzania as *in vitro* plantlets for use as improved parents. A selection based on harvest index, a highly heritable trait, and total biomass was made and 80 genotypes selected. These were planted in the second year in a controlled crossing block together with 54 local germplasm from the eastern and southern zones of the country. Emphasis was placed on local varieties with, or tolerance to, CBSD, which is a major disease of cassava in coastal east Africa and Mozambique. Over 40 000 crosses were made between the improved genotypes and the local varieties producing more than 60 000 seeds.

Sexual seeds obtained from crossing improved and local genotypes were planted in the screen house and transferred to the field 40 days after planting. Parental lines were also planted in the screen house from woody stakes. DNA was isolated from parental lines using the rapid mini-preparation method and evaluated with the five markers associated with the

CMD2 mediated resistance to CMD. Polymorphism in pair-wise combinations of the parental lines was observed with at least one of the five markers and will be used on the progeny. The phenotype of the progeny will be evaluated at three and six months after planting for resistance to CMD, CBSD and CGM. Markers are currently being tested for CGM resistance and are being developed for resistance to CBSD; when their utility is confirmed, they will also be used on progenies.

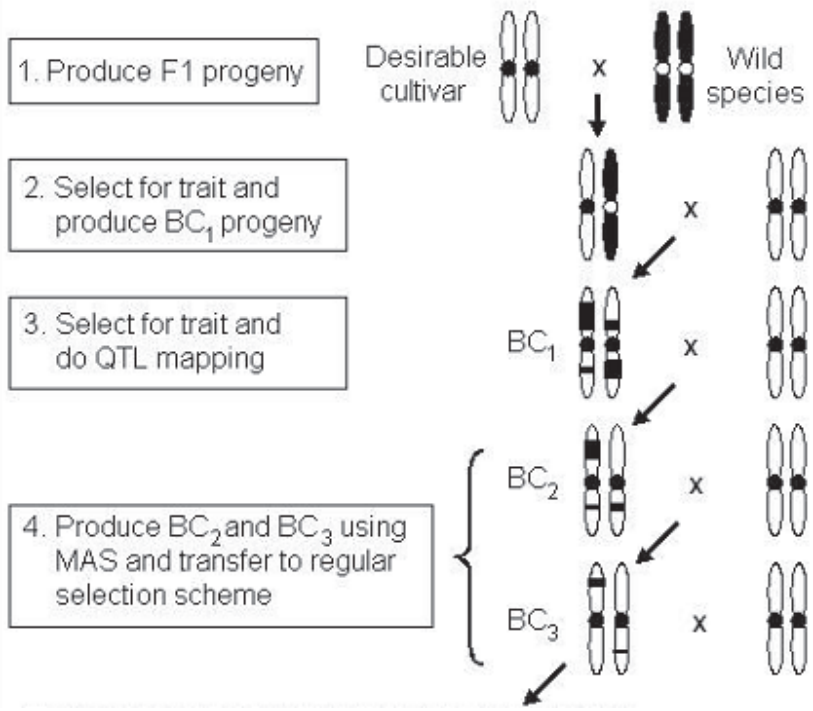
Using published broad sense heritability of 0.6 for CMD resistance (Hahn, Terry and Leuschner, 1980), it is expected that 24 000 symptomless genotypes will be analysed with markers associated with resistance to CMD. The gain of MAS will be the elimination of at least 38 400 (4 800 x 8 plants) that would have been carried to the single row trial stage (eight plant-rows per genotype), considering that breeders traditionally select 20 percent at the seedling trial stage. This represents a reduction of about 4 ha at the CET. If markers can be used to select for resistance to CGM and CBSD, then an additional number of genotypes can be eliminated from the CET leading to even greater savings. Using MAS for CMD alone would reduce the size of field trials by 50 percent. If additional second and third traits were included, reductions could be as high as 75 and 87.5 percent, respectively. Perhaps the most important advantage, however, comes from the increased genetic gain arising from higher heritabilities in these field evaluations with fewer genotypes.

MAS for transferring useful traits from wild relatives of cassava into the cultivated gene pool

Wild *Manihot* germplasm offers a wealth of useful genes for the cultivated *M. esculenta* species, but its use in regular breeding programmes is restricted by linkage drag and the long reproductive breeding cycle. For example, several accessions of *M.esculenta* sub spp. *flabellifolia*, *M. peruviana* and *M. tristis* have high levels of proteins (Nichols, 1947; Asiedu *et al*., 1992; CIAT, 2004). Low amylose content starch (3–5 percent) or waxy starch of relevance to the cassava starch industry has also been identified in two wild relatives of cassava, namely *M. crassisepala* and *M. chlorostricta.* The only source of dramatically delayed post-harvest physiological deterioration (PPD) has been identified in an interspecific hybrid between cassava and *M. walkerae*. The *M. walkerae* parent was collected in Mexico and held at the Washington University, St. Louis, United States of America (Bertram, 1993). It was brought to CIAT in 1998 in an attempt to use it in improving PPD. Furthermore, the only source of resistance to the cassava hornworm and the most widely deployed source of resistance to CMD were identified in fourth backcross generation progenies of *M. glaziovii* (Jennings, 1976; Chavarriaga *et al.,* 2004). Moderate to high levels of resistance to CGM, whiteflies and the cassava mealybug have been found in interspecific hybrids of *M. esculenta* sub spp. *flabellifolia.* The delayed PPD trait and resistance to the pests were successfully transferred to F_1 interspecific hybrids suggesting dominant or additive gene action of the gene(s) involved (CIAT, unpublished data).

The long reproductive cycle and lengthy time required to develop new cassava varieties (10–15 years) often discourages the use of wild species in most conventional cassava breeding programmes. However, the use of molecular markers to introgress a single target region of the genome can

FIGURE 6
Modified advanced backcross QTL scheme in the introgression of useful traits from wild relatives into cassava



save between two to four backcross generations (Frisch *et al.*, 1999). Indeed, it has been shown in several crops that the "tremendous genetic potential" locked up in wild relatives can be released more efficiently through the aid of new tools of molecular genetic maps and the advanced backcross QTL mapping scheme (ABC-QTL) (Tanksley and McCouch, 1997).

For several years now molecular marker tools and a modified ABC-QTL scheme have been tested in cassava at CIAT for the introgression of useful genes from wild relatives. The scheme entails generating BC_1 crosses and carrying out QTL mapping followed by selection of genotypes carrying the genome region of interest with minimum segments of the donor genome (Figure 6). The modified ABC-QTL is currently being used at CIAT to introgress genes for high protein content, waxy starch, delayed PPD, and resistance to whiteflies and the hornworm. The most advanced of these MAS projects is the introgression of high protein content from close wild relatives of cassava. Two BC_1 families of between 250 and 300 progenies were developed from two accessions of *M. esculenta* sub spp. *flabellifolia* OW284-1 and OW231-3, and the improved cassava variety from Thailand Rayong 60 (MTAI 8 in the germplasm collection). The BC_1 families were planted

in a CET for evaluation of root protein content at ten months. The grand parental lines of the BC_1 population were genotyped with over 800 simple sequence repeat (SSR) markers available for cassava and about 300 polymorphic markers were identified. The polymorphic markers are being assayed in the progenies after which QTL analysis will be conducted using the phenotypic protein and molecular marker data. Genotypes that have QTL for protein and a minimum of the donor parent genome will be selected and used for producing the BC_2 generation.

For introgression of naturally occurring mutant granule-bound starch synthase (GBSSI) for waxy starch in wild relatives, a more targeted approach was taken. Sequencing of the glycosyl transferase region of the GBSSI gene from the wild relatives and two cassava accessions identified four single nucleotide polymorphisms (SNPs) that differentiated the wild accessions from cassava. Allele-specific molecular markers unique to these SNPs were developed for selection of these alleles in a breeding scheme.

Genetic crosses were made between *M. chlorosticta* accession CW14-11 and MTAI8, and the resulting F_1 was back-crossed to MTAI8. The allele specific marker will be used together with other agronomic traits, particularly performance, to select for BC_1 that carry the mutant GBSS alleles for self-pollination to recover the waxy trait. The identification of natural mutants in a key gene and development of markers represent an innovative molecular tool to accelerate the introgression of favourable alleles from wild relatives into cassava. Backcross derivatives have also been developed from *M. walkerae* (MWal 001) for delayed post-harvest physiological deterioration; from MNG11 (a BC_4 derivative of *M. glaziovii)* for resistance to hornworm; and from *M. esculenta* sub spp. *flabellifolia* (FLA447-1) for resistance to whiteflies. Phenotypic and genetic mapping of these backcross populations are in progress to be followed by identification of QTL and selection of progenies to generate the next generation. MAS will later be used to combine these genes into progenitors for use as parents in breeding which, together with low cost marker technologies, will be distributed extensively to national programmes in Africa, Asia and Latin America to produce improved varieties.

Marker-assisted estimation of average heterozygosity during inbreeding of cassava

A principal use of molecular markers by private sector breeding companies is to accelerate the development of inbred lines. Cassava genotypes are heterozygous and very little inbreeding has been practised to date. However, inbred lines are better as parents as they do not have the confounding effect of dominance and carry lower levels of genetic load (undesirable alleles). Speed of inbreeding depends upon the average heterozygosity of the original parental lines, the homozygosity level of the selected genotypes at the end of the self-pollinating phase and the process of selection of progenies to be self-pollinated (Scotti *et al.,* 2000). Basically in the inbreeding process two events go together: phenotypically there is a decrease in vigour, which is correlated with the increased levels of homozygosity. While the aim is to select vigorous plants (tolerant to inbreeding), in the process plants may be selected that are less homozgygous than the expected average for their generation. It is expected that the first few cycles of self-pollination will result in a marked reduction of vigour (inbreeding depression associated with the genetic load of the parental lines); therefore, selection for

tolerance to inbreeding depression must be exerted. However, such selection is biased by the differences in homozygosity levels of segregating partially inbred genotypes. This highlights the need for a method to measure the level of heterozygosity in these partially inbred individuals and to use this in a co-variance correction in the selection of phenotypically vigorous genotypes. Molecular markers can be used to estimate the level of homozygosity of a given plant, enabling selection of plants with true tolerance to inbreeding.

Molecular markers can identify regions in the genome that are particularly related to the expression of heterosis and for measuring genetic distances among inbred lines to direct crosses with higher probabilities of high heterosis. Co-dominant SSR markers on a genome-wide basis are suitable for this purpose. The effect of self-pollination on vigour and heterozygosity was analysed in nine S_1 families, heterozygosity being estimated in the S_1 families by 100 mapped SSR markers that cover over 80 percent of the cassava genome and plant vigour by dry root yield and plant biomass. Results will assist in selecting the best performing and least heterozygous plants during inbreeding by identifying superior partially inbred parental lines. Molecular markers could also be used to delineate heterotic groups in cassava. Genetic resources of cassava have been characterized at the regional (Fregene *et al.*, 2003) and global (Hurtado *et al.*, 2005) levels. Highly differentiated groups of accessions were observed particularly among groups of materials from Guatemala and Africa and they may represent heterotic pools. These groupings are being tested based on molecular markers by genetic crossing between and within the groups as a first step to define heterotic patterns for a more systematic improvement of combining ability via recurrent reciprocal selection.

Other potential MAS targets

Several other traits for which MAS can be applied to increase efficiency of breeding include:

Beta-carotene

CIAT and a number of partners are involved in a project to produce cassava varieties with higher levels of β-carotene in yellow roots. This is one way of combating the deficiency of this key micronutrient in areas where cassava is a major staple. The experimental approach to increasing cassava β-carotene content includes conventional breeding and genetic transformation. The discovery of a wide segregation pattern of root colour in two S_1 families from the Colombian landrace MCOL 72 (cross code AM 273) and MTAI 8 (AM 320) was the basis for molecular genetic analysis of β-carotene content in cassava. Three markers, SSRY251, NS980 and SSRY330, were found to be associated with β–carotene content. These are in the same region of the genome and together explain >80 percent of phenotypic variation for β-carotene content in the population used for this study. The homozygous state of certain alleles of these markers translates into higher β-carotene content, suggesting that breeding for this trait can benefit from molecular markers to assist in combining favourable alleles in breeding populations. The work is continuing with the search for additional favourable alleles in yellow-rooted germplasm to give the best possible phenotypic expression of the trait.

Cyanogenic potential

A collaborative project between the Swedish University of Agricultural Sciences (SLU), Uppsala, the Medical Biotechnology

Laboratories (MBL), Kampala, and CIAT, is aimed at the genetic mapping of CNP in cassava. An S_1 family-AM 320, derived from the bitter variety MTAI 8 is the basis for the study. This family has been evaluated for cyanogenic glucoside content and has been genotyped with more than 200 diversity array technology (DarT) markers at CAMBIA, Australia, and 150 SSR markers at CIAT. The discovery of molecular markers for CNP will provide a tool to select efficiently for low cyanogenic potential in cassava. Also ongoing is the genetic mapping of the two cytochrome P450 genes CYP79D1 and D2 that catalyse the rate-limiting step of the biosynthesis of the cyanogenic glucosides, linamarin in the S_1 family AM 320. The group is also looking for an association with QTL for CNP. It is expected that markers associated with CNP will be identified at the end of the study.

Dry matter content

Few key traits in cassava hold greater potential for increasing cost-effectiveness via MAS than root dry matter content (DMC). This trait is usually measured at the end of the growth cycle. A number of genetic and environmental effects influence DMC. It is usually highest before the onset of rains, but drops after the rains begin as the plant mobilizes starch from the roots for re-growth of leaves (Byrne, 1984). Defoliation from pest and disease attacks can lower DMC. Breeding programmes have been quite successful in improving DMC, especially for industrial markets. The entry point for developing markers associated with DMC was recent diallel experiments (Jaramillo *et al.*, 2005; Calle *et al.*, 2005; Pérez *et a*l., 2005a, b; Cach *et al*., 2005b). Diallels, in this case made up of 90 families, are an ideal method to identify genes controlling DMC that are useful in many genetic backgrounds. Estimates of general and specific combining ability (SCA and GCA, respectively) for many traits of agronomic interest were calculated, with emphasis on DMC. Based on GCA estimates, parents were selected to generate larger-sized progenies for DMC mapping. Sizes of families in the original diallel experiment were about 30 progenies, which is rather small for genetic mapping. Parallel to the development of mapping populations was the search for markers associated with DMC using two F_1 families, GM 312 and GM 313, selected from the diallel experiment having parents with high GCA for DMC.

Initial marker analysis using bulked segregant analysis led to the discovery of two molecular genetic markers, SSRY160 and SSRY150, which explain about 30 and 18 percent, respectively, of phenotypic variance for DMC. These markers are being analysed on approximately 700 genotypes derived from 23 crosses with parents having high GCA for DMC in order to confirm their utility across genetic backgrounds. Parallel to this, larger families are being developed from selected parents for QTL mapping of DMC.

Disadvantages of MAS

Perhaps the greatest disadvantage of MAS is the time and financial investment required to develop markers that are widely applicable for traits of agronomic importance. Often a marker developed in one or a few related genotypes will not work for other genotypes in a breeding scheme due to allelic effects. Furthermore, development of markers, particularly for QTL, is complicated by epistatic interactions and the critical need for good quality phenotypic data. Several ways around this

problem have been proposed, such as the use of candidate genes involved in the traits directly as selectable markers without the need for laborious gene tagging experiments. However, unravelling the genetics and the development of markers for such traits is still many years down the road. New methods of association mapping and linkage disequilibrium mapping that rely upon non-random association of candidate genes or markers on a high resolution map with a phenotype of interest in a non-structured collection of genotypes have been used extensively in human medicine to identify genes involved in disease (Cardon and Bell, 2001). Given the enormous difficulties of quantitative mapping in humans and the success of association mapping, these methods have also been proposed as ways around the problems in developing markers for low heritability traits in plants (Gaut and Long, 2003). The development of (partially) inbred cassava genetic stocks will certainly accelerate the application of MAS for the genetic improvement of the crop.

CONCLUSIONS

Given limited resources, further prioritization of traits is needed for the development of markers if they do not already exist. Top priority should be given to MAS for the most important pests and diseases prevalent in the region for which durable sources of resistance genes exist. Priority should also be given to DMC as this is another trait that, although having a high narrow sense heritability at the time of evaluation (usually after the onset of the rains to permit planting immediately thereafter), is significantly affected by non-genetic factors and is not as highly heritable. There are several initiatives to assist national programmes acquire new molecular tools to increase the cost-effectiveness of breeding. Prominent among these are the "molecular breeding communities of practice" project of the Generation Challenge Programme (GCP, www.generationcp.org) and the Rockefeller Foundation-funded African Molecular Marker Network (AMMANET, www.africancrops.net/ammanet). Both have training programmes on molecular breeding that are open to national programme scientists. The CIAT cassava project has also developed a Web-based database resource including protocols, populations, and markers for MAS in cassava that can easily be accessed by national programmes (www.ciat.cgiar.org/mascas).

Cassava and common beans: contrasts

Cassava and beans are similar with respect to the modest level of research input they have enjoyed over the past three to five decades. Both have been part of the research agenda of CIAT and of the CGIAR for nearly thirty years, and especially beans have benefited from inputs from laboratories and programmes in the United States of America and, to a lesser degree, Europe. However, research investments for high-scale genomics through marker development in these crops has been far less than for the "super crops" like maize, rice or soybean that enjoy participation by the private sector, but are more than minor orphan crops with local usage in the tropics.

Yet biologically, these two crops are widely contrasting. Cassava is a perennial versus beans, which are short-season annuals, although climbing beans at high altitudes can be similar to cassava in growth cycle. Beans are an autogamous seed crop while cassava is an allogamous crop with vegetative propagation. Accompanying this latter dichotomy are differences in gene action. Beans present largely additive gene

action, while cassava expresses important components of dominance and epistatic action. Finally, cassava as a clonal crop can fix heterotic combinations, while a lack of genetic male sterility or apomixis systems in common bean have curtailed the development of a hybrid industry for this seed crop even though heterosis is observed.

In spite of their biological and other differences, the results of several years experience with MAS in beans and cassava are surprisingly similar. In both crops, MAS is being employed principally to bolster phenotypic selection for disease resistance genes. Disease resistance is often governed by relatively few genes, and phenotypic data are obtained more easily. On the other hand, MAS for more complex traits has yet to find ready application. While there are candidates for such traits in both crops (root bulking in cassava; low phosphorus or drought tolerance in beans), the complexity of these traits has made the identification of reliable markers more difficult and has delayed application. Obtaining reliable phenotypic data for complex traits is especially difficult and is often the biggest bottleneck to eventual application of MAS. In the case of cassava, no inbred parents have been used to date for the development of molecular markers, making the genetic analysis more difficult.

However, some differences in the application of MAS for the two crops may be noted, arising from the form of reproduction of each crop. The time frame to select cassava clones through multilocational trials is about six to seven years. During this period and with each step the number of genotypes is reduced as a result of the selection exerted, but the genotype of each individual clone remains stable. In the case of beans from the F_1 until stabilization of pure lines there is an intense segregation process in the early generations which tapers off in later generations. In both crops MAS can be used in the early stages of the selection process but with different objectives. In cassava, MAS can help to select early on the clone that will ultimately be released, whereas in beans MAS is used to "direct" the segregation process in the more desirable direction. Although maps with significant saturation are available for both crops, these have been constructed over several years, employing genotypes (in the case of beans) from different gene pools with wide polymorphism. A small proportion of these markers (often 20–30 percent) is polymorphic in other hybrid combinations among the genotypes within the same gene pool or race that have been created to tag a specific trait. Thus, genome coverage is often still not optimal for the high quality QTL analysis that is usually needed for complex traits.

RECOMMENDATIONS

Careful prioritization of traits, marker system and genetic stocks for MAS

The limited resources available for cassava or bean research require a judicious allocation of efforts. In the past 10–20 years there has been increased investment in molecular marker research in both crops. However, a considerable proportion of that research was directed at demonstrating the usefulness of different techniques, e.g. RAPD, restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), etc. Over this period there has been an ever-changing set of technologies but relatively little actual benefit derived from their application. There is a trade-off between being on the cutting edge with the newest technologies and "sticking it out" with an "outdated" technology

until some benefit is extracted from it. On the other hand, far too much effort has been expended in the identification of markers for traits without carrying these through to application. Often gene tagging is a component of a short-term project, and does not receive the necessary follow up in implementation. In each case, the essential question is: what are the key genes for each crop? And once defined, which genes merit the investment to develop molecular markers? For investments in molecular marker development to yield results, it is important that traits are chosen for which molecular breeding has both a clear advantage over field-based selection and is feasible in the short to medium term. It is also important that emphasis be given to selecting the relevant crosses, pedigrees and populations in which to practise MAS, and to have in place appropriate phenotyping strategies for the confirmation of MAS results. In this regard, the use of parental surveys of many of the genotypes involved in a given breeding programme is an important first step in implementing MAS.

Short- and long-term research related to MAS

The present research structure that is normally based on short-term projects, usually of three years' duration, can seldom be expected to deliver results of usable markers for complex traits. Such short-term projects that seek to establish the basis for MAS or to implement selection should limit their objectives to simply inherited traits. On the other hand, longer-term funding either of a programmatic or successive project funding nature, must be obtained to address more complex traits governed by QTL as these would normally require at least two phases of three-year projects. The first phase might be expected to reveal the inheritance of a given trait, establishing the location and numbers of QTL, while a second phase would be required to validate these over more environments and to find markers that are polymorphic over a wide number of genotypes and therefore widely useful for breeding, as well as adapted to rapid laboratory techniques. A medium- to long-term investment likewise implies careful prioritization of such traits, with regard to potential impact and the eventual need for MAS. These reflections are based upon presently available laboratory techniques, but as techniques for more detailed and widespread evaluation of loci and genotypes are developed (e.g. gene chips for analysis of multiple loci), conclusions could change significantly.

Scaling-up technologies

After the development of molecular markers for a trait and their initial implementation, a period of scaling-up in use of the specific markers is necessary. Sometimes this involves changes to MAS protocols, in the marker detection technique or in the markers themselves. Marker re-design has been a common element of scaling-up exercises and can involve something as simple as changing a PCR fragment size to implementing a SNP assay for the actual sequence differences between alleles. Technologies that speed up the implementation process and lower the costs associated with scaling-up are crucial to the success of MAS and are often neglected.

Development of markers that are useful in a large number of crosses

Often a marker developed for a particular trait in one or a few related genotypes will not work for other genotypes with high value of the trait due to differences in gene or allelic effects. Unravelling the genetics of

major traits of agronomic interest even in a subset of elite parents used for breeding is beyond the resources available for bean and cassava research. Association mapping and linkage disequilibrium mapping, which rely upon non-random association of candidate genes or markers on a high resolution map with a phenotype of interest in a non-structured collection of genotypes, have been proposed as a way around this problem. Association mapping can be used to discover new marker-trait associations or to validate associations that were found through conventional genetic mapping. The GCP is facilitating association mapping of traits of agronomic importance in cassava and beans with the goal of discovering more useful markers for a wider range of genotypes.

The need to strike a balance between MAS and field-based selection

Occasionally the question is raised: which is better, MAS or conventional selection? This very question betrays a false dichotomy that hinders progress. By itself, MAS is seldom an adequate selection tool and therefore must be combined with conventional phenotypic selection. The objective should be to develop the optimal balance between conventional and molecular breeding, and the "best" balance will be unique to each situation, crop, selection scheme, environment and opportunities for different selection methods. More emphasis is needed on combined selection systems, rather than viewing MAS as a replacement for phenotypic or field selection.

REFERENCES

Afanador, L.K. & Hadley, S.D. 1993. Adoption of a mini-prep DNA extraction method for RAPD marker analysis in common bean. *Bean Improv. Coop*. 35: 10–11.

Akano, A.O., Dixon, A.G.O., Mba, C., Barrera, E. & Fregene, M. 2002. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105: 521–525.

Asiedu, R., Hahn, S.K., Bai, K.V. & Dixon, A.G.O. 1992. Introgression of genes from wild relatives into cassava. *In* M.O. Akoroda & O.B. Arene, eds. *Proc. 4th. Triennial Symp. Internat. Soc. Trop. Root Crops-Africa Branch,* pp. 89-91. Nigeria, ISTRC-AB/IDRC/CTA/IITA.

Awale, H.E. &. Kelly, J.D. 2001. Development of SCAR markers linked to Co-42 gene in common bean. *Ann. Rep. Bean Improv. Coop*. 44: 119–120.

Babu, L. & Chatterjee, S.R. 1999. Protein content and amino acid composition of cassava tubers and leaves. *J. Root Crops* .25: 163–168.

Beaver, J.S., Rosas, J.C., Myers, J., Acosta, J., Kelly, J.D., Nchimbi-Msolla, S.M., Misangu, R., Bokosi, J., Temple, S., Aranud-Santana, E. & Coyne, D.P. 2003. Contributions of the bean/cowpea CRSP to cultivar and germplasm development in common bean. *Field Crops Res*. 82: 87–102.

Beebe, S., Velasco, A. & Pedraza, F. 1999. Marcaje de genes para rendimiento en condiciones de alto y bajo fósforo en las accesiones de frijol G21212 y BAT 881. *VI Reunião Nacional de Pesquisa de Feijão*. Salvador, Brazil.

Beebe, S., Skroch, P.W., Tohme, J., Duque, M.C., Pedraza, F. & Nienhuis, J. 2000. Structure of genetic diversity among common bean landraces of Mesoamerican origin based on correspondence analysis of RAPD. *Crop Sci.* 40: 264–273.

Beebe, S., Rengifo, J., Gaitan, E., Duque, M.C. & Tohme, J. 2001. Diversity and origin of Andean landraces of common bean. *Crop Sci.* 41: 854–862.

Beebe, S.E., Rojas, M., Yan, X., Blair, M.W., Pedraza, F., Muñoz, F., Tohme, J. & Lynch, J.P. 2006. Quantitative trait loci for root architecture traits correlated with phosphorus acquisition in common bean. *Crop Sci.* 46: 413–423.

Bellotti, A.C. 2002. Arthropod pests. *In* R.J. Hillocks, J.M. Thresh & A.C. Bellotti, eds. *Cassava: biology, production and utilization,* pp. 209-235. Wallingford, UK, CABI Publishing.

Bellotti, A.C., Arias B., Vargas, V.O., Reyes H.J.A. & Guerrero, J.M. 2002. Insectos y ácaros dañinos a la yuca y su control. *In* B. Ospina & H. Ceballos, eds. *La yuca en el tercer milenio. Sistemas modernos de producción, procesamiento, utilización y comercialización,* CIAT Publication No. 327, pp. 160-203. Cali, Colombia.

Bertram, R.B. 1993. *Application of molecular techniques resources of cassava (*Manihot esculenta *Crantz, Euphorfiabeae): interspecific evolutionary relationships and intraspecific characterization.* University of Maryland, USA. (Ph.D. thesis).

Blair, M.W. & Beaver, J.S. 1993. Inheritance of bean golden mosaic resistance from bean genotype A 429. *Ann. Rep. Bean Improv. Coop.* 36: 143–144.

Blair, M.W., Iriarte, G. & Beebe, S.E. 2003. QTL analysis of an Andean advanced backcross population for yield traits derived from wild *P. vulgaris*. *Ann. Rep. Bean Improv. Coop.* 46: 7–9.

Blair, M.W., Pedraza, F., Buendía, H.F., Gaitán-Solís, E., Beebe, S.E., Gepts, P. & Tohme, J. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.) *Theor. Appl. Genet*. 107: 1362–1374.

Blair, M.W., Beaver, J.S., Nin, J.C., Prophete, E. & Singh, S.P. 2006. Registration of PR9745–232 and RMC-3 red-mottled dry bean germplasm lines with resistance to bean golden yellow mosaic virus. *Crop Sci.* 46: 1000–1002.

Broughton, W.J., Hernandez, G., Blair, M.W., Beebe, S.E., Gepts, P. & Vanderleyden, J. 2003. Beans (*Phaseolus* spp.)–model food legumes. *Plant & Soil* 252: 55–128.

Buendia, H.F., Beebe, S.E., Blair, M.W., Teran, H. & Pedraza, F. 2003. Identificación de marcadores moleculares asociados a genes de rendimiento en una población RC2F3.7 de frijol común *Phaseolus vulgaris* L. (DOR390 x G19892). *Fitotecnia Colombiana* 3: 57–64.

Buitrago, A.J.A. 1990. *La yuca en la alimentación animal*, p. 446. Cali, Colombia, CIAT.

Byrne, D. 1984. Breeding cassava*. Plant Breeding Rev.* 2: 73–134.

Cach, N.T., Lenis, J.I., Perez, J.C., Morante, N., Calle, F. & Ceballos, H. 2005a. Inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) for sub-humid conditions. *Plant Breed.* 125(2): 177–182.

Cach, N.T., Perez, J.C., Lenis, J.I., Calle, F., Morante, N. & Ceballos, H. 2005b. Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for sub-humid conditions. *J. Heredity* 96: 586–592.

Calle, F., Perez, J.C. Gaitán, W., Morante, N., Ceballos, H., Llano, G. & Alvarez, E. 2005. Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soil savannas. *Euphytica* 144: 177–186.

Cardon, L.R. & Bell, J.I. 2001. Association study design for complex disease. *Nature Revs. Genet*. 2: 91–99.

Carvalho, G.A., Paular, T.J., Alzate-Marin, A.L., Nietsche, S., Barros, E.G. & Moreira. M.A. 1998. Herança da resistência da linhagem AND 277 de feijoeiro-comun à raça 63-23 de *Phaeoisariopsis griseola* e identificação de marcador RAPD ligado ao gene de resistência. *Fitopatol. Bras*. 23: 482–485.

Ceballos, H., Iglesias, C.A., Pérez, J.C. & Dixon, A.G.O. 2004. Cassava breeding: opportunities and challenges. *Plant Mol. Biol*. 56: 503–515.

Ceballos, H., Fregene, M., Lentini, Z.J., Sanchez, T., Puentes, Y.I., Perez, J.C., Rosero, A. & Tofiño, A.P. 2006. Development and identification of high-value cassava clones. *Acta Hort.* 703:63–70.

Chavarriaga, P., Prieto, S., Herrera, C.J., López, D., Bellotti, A. & Tohme, J. 2004. Screening transgenics unveils apparent resistance to hornworm (*E. ello*) in the non-transgenic, African cassava clone 60444. *In A.* Alves & J. Tohme, eds. Adding value to a small-farmer crop. *Proc. 6th. Internat. Scientific Mtg. Cassava Biotech. Network,* Book of Abstracts, p.4. Cali, Colombia, CIAT.

Chávez, A.L., Sánchez, T., Jaramillo, G., Bedoya, J.M., Echeverry, J., Bolaños, E.A., Ceballos, H. & Iglesias, C.A. 2005. Variation of quality traits in cassava roots evaluated in landraces and improved clones. *Euphytica* 143: 125–133.

CIAT (International Center for Tropical Agriculture). 1997. Assessing and enhancing agrobiodiversity through biotechnology. *Annual Report*, 1994-1997, pp. 63–65. Cali, Colombia.

CIAT. 2001. Bean improvement for sustainable productivity, input use efficiency and poverty alleviation. *Annual Report*, pp. 37–38. Cali, Colombia.

CIAT. 2002. Bean improvement for sustainable productivity, input use efficiency and poverty alleviation. *Annual Report*, pp. 104–107. Cali, Colombia.

CIAT. 2003. Bean improvement for sustainable productivity, input use efficiency and poverty alleviation.. *Annual Report*, pp. 32–33 and 67–75. Cali, Colombia.

CIAT. 2004. Bean improvement for sustainable productivity, input use efficiency and poverty alleviation, *Annual Report*, pp. 84 and 115–117. Cali, Colombia.

FAO. 2001. *Food Outlook*. December 1999. Rome.

Ferreira, C.F., Borém, A., Carvalho, G.A., Nietsche, S., Paula, T.J., de Barros, E.G. & Moreira, M.A. 2000. Inheritance of angular leaf spot resistance in common bean and identification of a RAPD marker linked to a resistance gene. *Crop Sci*. 40: 1130–1133.

Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., Tohme, J., & Bonierbale, M. 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 95: 431–441.

Fregene, M., Suarez, M., Mkumbira, J., Kulembeka, H., Ndedya, E., Kulaya, A., Mitchel, S., Gullberg, U., Rosling, H., Dixon, A. & Kresovich, S. 2003. Simple sequence repeat (SSR) diversity of cassava (*Manihot esculenta* Crantz) landraces: genetic structure in a predominantly asexually propagated *Crop Theor. Appl. Genet*. 107: 1083–1093.

Frei, A., Blair, M.W., Cardona, C., Beebe, S.E., Gu, H. & Dorn, S. 2005. Identification of quantitative trait loci for resistance to *Thrips palmi* Karny in common bean (*Phaseolus vulgaris* L.). *Crop Sci.* 45: 379–387.

Frisch, M., Bohn, M. & Melchinger, A.E. 1999. Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci*. 39: 1295–1301.

Gaut, B.S. & Long, A.D. 2003. The lowdown on linkage disequilibrum. *Plant Cell* 15: 1502–1506.

Gepts, P. 1988. A middle American and an Andean common bean gene pool. *In* P. Gepts, ed. *Genetic resources of* Phaseolus *beans,* pp. 375-390. Dordrecht, Netherlands, Kluwer Academic Publishers.

Gimelfarb, A. & Lande, R. 1994. Simulation of marker-assisted selection in hybrid populations. *Genetical Res*. 63: 39–47.

Hahn, S.K., Howland, A.K. & Terry, E.R. 1980. Correlated resistance to cassava to mosaic and bacterial blight diseases. *Euphytica* 29: 305–311.

Hahn, S.K., Terry, E.R. & Leuschner, K. 1980. Breeding cassava for resistance to cassava mosaic disease. *Euphytica* 29: 673–683.

Haley, S.D., Afanador, L. & Kelly, J.D. 1994. Identification and application of a random amplified polymorphic DNA marker for the I gene (potyvirus resistance) in common bean. *Phytopathology* 84: 157–160.

Hallauer, A.R. & Miranda Fo., J.B. 1988. *Quantitative genetics in maize breeding, second edition.* USA, Iowa State University Press.

Hershey, C.H. 1984. Breeding cassava for adaptation to stress conditions: development of a methodology, pp. 303-314. *Proc. 6th Internat. Symp. Soc. Trop. Root Crops,* Lima, Peru.

Hillocks, R.J. & Wydra, K. 2002. Bacterial, fungal and nematode diseases. *In* R.J. Hillocks, J.M. Thresh & A.C. Bellotti, eds. *Cassava: biology, production and utilization,* pp. 261–280. USA and UK, CABI Publishing.

Hospital, F. & Charcosset, M. 1997. Marker-assisted introgression of quantitative loci. *Genetics* 147: 1469–1485.

Hospital, F, Chevalet, C. & Mulsant, P. 1992. Using markers in gene introgression breeding programs. *Genetics* 132: 1119–1210.

Hurtado L., P.X., Ospina, C., Marín, J., Buitrago R.Ch., Fregene, M.A., De Vicente, M.C., Wongtiem, P., Killian, A. & Wenzl, P. 2005. Comparación de dos marcadores moleculares: Microarreglos-microsatélites para el análisis de la diversidad estructural de yuca Manihot esculenta Crantz [resumen] [CD-ROM]. In *Congreso Asociación Colombiana de Fitomejoramiento y Producción de Cultivos* (9, 2005, Palmira, Colombia). [Resúmenes] [CD-ROM]. Corporación Colombiana de Investigación Agropecuaria (CORPOICA), Cali, CO. p. 138.

Ikeda, N., Bautista, N.S., Yamada, T., Kamijima, O. & Ishii, T. 2001. Ultra-simple DNA extraction method for marker-assisted selection using microsatellite markers in rice. *Plant Mol. Bio. Reporter* 19: 27–32.

Islam, F.M.A., Basford, K.E., Redden, R.J., Gonzalez, A.V., Kroonenberg, P.M. & Beebe, S. 2002. Genetic variability in cultivated common bean beyond the two major gene pools. *Genetic Resources and Crop Evolution* 49(3): 271–283.

Jaramillo, G., Morante, N., Pérez, J.C., Calle, F., Ceballos, H., Arias, B. & Bellotti, A. C. 2005. Diallel analysis in cassava adapted to the midaltitude valleys environments. *Crop Sci.* 45: 1058–1063.

Jennings, D.L. 1976. Breeding for resistance to African cassava mosaic. African cassava mosaic. *Report Interdiscipl. Workshop* IDRC 071e, pp. 39–44. Muguga, Kenya.

Jennings, D.L. & Iglesias, C. 2002. Breeding for crop improvement, pp. 149–166. *In* R.J. Hillocks, J.M. Thresh & A.C. Bellotti, eds. *Cassava: biology, production and utilization*. UK, CABI Publishing.

Johnson, W.C., Guzman, P., Mandala, D., Mkandawire, A.B.C., Temple, S., Gilbertson, R.L. & Gepts, P. 1997. Molecular tagging of the *bc*-3 gene for introgression into Andean common bean. *Crop Sci.* 37: 248–254.

Jung, G., Skroch, P.W., Coyne, D.P., Nienhuis, J., Aranud-Santana E., Ariyarathne, H.M., Kaeppler, S.M. & Bassett, M. 1997. Molecular marker-based genetic analysis of tepary bean-derived common bacterial blight resistance in different development stages of common bean. *J. Am.. Soc. Hort. Sci*. 122: 329–337.

Kawano, K. 1980. Cassava. *In* W.R. Fehr & H.H. Hadley, eds. *Hybridization of crop plants,* pp 225-233. Madison, WI, USA, ASA and CSSA.

Kawano, K. 2003. Thirty years of cassava breeding for productivity- biological and social factors for success. *Crop Sci.* 43: 1325–1335.

Kawano K., Narintaraporn, K., Narintaraporn, P., Sarakarn, S., Limsila, A., Limsila, J., Suparhan, D., Sarawat, V. & Watananonta, W. 1998. Yield improvement in a multistage breeding program for cassava. *Crop Sci.* 38: 325–332.

Kelly, J.D. & Adams, M.W. 1987. Phenotypic recurrent selection in ideotype breeding of pinto beans. *Euphytica* 36: 69–80.

Kelly, J.D., Schneider, K.A. & Kolkman, J.M. 1999. Breeding to improve yield. *In* S.P. Singh, ed. *Common bean improvement in the twenty-first century. Developments in plant breeding*, Vol. 7, pp. 185–222. Dordrecht, Netherlands, Kluwer Academic Publishers.

Kelly, J.D, Gepts, P., Miklas, P.N. & Coyne, D.P. 2003. Tagging and mapping of genes and QTL and molecular-marker assisted selection for traits of economic importance in bean and cowpea. *Field Crops Res*. 82: 135–154.

Klimyuk, V.I., Carroll, B.J., Thomas, C.M. & Jones, J.D.G. 1993. Alkali treatment for rapid preparation of plant material for reliable PCR analysis. *Plant J.* 3: 493–494.

Kullaya, A., Mtunda, K., Kulembeka, H., Ferguson, M., Marin, J., Ospina, C., Barrera, E., Jarvis, A., Morante, M., Ceballos, H., Tohme, J. & Fregene, M. 2004. Molecular marker-assisted and farmer participatory improvement of cassava germplasm for farmer/market preferred traits in Tanzania. *6th. Internat. Scientific Mtg. Cassava Biotech. Network.* Cali, Colombia, CIAT.

Legg, J.P. & Ogwal, S. 1998. Changes in the incidence of African cassava mosaic virus disease and the abundance of its whitefly vector along south-north transects in Uganda. *J. Appl. Entom*. 122: 169–178.

Liao, H., Yan, X., Rubio, G., Beebe, S.E., Blair, M.W. & Lynch, J.P. 2004. Basal root gravitropism and phosphorus acquisition efficiency in common bean. *Funct. Plant Biol*. 31: 1–12.

Mahuku, G., Montoya, C., Henríquez, M.A., Jara, C., Terán, H. & Beebe, S. 2004. Inheritance and characterization of angular leaf spot resistance gene in common bean accession G 10474 and identification of an AFLP marker linked to the resistance gene. *Crop Sci.* 44: 1817–1824.

Mba, R.E.C., Stephenson, P., Edwards, K., Melzer, S., Nkumbira, J., Gullberg, U., Apel, K., Gale, M., Tohme, J. & Fregene, M. 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards a SSR-based molecular genetic map of cassava**.** *Theor. Appl. Genet.* 102: 21–31.

Melotto, M., Afanador, L. & Kelly, J.D. 1996. Development of a SCAR marker linked to the *I* gene in common bean. *Genome* 39: 1216–1219.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

Miklas, P.N., Larsen, R.C., Riley, R. & Kelly, J.D. 2000a. Potential marker-assisted selection for *bc-1*^{2} resistance to bean common mosaic potyvirus in common bean. *Euphytica* 116(3): 211–219.

Miklas, P.N., Smith, J.R., Riley, R., Grafton, K.F., Singh, S.P., Jung, G. & Coyne, D.P. 2000b. Marker-assisted breeding for pyramided resistance to common bacterial blight in common bean. *Ann. Rep. Bean Improv. Coop*. 43: 39–40.

Miklas, P.N., Stone, V., Daly, M.J., Stavely, J.R., Steadman, J.R., Bassett, M.J., Delorme, R. & Beaver, J.S. 2000c. Bacterial, fungal, and viral disease resistance loci mapped in a recombinant inbred common bean population ('Dorado'/XAN 176). *J. Am. Soc. Hort. Sci.* 125: 476–481.

Miklas, P.N., Kelly, J.D., Beebe, S.E. & Blair, M.W. 2006. Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. *Euphytica* 147: 105–131.

Muñoz, L.C., Blair, M.W., Duque, M.C., Roca, W. & Tohme, J. 2004. Level of introgression in interspecific (*Phaseolus vulgaris* x *P. acutifolius*) congruity-backcross lines. *Crop Sci.* 44: 637–645.

Nichols, R.F.W. 1947. Breeding cassava for virus resistance. *East Afr. Agric. J.* 12: 184–194.

Nietsche, S., Borém, A., Carvalho, G.A., Rocha, C., Paula, T.J., de Barros, E.G. & Moreira, M.A. 2000. RAPD and SCAR markers linked to a gene coferring resistance to angular leaf spot in common bean. *J. Phytopathol.* 148: 117–121.

Nobuyuki, I., Bautista, N.S., Yamada, T., Kamijima, O. & Ishi, T. 2000. Ultra simple DNA extraction method for marker-assisted selection using microsatellite markers in rice. *Plant Mol. Biol. Reporter* 18: 1–6.

Okogbenin, E., Porto, M.C.M & Dixon, A.G.O. 1998. Influence of planting season on incidence and severity of African cassava mosaic disease in the subhumic zone of Nigeria. *In* I.J. Ekanayake, ed. *Root crops and poverty alleviation*, pp. 338–392. Ibadan, Nigeria, IITA.

Park, S.O., Coyne, D.P., Mutlu, N., Jung, G. & Steadman, L.R. 1999. Confirmation of molecular markers and flower color associated with QTL for resistance to common bacterial blight in common beans. *J. Amer. Soc. Hort. Sci.* 124: 519–526.

Pastor-Corrales M.A., Jara, C. & Singh, S. 1998. Pathogenic variation in, source of, and breeding for resistance to *Phaeoisariopsis griseola* causing angular leaf spot in common bean. *Euphytica* 103: 161–171.

Pérez, J.C., Ceballos, H., Jaramillo, G., Morante, N., Calle, F., Arias, B. & Bellotti, A.C. 2005a. Epistasis in cassava (*Manihot esculenta* Crantz) adapted to mid-altitude valley environment. *Crop Sci.* 45: 1491–1496.

Pérez, J.C., Ceballos, H., Calle, F., Morante, N., Gaitán, W., Llano, G. & Alvarez, E. 2005b. Within-family genetic variation and epistasis in cassava (*Manihot esculenta* Crantz) adapted to the acid-soils environment. *Euphytica* 145: 77–85.

Polston, J.E. & Anderson, P.K. 1997. The emergence of whitefly transmitted geminiviruses in tomato in the western hemisphere. *Plant Disease* 81: 1357–1369.

Rajendran, P.G., Ravindran, C.S., Nair, S.G. & Nayar, T.V.R. 2000. True cassava seeds (TCS) for rapid spread of the crop in non-traditional areas. Thiruvananthapuram, Kerala, India, Central Tuber Crops Research Institute. 22 pp.

Santana, G.E., Blair, M.W., Morales, F., Mahuku, G., Jara, C. & Castaño, M. 2004. Uso de tecnicas clasicas y avanzadas para identificar genotipos de frijol resistentes a antracnosis y mosaico común. *Fitotecnia Colombiana* 4: 44–54.

Sartorato, A., Nietsche, S., Barros, E.G. & Moreira, M.A. 1999. SCAR marker linked to angular leaf spot resistance gene in common bean. *Bean Improv. Coop.* 42: 23–24.

Schneider, K.A., Brothers, M.E. & Kelly, J.D. 1997. Marker-assisted selection to improve drought tolerance in common bean. *Crop Sci.* 37: 51–60.

Scotti, C., Pupilli, F., Salvi, S. & Arcioni, S. 2000. Variation in Vigot and in RFLP-estimated heterozygosity by selfing tetraploid alfalfa: new perspectives for the use of selfing in alfalfa breeding. *Theor. Appl. Genet.* 101: 120–125.

Shull, G.F. 1952. Beginnings of the heterosis concept. *In* J.W. Gowen, ed. *Heterosis,* pp. 14–48. Ames, IA, USA, Iowa State College Press.

Singh, S.P. 1999. Integrated genetic improvement. *In* S.P. Singh, ed. *Common bean improvement in the twenty-first century. Developments in plant breeding*, Vol. 7, pp. 133–166. Dordrecht, Netherlands, Kluwer Academic Publishers.

Singh, S.P., Gepts, P. &. Debouck, D.G. 1991. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot*. 45: 379–396.

Singh, S.P., Cardona, C., Morales, F.J., Pastor-Corrales, M.A. & Voysest, O. 1998. Gamete selection for upright carioca bean with resistance to five diseases and a leafhopper. *Crop Sci.* 38: 666–672.

Stam, P., & Zeven, A.C. 1981. The theoretical proportion of donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* 30: 227–238.

Storey, H.H. & Nichols, R.F.W. 1938. Studies on the mosaic disease of cassava. *Ann. Appl. Bio*. 25: 709–806.

Sullivan, J.G. & Bliss, F.A. 1983. Expression of enhanced seed protein content in inbred backcross lines of common bean. *J. Am. Soc. Hort. Sci.* 108: 787–791.

Swanson, M.M. & Harrison, B.D. 1994. Properties, relationships and distribution of cassava mosaic geminiviruses. *Trop. Sci.* 34: 15–25.

Tanksley, S.D. & McCouch, S.R. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063–1066.

Urrea, C.A., Miklas, P.N., Beaver, J.S. & Riley. R.H. 1996. A codominant randomly amplified polymorphic DNA (RAPD) marker useful for indirect selection of BGMV resistance in common bean. *J. Am. Soc. Hort. Sci*. 121: 1035–1039.

Vallejo, V. & Kelly, J.D. 2001. Development of a SCAR marker linked to Co-5 gene in common bean. *Ann. Rept. Bean Improv. Coop*. 44: 121–122.

Van Schoonhoven, A. & Pastor-Corrales, M.A. 1987. Standard system for the evaluation of bean germplasm. Cali, Colombia, CIAT.

Yan, X., Liao, H., Beebe, S.E., Blair, M.W. & Lynch, J.P. 2005. Molecular mapping of QTLs associated with root hairs and acid exudation as related to phosphorus uptake in common bean. *Plant & Soil* 265: 17–29.

Young, R.A., Melotto, M., Nodari, R.O. & Kelly, J.D. 1998. Marker assisted dissection of the oligogenic anthracnose resistance in the common bean cultivar G2333. *Theor. Appl. Genet*. 96: 87–94.

CHAPTER 8

Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors

Michel Ragot and Michael Lee

SUMMARY

More than twenty-five years after the advent of DNA markers, marker-assisted selection (MAS) has become a routine component of some private maize breeding programmes. Line conversion has been one of the most productive applications of MAS in maize breeding, reducing time to market and resulting in countless numbers of commercial products. Recently, applications of MAS for forward breeding have been shown to increase significantly the rate of genetic gain when compared with conventional breeding. Costs associated with MAS are still very high. Further improvements in marker technologies, data handling and analysis, phenotyping and nursery operations are needed to realize the full benefits of MAS for private maize breeding programmes and to allow the transfer of proven approaches and protocols to public breeding programmes in developing countries.

INTRODUCTION

The ability to identify genetic components of traits, particularly quantitative traits, in Mendelian factors, and to monitor or direct their changes during breeding through the use of DNA-based markers has created much enthusiasm. Claims were sometimes made that marker-assisted selection (MAS) would rapidly replace phenotypic selection and dramatically reduce the time required to develop commercial varieties (Mazur, 1995). At the turn of this century, phenotypic selection was still the approach on which maize breeding programmes mostly relied to develop new and improved cultivars while MAS had contributed to advances in introgression, or backcross breeding (Ragot *et al.,* 1995; Ho, McCouch and Smith, 2002; Ribaut, Jiang and Hoisington, 2002; Morris *et al.,* 2003). Overly optimistic statements and exaggerated promises about the power of MAS to improve complex traits created excessively high and largely unfulfilled hopes and prompted a wave of cautious and sometimes pessimistic views (Melchinger, Utz and Schön, 1998; Young, 1999; Goodman and Carson, 2000; Bernardo, 2001).

Recently, multinational corporations with large maize breeding programmes reported the routine and successful use of MAS (Johnson, 2004; Niebur *et al.,* 2004; Eathington, 2005; Crosbie *et al.,* 2006). Rates of genetic gain twice as high as those achieved through conventional breeding were reported for MAS in maize. Accounts were also given of a number of MAS-derived single-cross (i.e. simple) hybrids being currently on the market. Although too little is known about the methods (e.g. breeding schemes, mathematical algorithms) and tools (e.g. marker technologies, computer programs, databases) used to develop these hybrids, these results have raised confidence in the ability of MAS to increase the rate of genetic gain over what can be achieved through conventional breeding. As technologies evolve and marker genotypes become less expensive, MAS becomes increasingly within the reach of developing countries. Whenever necessary, transfer of methods or tools from private companies to developing countries should be made possible while preserving the commercial interests of the companies concerned, thereby contributing to increasing the rate of genetic gain where it is most needed.

Much has happened in maize breeding since Stuber and Moll (1972) first reported that selection for grain yield in maize had resulted in changes in allele frequencies at several isozyme loci throughout the genome. In so doing, they essentially laid the grounds for MAS in maize. Indeed, if phenotypic selection could produce a change in marker allele frequencies, then why could deliberately altering marker allele frequencies at specific loci not produce predictable phenotypic changes for one or several traits?

The objectives of this chapter are to provide the scientific community and decision-makers with information on the current status of MAS in maize breeding programmes, including the major steps that led to it, and to provide suggestions to developing countries for deploying the technology and methods involved in an efficient, cost-effective and realistic manner.

HOW HAS MAS BEEN USED BY THE PRIVATE SECTOR TO IMPROVE THE MAIZE CROP?

Applications of DNA markers in private maize breeding programmes started in the 1980s with the identification of DNA clones used to detect restriction fragment length polymorphisms (RFLPs)

in the nuclear genome. As described below, the methods used to detect RFLPs were incompatible with the magnitude, speed and efficiency of all but a few aspects of selection in maize breeding programmes. Gradually, however, the methods used to detect DNA polymorphisms and to create meaningful information from DNA marker and phenotypic data sets have evolved to the point where they are routine components of some maize breeding programmes in the private sector.

Selection occurs at various stages in maize breeding programmes. The first opportunity arises when choosing inbred lines to mate as parents of new populations. In some programmes, all such inbreds are genotyped systematically at DNA marker loci (Smith and Smith, 1992). If the marker loci are sufficiently close on genetic or physical maps then reasonably good inferences may be made about the inbred's haplotype. Such information is used to establish identity, resolve disagreements related to germplasm ownership and acquisition, enforce laws intended to encourage genetic diversity of the hybrids and avoid using inbreds that contain transgenes which may violate regulatory considerations and restrictions. These selection practices, while admittedly not conventional MAS, have led to improvements in the maize crop by enabling more informed stewardship and deployment of genetic resources and by providing a degree of protection of intellectual property and related investments in maize breeding.

Unquestionably, the most pervasive and direct use of MAS in maize by the private sector has been with backcrossing of transgenes into elite inbred lines, the direct parents of the commercial hybrids (Ragot *et al.,* 1995; Crosbie *et al.,* 2006). Currently, the most widely deployed transgenes and combinations thereof (i.e. gene stacks) are for resistance to herbicides or insects (e.g. *Ostrinia* and *Diabrotica*). As the commercial maize crop of any region, maturity zone, market or country is not yet uniform or homogeneous for any transgene, maize breeders have elected to develop near-isogenic versions (transgenic and non-transgenic) of elite inbreds and commercial hybrids in order to satisfy combinations of licensing agreements, agronomic practices, regulatory requirements, market demands and product development schemes. This has required companies to have two parallel maize breeding programmes, transgenic and non-transgenic. In this manner, marker-assisted backcrossing (MABC) of transgenes, and to a lesser degree, of native genes and quantitative trait loci (QTL) for other traits, has expedited the development of commercial hybrids.

More recently, marker-assisted recurrent selection (MARS) schemes and infrastructure have been developed for "forward breeding" of native genes and QTL for relatively complex traits such as disease resistance, abiotic stress tolerance and grain yield (Ribaut and Betrán, 1999; Ragot *et al.,* 2000; Ribaut, Jiang and Hoisington, 2000; Eathington, 2005; Crosbie *et al.,* 2006). Simulation studies suggested that MAS could be effective for such traits under certain conditions (Edwards and Page, 1994; Gimelfarb and Lande, 1994), but the initial empirical attempts at such selection were not successful (Stromberg, Dudley and Rufener, 1994; Openshaw and Frascaroli, 1997; Holland, 2004; Moreau, Charcosset and Gallais, 2004) except in the special case of sweetcorn (Edwards and Johnson, 1994; Yousef and Juvik, 2001). The success reported for sweetcorn is due to the fact that the genetic base of sweetcorn is extremely narrow relative to dent or flint maize; thus

predicted gains and extrapolations across populations are more reliable. Also, phenotypic analyses of many traits in a sweetcorn breeding programme are extremely expensive because they involve processing large volumes of grain; therefore, MAS would be relatively inexpensive and effective under such circumstances. However, subsequent developments in technology, refinements in analytical methods and improvements in experimental designs have been assembled into a process that has shown promise for some reference populations of dent maize (Ragot *et al.,* 2000; Johnson, 2004; Crosbie *et al.,* 2006) as improvement in grain yield from MAS often exceeded that from non-MAS approaches. Presumably, such results will lead to the development of new and superior inbred lines and commercial hybrids in a cost-effective manner. While the impact of such MAS has not yet been fully realized in the maize crop, the methods have been employed to various degrees by programmes in the private sector that have the necessary infrastructure.

The potential for MAS to contribute to improvements in the maize crop should increase in parallel with our understanding of the relationships among genomes, the environment and phenotypes. Candidate transgenes will be developed on a regular basis and their contributions to maize improvement will be realized in the most efficient manner with MAS. Likewise, the identification of candidate native genes and their gene products and functions, and of other DNA sequences (e.g. miRNA, matrix attachment and regulatory regions), will improve the power of methods such as association mapping and genome scans to assess their genotypic value in the context of defined reference populations of significance to maize breeding (Thornsberry *et al.,* 2001; Rafalski, 2002; Niebur *et al.,* 2004; Varshney, Graner and Sorrels, 2005). Beyond its use in MARS schemes, this information might make it reasonable to reconsider ideas such as methods for predicting hybrid performance that may have been limited by the amount and type of information and by the design of the experiment when they were initially evaluated (Bernardo, 1994).

METHODOLOGY AND DESIGN OF BREEDING PROGRAMMES SUPPORTED BY MAS

As expected, private sector maize programmes focus entirely on inbred-hybrid breeding schemes intended to develop elite inbred lines that enable the profitable production of commercial F_1 hybrids. To a large extent, MAS breeding programmes use the same designs and methods known to maize breeders for decades and generic descriptions of these have been published (Hallauer and Miranda, 1981; Sprague and Dudley, 1988; Bernardo, 2002). When MAS is included in the breeding programme, the significant differences are, of course, the availability of genotypic data at different stages of selection and some knowledge of the relationships between the genotypic and phenotypic data sets for the reference population(s) in the target environment(s).

In contrast to conventional breeding schemes, the methods and design of infrastructure needed to support MAS have been the areas of greatest change. In order to utilize MAS, companies had to make significant investments to assemble or modify various aspects of infrastructure such as methods to detect DNA polymorphism, manage information, or analyse and track samples, software to relate genotype with phenotype, and off-season or continuous nurseries. These components had to be integrated with each

other and with breeding activities, which meant that scientists needed to learn how and when MAS provided a comparative advantage over other methods.

MAS: enabling methods, tools and infrastructure

Perhaps the component of infrastructure in greatest need of development was related to the acquisition of genotypic data (i.e. DNA markers). Although the concept of associating markers with quantitative traits was not new (Sax, 1923), the discovery reported by Stuber and Moll (1972) was very significant. Stuber and Moll (1972) described for the first time associations between molecular markers and quantitative traits while previous associations had been based on morphological markers (Sax, 1923). The advantages of molecular over morphological markers soon became obvious and detailed descriptions of these advantages were published by Tanksley *et al.* (1989) and Stuber (1992).

Two of these advantages are of particular importance. First, molecular marker genotypes can usually be obtained from any plant tissue, even from young seedlings or kernels, while morphological markers frequently require the observation of whole, mature plants. Selection can therefore occur earlier in the plant's cycle when using molecular markers than when using morphological markers. The ability to conduct early selection, possibly before flowering, can have a tremendous impact on the rate of genetic gain of a breeding programme and therefore constitutes a very significant advantage of molecular over morphological markers.

Second, molecular markers are neutral markers. They are not affected by environmental or growing conditions. They are not affected by the genetic background either, nor do they affect phenotypes. The expression of morphological traits, by contrast, can be dependent on environmental or growing conditions. In addition, epistasic interactions are often observed among morphological marker loci or between morphological marker loci and the genetic background. These epistatic interactions prevent distinguishing all genotypes associated with morphological markers and further limit the number of morphological markers that can be studied simultaneously.

Although isozyme markers had many advantages over morphological markers, the lack of a sufficient number of polymorphic loci limited their use for MAS (Goodman *et al.,* 1980). Nevertheless, isozyme markers are still used for quality control during seed production.

RFLPs (Botstein *et al.,* 1980) are based on DNA polymorphisms detected through restriction nuclease digestions followed by DNA blot hybridizations. The abundance and high level of polymorphism of RFLPs, especially in maize, allowed the construction of extensive maize genetic maps (Helentjaris *et al.,* 1986; Burr *et al.,* 1988; Hoisington, 1989; Coe *et al.,* 1995; Davis *et al.,* 1999) as well as the identification and mapping of many QTL.

Being robust, reproducible and codominant, RFLPs are perfectly suited for genetic studies as well as for MAS applications. Their two main disadvantages are the large quantities of DNA required, and the difficulty to miniaturize and automate. Nevertheless, RFLPs were quickly adopted and represented the marker system of choice for many plant species including maize throughout the 1980s and during much of the 1990s.

The development of the polymerase chain reaction (PCR) (Saiki *et al.,* 1988) turned out to be a major breakthrough in

molecular marker technology. PCR-based markers require little DNA, allowing sampling of young seedlings and very early selection and thereby optimization of breeding schemes. PCR-based marker protocols are very amenable to automation and miniaturization and improvements to protocols resulted in considerable reductions in both cost and time required to produce data points. The first two PCR-based marker systems were random amplified polymorphic DNA (RAPDs), and amplified fragment length polymorphisms (AFLPs). Detailed descriptions and critical assessments of these two systems can be found in Welsh and McClelland (1990), Williams *et al.* (1990), Penner *et al.* (1993), Ragot and Hoisington (1993), Skroch and Nienhuis (1995) and Jones *et al.* (1997) for RAPDs, and in Vos *et al.* (1995), Jones *et al.* (1997) and Castiglioni *et al.* (1999) for AFLPs. They are also described in other chapters of this book.

Simple sequence repeats (SSRs) or microsatellites rapidly became the marker of choice in maize, almost entirely displacing RFLPs and previously developed PCR-based marker systems. Polymorphism of SSRs is due to variable numbers of short tandem repeats, often two or three base pairs in length and usually flanked by unique regions (Tautz, 1989). SSRs are very reproducible (Jones *et al.,* 1997) and co-dominant (Shattuck-Eidens *et al.,* 1990; Senior and Heun, 1993; Senior *et al.,* 1996) and are therefore very suitable for maize MAS applications.

Many additional variations of PCR-based marker systems have been developed and a thorough review can be found in Mohan *et al.* (1997).

All the DNA-based marker systems described to date are gel-based systems, a major constraint for automation. Single nucleotide polymorphisms (SNPs) (Lindblad-Toh *et al.,* 2000) can be revealed in many ways including allele-specific PCR, primer extension approaches, or DNA chips, all of which are not gel-based. SNPs can generally be scored as co-dominant markers, except in the case of insertion-deletion polymorphisms. Although allelic diversity at SNPs is usually limited to two alleles, this limitation can be offset by the abundance of SNPs and the analysis of haplotypes, combinations of genotypes at several neighbouring SNPs. Haplotype analyses increase informativeness (Ching *et al.,* 2002), although at some expense because two to four SNPs have to be genotyped where one SSR sufficed. SNP genotyping can be highly miniaturized and automated, thereby reducing the cost and allowing the production of very large numbers of data points. With genetic maps containing several thousand mapped SNPs, these have become the marker of choice for private maize MAS programmes.

DNA marker technology has been a dynamic and often expensive component of the infrastructure needed for MAS. For example, one corporation indicated having spent tens of millions of United States dollars to develop an automated system for detecting RAPDs, a technology that was never suited for MAS in a large maize breeding programme. Later, another corporation spent an even greater amount of money to acquire technology for matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) analysis of amplified DNA fragments. These technologies were either rapidly replaced or never used. Such decisions would have bankrupted most national maize programmes or a couple of centres belonging to the Consultative Group on International Agricultural Research (CGIAR). Fortunately, this area of infrastructure has matured somewhat

and become more stable so that start-up and operating costs, while still high for some programmes, are more predictable.

Statistical methods and related software have also been areas of significant development, especially for the detection and description of putative QTL. QTL, which are nothing more than associations between markers and traits, were first described using simple association tests between trait values and marker genotypes (Stuber and Moll, 1972). These tests consider each marker locus independently and neither require nor take advantage of the existence of genetic maps. Statistical methods have been developed that take advantage of the existence of genetic maps (see review by Manly and Olson, 1999). These statistical methods, simple interval mapping (Lander and Botstein, 1989) and composite interval mapping (Jansen, 1993; Zeng, 1993, 1994), test the existence of associations between hypothetical marker genotypes and trait values at several points in intervals between pairs of adjacent marker loci on the genetic map, allowing the positioning of QTL on these genetic maps. All of the previous methods are based on single QTL models. Other statistical methods have been developed that simultaneously test the presence of several QTL in the genome (Kao, Zeng and Teasdale, 1999).

Many software packages are available for QTL mapping and based on one or several of the statistical methods developed to date. No two packages are exactly alike and all have specific strengths and weaknesses with respect to particular situations, making it sometimes beneficial to use more than one package to perform QTL mapping analyses. The software packages most commonly used for QTL mapping in maize include QTL Cartographer (Basten, Weir and Zeng, 1994), MapQTL (van Ooijen and Maliepaard, 1996), and PLABQTL (Utz and Melchinger, 1996). All of these only handle bi-allelic populations, while MCQTL (Jourjon *et al.,* 2005) also performs QTL mapping in multi-allelic situations, including bi-parental populations made from segregating parents, or sets of bi-parental, bi-allelic populations.

More recently, methods based on Bayesian analysis (Jansen, Jannink and Beavis, 2003; Gelman *et al.,* 2004) and association (Varshney, Graner and Sorrels, 2005) or *in silico* mapping (Parisseaux and Bernardo, 2004) have been proposed as more powerful and refined approaches to assess the relationships between genotype and phenotype that are needed for MAS. Methods of Bayesian analysis should be less affected by the uncertainties of QTL effects and locations and produce better estimates of those parameters in MAS. Association mapping approaches are particularly useful to validate the relevance of genes and alleles in specific germplasm such as that used by maize breeders. *In silico* mapping takes advantage of the pedigree relationships among individuals to structure the population used to establish marker-trait associations. This approach, which is highly complex due to the population structure resulting from pedigree breeding, is particularly appropriate for maize where data across many years and environments are available for large sets of related individuals. Certainly, as the annotation of genomes gradually improves, such methods will be common components of breeding programmes. Currently, the applications of methods such as association mapping for MAS are hindered by the fact that a very low percentage of the genes in crop plants have a function assigned to them on the basis of direct experimentation. However, this impoverished situation

is being enriched through a variety of projects on functional genomics.

In sharp contrast to the many methods and software packages developed for QTL identification and mapping, little has been published for MAS. This paucity of information on MAS tools most likely reflects both the low level of activity in the public sector and the fully proprietary nature of developments in the private sector.

In parallel with advancements in DNA technology and statistical methods, private sector programmes have enhanced the capabilities and capacities of their continuous nurseries. Such nurseries have been used for decades by programmes in both the private and public sectors. In order to conduct MAS to its greatest advantage, continuous nurseries had to be managed, equipped and staffed in new ways so that the plants complete their life cycle as quickly as possible and that the genotypic data (and sometimes some phenotypic data) needed for MAS may be collected at each sexual generation. Three to four sexual generations per year may be completed at such nurseries.

These activities and the continuous collection of both genotypic and phenotypic data in the target environment and their integrated analyses create huge data sets that must be analysed quickly and related to other large extant data sets. Data management and bioinformatics for breeders have therefore become critical components of the infrastructure needed to use MAS. Prior to the advent of MAS, some large private breeding programmes had established a group of dedicated data managers to assist with research and marketing, and with the arrival of genomics and MAS the need for such dedicated specialists has increased greatly.

Once the basic infrastructure had been established to complement the activities of maize breeders, programmes were ready to implement several basic aspects of MAS; many of which are derived from well established methods and principles of maize breeding.

MAS-based breeding

Selection occurs at various stages in maize breeding programmes. The first opportunity for selection is the choice of inbred lines to mate as parents of new populations. Prior to the advent of DNA marker data, the selection of such parents would be based on a combination of phenotypic assessments, pedigree information, breeding records and chance (Hallauer and Miranda, 1981; Sprague and Dudley, 1988). In some programmes today, all such inbreds are genotyped systematically at DNA marker loci. Depending on the resources and objectives, the degree of genotyping may range from a low density of marker loci (e.g. SNPs in candidate genes) to higher density whole genome scans (Varshney, Graner and Sorrels, 2005). These genotypic data, alone or integrated with phenotypic information, may reveal novel aspects of maize gene pools, heterotic groups, haplotype evolution, gene content and parents used in MAS for specific target environments (Fu and Dooner, 2002; Niebur *et al.,* 2004; Crosbie *et al.,* 2006). When properly integrated with phenotypic information and functional genomics, genotypic data of inbred lines should allow breeders to choose parents that, when mated, should provide populations or gene pools enriched for the more desirable combinations of favourable alleles. Such a starting point is a huge advantage in plant breeding because it increases the probability of selecting progeny that are superior to the parents and that approximate a predicted optimum genotype.

MABC is certainly the form of MAS with the most immediate and obvious benefits for maize breeding. MABC is used for three main purposes: selection of transgenes (or of native DNA sequences of the maize genome, whether genes or QTL), elimination of unwanted regions of the donor-parent genome linked to the transgene and selection of unlinked regions of the recurrent-parent genome. With the exception of DNA markers and transgenes, these have been the same goals of backcross breeding since the inception of that method decades ago (Fehr, 1987). Of course, DNA markers enable breeders to identify progeny that contain the desired recombinant chromosomes and donor-parent genome in a more direct manner. Also, MABC facilitates the process of combining more than one transgene in a given inbred line (e.g. "gene or trait stacking or pyramiding"). This reduces the number of generations needed to reach certain stages of a breeding programme and reduces the time needed to produce commercial hybrids for the market. Generic MABC schemes suitable for maize breeding programmes have been described in detail for single genes (Hospital, Chevalet and Mulsant, 1992; Ragot *et al.,* 1995; Frisch, Bohn and Melchinger, 1999a, 1999b; Frisch and Melchinger, 2001a; Hospital, 2001; Ribaut, Jiang and Hoisington, 2002), for QTL (Hospital and Charcosset, 1997; Bouchez *et al.,* 2002) and for gene stacks (Frisch and Melchinger, 2001b). Versions of such schemes have been used in maize breeding programmes in the private sector, often at their continuous nurseries (Ragot *et al.,* 1995). Most recently, MABC has also been adopted as a tool to develop sets of near-isogenic lines (NILs) for genomics research (Peleman and van der Voort, 2003).

Theoretical and simulation studies have been conducted to identify the most efficient MABC protocols. Parameters most commonly studied include the number of individuals genotyped at each generation, the number of markers used, relative selection pressure for recombination around the target locus or global recovery of recurrent parent genome and the number of individuals selected at any generation. Optimal values for each of the above depend on the objective of the MABC approach in terms of quality (required level of recurrent parent genome recovery), speed (fastest possible conversion or set number of generations) and resources (unlimited or limited). While the fastest and highest quality MABC approaches have the most expensive protocols, less intensive approaches can result in significant time savings and quality improvements when compared with conventional backcrossing approaches and at a fraction of the cost of the most expensive MABC protocols.

Frisch, Bohn and Melchinger (1999b) showed that to minimize linkage drag around the target locus (loci), selection of recombination events close to the target locus (loci) should be conducted in the early backcross generations. Frisch and Melchinger (2001a) and Ribaut, Jiang and Hoisington (2002) further demonstrated that minimizing linkage drag around the target locus requires very large numbers of individuals (possibly hundreds) to be genotyped. Hospital and Charcosset (1997) proposed a selection scheme based on selecting a single individual to be backcrossed. By contrast, Frisch and Melchinger (2001a) proposed selecting several individuals and determining the family size of their backcross progeny based on the individuals' genotypes. By using varying rather than constant numbers of individuals or markers at the different backcross generations, it was shown that the number

of marker data points required could be reduced and thus the efficiency of MABC improved (Hospital, Chevalet and Mulsant, 1992; Frisch, Bohn and Melchinger, 1999b). Several studies also showed that using a limited number of markers on non-carrier chromosomes was sufficient to recover more that 95 percent of the recurrent parent genome in three or fewer backcross generations (Hospital, Chevalet and Mulsant, 1992; Visscher, Haley and Thompson, 1996; Servin and Hospital, 2002).

One of the most important lessons from the various theoretical and simulation studies of MABC is that the effects of the different MABC parameters are not independent of each other. With maize, large backcross populations can be generated from a single plant when that plant is used as the male and recurrent parent plants are used as females. Marker systems in maize are also such that very large amounts of marker data can be generated on plants before flowering. Potential MABC protocols are almost endless in maize and identifying the most efficient is only possible on a case-by-case basis. For example, while achieving almost complete recovery of the recurrent parent's genome is necessary for registering backcross-derived lines and hybrids in many European countries, partial recovery might be sufficient to improve the agronomic performance of varieties in developing countries. The optimal MABC protocols for these two strikingly different objectives will be very different. Protocols for the first objective will involve background selection and the use of background markers very close to the target locus (loci). Protocols for the second objective might involve markers for the target locus (loci) only, while relying on successive backcross generations to recover an adequate amount of recurrent parent genome.

Successful examples of MABC in maize include backcrossing of transgenes (Ragot *et al.,* 1995), and QTL for insect resistance (Willcox *et al.,* 2002), flowering maturity (Ragot *et al.,* 2000; Bouchez *et al.,* 2002) and grain yield (Ho, McCouch and Smith, 2002).

Methods of "forward breeding" with DNA markers have also been proposed and implemented by maize breeding programmes. As with the pedigree-based methods of maize breeding favoured by the private sector, many of the "new" methods that utilize genetic data from DNA markers integrated with phenotypic data are essentially a form of recurrent selection, a method that has been in use for several decades (Hallauer and Miranda, 1981). The key advantages of the new versions of recurrent selection are, of course, the availability of genetic data for all progeny at each generation of selection, the integration of genotypic and phenotypic data, and the rapid cycling of generations of selection and information-directed matings at continuous nurseries.

At least two distinct forms of forward breeding with MAS have been proposed: single large-scale MAS (SLS-MAS) (Ribaut and Betrán, 1999) and MARS (Edwards and Johnson, 1994; Lee, 1995; Stam, 1995). A key difference between the methods is that SLS-MAS employs DNA markers at only one generation and attempts to retain genetic variation in regions of the genome unlinked to the DNA markers, while MARS uses markers at each generation, exhausting genetic variation in most regions of the genome. Versions of both SLS-MAS and MARS have been used by breeding programmes in the private sector (Johnson, 2004; Eathington, 2005; Crosbie *et al.,* 2006).

SLS-MAS is of particular interest in pedigree breeding as it consists of screening

and selecting individuals at a few loci at early generations, usually F_2 or F_3, (Eathington, Dudley and Rufener, 1997), using large populations (Ribaut and Betrán, 1999). Individuals displaying homozygous favourable genotypes at the loci of interest are selected and self-pollinated while others are discarded. Self-pollinated progeny of the selected plants then proceed normally through subsequent steps of pedigree breeding. Screening large populations is necessary to ensure that genetic diversity is maintained at regions not under genotypic selection, thereby allowing further phenotypic selection to be conducted. Loci at which marker selection operates can be QTL as described by Ribaut and Betrán (1999). SLS-MAS is thus limited by issues such as the precision of the QTL parameters (position, effect), and relevance of the QTL across environments or gene pools. SLS-MAS can also be conducted for genes, eliminating many of the limitations pertaining to QTL. Although a powerful approach adopted in several species (barley, soybean, sunflower, wheat) to enrich breeding populations at a few loci (Crosbie *et al.,* 2006), SLS-MAS does not appear to have been widely implemented in maize breeding programmes.

MARS targets all traits of importance in a breeding programme and for which genetic information can be obtained. Genetic information is usually obtained from QTL analyses performed on experimental populations and comes in the form of maps of QTL with their corresponding effects. If the QTL mapping analysis is conducted based on a bi-parental population, the sign of the effect at each QTL indicates which of the two parents carried the favourable allele at that QTL. As both parents often contribute favourable alleles, the ideal genotype is a mosaic of chromosomal segments from the two parents. This assumes that the goal is to obtain individuals with as many accumulated favourable alleles as possible, a different goal from that of marker-assisted population improvement as studied elsewhere (Lande and Thompson, 1990; Gimelfarb and Lande, 1994; Gallais, Dillmann and Hospital, 1997; Hospital, Chevalet and Mulsant, 1997; Knapp, 1998; Moreau *et al.,* 1998; Xie and Xu, 1998). Population improvement schemes are generally based on the random mating of selected individuals while the scheme proposed here is based on directed recombination between specific individuals. As reported by Stam (1995), the ideal genotype, defined as the mosaic of favourable chromosomal segments from two parents, will usually never occur in any F_n population of realistic size. It is, however, possible to design a breeding scheme to produce or approach this ideal genotype based on individuals of the experimental population. This breeding scheme could involve several successive generations of crossing individuals (Stam, 1995; Peleman and van der Voort, 2003) and would therefore constitute what is referred to as MARS or genotype construction. This idea can be extended to situations where favourable alleles come from more than two parents (Stam, 1995; Peleman and van der Voort, 2003).

Van Berloo and Stam (1998, 2001) and Charmet *et al.* (1999) developed computer simulations around this idea and assessed the relative merits of marker-assisted genotype construction over phenotypic selec-tion. MARS was simulated in an experimental population where QTL had been mapped. Index (genetic) values were computed for each individual based on its genotypes at QTL-flanking markers (van Berloo and Stam, 1998, 2001). All simulation studies of MARS found that

it was generally superior to phenotypic selection in accumulating favourable alleles in one individual (van Berloo and Stam, 1998, 2001; Charmet *et al.,* 1999). MARS appeared to take better advantage of the genetic diversity present in the populations to which it was applied than phenotypic selection. Simulation research conducted by van Berloo and Stam (2001) showed that MARS was between 3 and almost 20 percent more efficient than phenotypic selection. The advantage of MARS over phenotypic selection was greater when the population under selection was larger or more heterozygous (BC_1s or F_2s vs. RILs, recombinant inbred lines, or DHs, doubled haploids). Although van Berloo and Stam (2001) limited their simulations to populations of up to 200 individuals, their results seem to indicate that the relative advantage of marker-assisted over phenotypic selection would keep increasing as population size increased. The same simulation studies showed that the advantage of marker-assisted over phenotypic selection was larger when dominant QTL were involved in the selection index, or when trait heritability was low in the case of selection for a single trait (van Berloo and Stam, 1998, 2001). These latter observations are of little relevance to most commercial maize breeding programmes, the goal of which is generally the development of inbred lines improved for several traits that will be later combined into superior hybrid varieties. They should, however, increase the appeal of MARS approaches for breeding programmes aimed at developing open-pollinated varieties.

Simulations have also addressed the impact of the amount and quality of QTL information on selection efficiency. Simulation and empirical studies (Beavis, 1994, 1999) showed that QTL mapping experiments based on segregating populations of less than 500 individuals generally revealed only a subset of all QTL affecting the complex traits segregating in these populations. Quantitative trait loci information used in subsequent MARS was therefore necessarily incomplete. Van Berloo and Stam (2001) showed that the relative advantage of MARS over phenotypic selection decreased rapidly when the fraction of the total genotypic variance explained by the QTL included in the selection index decreased. By contrast (van Berloo and Stam, 1998; Charmet *et al.,* 1999), the efficiency of MARS seems to be rather robust to the well-documented (Lee, 1995) uncertainty of QTL genetic locations. The use of genotypic information at markers flanking the QTL possibly explains this observation.

The cost efficiency of MARS was also investigated through simulation (Moreau *et al.,* 2000; Xie and Xu, 1998). When simulating selection for a single trait, Moreau *et al.* (2000) found that, irrespective of the heritability of the trait, MARS was always more cost efficient than phenotypic selection if the cost of genotyping was less than that of evaluating one individual in one plot. When simulating simultaneous selection for multiple traits, Xie and Xu (1998) found that MARS was more cost efficient than phenotypic selection if the cost of genotyping was less than that of phenotyping one individual for all traits. These studies were based on a single generation of MARS. Also, they did not take into consideration any factors besides genotyping and phenotyping costs, although factors influencing the length of a selection cycle or the number of cycles that can be completed in a year can obviously affect the relative economic merits of marker-assisted and phenotypic selection.

In contrast to the abundance of QTL mapping reports, very few accounts of MARS experiments are found in the literature. Moreau, Charcosset and Gallais (2004) compared phenotypic, marker-only, and combined recurrent selection for grain yield and grain moisture at harvest over several cycles and years in maize. Combined selection was based both on phenotypic and marker information while marker-only selection was based on marker information only. Both the marker-only and the combined selection methods constitute MARS approaches. Several combinations of these three methods of selection were applied to the segregating population that served to map the QTL used in marker-based selection indices. Over the six years of the experiment, two cycles of phenotypic selection, two cycles of combined selection, one cycle of combined selection followed by two cycles of marker-only selection, and one cycle of marker-only selection were conducted in parallel. A reassessment of the positions and effects of QTL was conducted after the first cycle for the three schemes containing multiple cycles. All MARS methods were more efficient than phenotypic selection to increase the frequency of favourable alleles at QTL. Nevertheless, Moreau, Charcosset and Gallais (2004) reported no significant difference between marker-assisted and phenotypic selection on the multitrait performance index, although all MARS methods resulted in genetic gain for both grain yield and grain moisture while phenotypic selection resulted in genetic gain for grain yield but an unfavourable evolution of grain moisture. This disappointing result was tentatively explained by the high heritability of the traits, favourable to phenotypic selection, while the percentage of total phenotypic variance explained by the QTL detected for both traits was only about 50 percent. One very encouraging result of this experiment, although Moreau, Charcosset and Gallais (2004) failed to present it as such, was that the first cycle of marker-only selection was as efficient as phenotypic or combined selection in delivering genetic gain. Two conclusions can be drawn from this observation. First, the QTL identified in the initial experimental population were in general not artefacts. Second, selection pressure applied at these QTL, and aimed at fixing alleles identified as favourable, resulted in a change in performance of the selected population in the desired direction when compared with the initial population.

A similar experiment, although based solely on marker-only recurrent selection, was reported by Openshaw and Frascaroli (1997). They conducted MARS in maize simultaneously for four traits, for each of which about ten QTL had been identified. They showed that genetic gain had been achieved in the first cycle of MARS, but that later cycles did not result in any gain. Possible explanations given for these results included uncertainties about QTL parameters (location and effect), interaction effects (epistasis, genetic x environment interaction), and the fact that selection was based on single markers rather than chromosomal segments (Openshaw and Frascaroli, 1997).

Recent communications from several private MARS research programmes (Ragot *et al.*, 2000; Eathington, 2005; Crosbie *et al.*, 2006) revealed large-scale successful applications in maize. Accounts were given of commercial maize hybrids for which at least one of the parental lines was derived through MARS. Eathington (2005) and Crosbie *et al.* (2006) reported that the rates of genetic gain achieved through MARS were about twice those

of phenotypic selection in some reference populations. Marker-only recurrent selection schemes have been implemented for a variety of traits including grain yield and grain moisture (Eathington, 2005), or abiotic stress tolerance (Ragot *et al.,* 2000), and multiple traits are being targeted simultaneously. Selection indices were apparently based on 10 to probably more than 50 loci, these being either QTL identified in the experimental population where MARS was being initiated, QTL identified in other populations, or genes. Marker genotypes are generated for all markers flanking QTL included in the selection indices (Ragot *et al.,* 2000). Plants are genotyped at each cycle and specific combinations of plants are selected for crossing, as proposed by van Berloo and Stam (1998). Several, probably three to four, cycles or MARS are conducted per year using continuous nurseries. In maize, early versions of such schemes have been tested and implemented (Johnson, 2004; Crosbie *et al.,* 2006).

Results reported in these recent communications about private MARS experiments (Ragot *et al.,* 2000; Eathington, 2005) are in sharp contrast to those in earlier publications (Openshaw and Frascaroli, 1997; Moreau, Charcosset and Gallais, 2004). Several factors can explain these discrepancies:

- *Size of the populations submitted to selection at each cycle*. Given reports that increasing population size should result in higher genetic gain through MARS (van Berloo and Stam, 2001,) it is likely that populations submitted to selection in private programmes are rather large, larger than the 160 and 300 individuals reported respectively by Openshaw and Frascaroli (1997) and Moreau, Charcosset and Gallais (2004).
- *Use of flanking versus single markers*. The use of flanking markers for QTL under selection allows better prediction of the genotype at the QTL than when using single markers. When single markers are used, recombination events that occur between the marker and the QTL lead to loss of linkage between the marker and the QTL much faster than when flanking markers are used, thereby rapidly reducing the predictive power of the single marker.
- *Early selection, pre-flowering*. The ability to select plants before flowering ensures optimal mating schemes as the genotypes of plants being selfed or intercrossed are fully known. However, this is not the case when selection cannot take place before flowering and involves intercrossing selfed progenies of selected plants, the genotypes of which might have drifted significantly from those of their genotyped parents.
- *Number of generations per year*. To the authors' knowledge, none of the simulation or experimental studies of MARS has assessed the effects of cycle length on its efficiency despite its direct relationship to the rate of genetic gain. In maize, cycle length can be reduced three- to six-fold when using marker-only recurrent selection compared with phenotypic recurrent selection. Consequently, marker-only recurrent selection will be superior to phenotypic selection as soon as the genetic gain achieved through one cycle of MARS is, respectively, more than a third or a sixth of that achieved through one cycle of phenotypic selection. Private maize breeding programmes have access to off-season nurseries. Furthermore, they have often established efficient continuous nurseries where three to four generations of maize can be grown per year. The use of such nurseries allows them to carry MARS continuously, i.e.

with up to four cycles per year, whereas phenotypic recurrent selection is limited to one cycle per year at most. The impact on the rate of genetic gain of such an implementation of MARS might be very positive even if MARS did not present any advantage over phenotypic selection on a per-cycle basis.

- *Cost of marker data points*. Large private companies have made considerable efforts to reduce both the cost of marker data points and the cost of experimental field plots. The ratio of cost of marker data point to cost of experimental field plot is most likely lower in large private breeding programmes than in most public research laboratories or small private programmes, potentially leading to different views on the economic efficiency of MARS.

Marker-based and phenotypic selection can be mobilized in many different ways, with respect to each other, in marker-assisted breeding schemes. Marker and phenotypic information can be used either simultaneously or sequentially. Selection of parents for breeding populations can be made using marker information alone, phenotypic information alone, or a combination of each. Selection of individuals in a backcross programme can be made on the sole basis of either marker or phenotypic information, or using both. Advancement of individuals in a line development programme can also be made at each generation on the basis of either marker information only, phenotypic information only, or a combination of each. In order to maximize the rate of genetic gain it is likely that MAS breeding schemes such as MABC and MARS will involve generations of marker-only selection conducted at continuous nurseries. The advent of improved methods of producing doubled haploids will certainly further influence the way marker-based and phenotypic selection are mobilized with respect to each other.

In spite of the development of marker-only selection and regardless of the underlying technology and breeding scheme, high-quality phenotyping remains vital and without substitute at several stages; but it may become more focused. Phenotypic evaluation remains the ultimate screen before any cultivar is released. MAS-derived lines and hybrids that meet phenotypic requirements are selected for further evaluation and selection on the basis of their phenotypic value, while those that do not are discarded. Phenotypic evaluation is also critical to establish marker-trait associations or perform the candidate gene validations required to conduct MAS. Here, high quality phenotyping is necessary. Phenotyping protocols will therefore likely be different from those commonly used for phenotypic selection. Experiments may be conducted that involve side-by-side comparisons of different treatments such as water stress or nitrogen fertilization levels to dissect complex traits into their components and facilitate the elucidation of their genetic basis.

Enhancements of such approaches to maize breeding will be based on the incorporation of improved methods of producing doubled haploid inbred lines, information from functional genomics and by learning how to incorporate favourable native genetic variation systematically after MAS has reduced the genetic variation in the original reference populations to unacceptable levels.

ADVANTAGES AND LIMITATIONS OF MAS IN MAIZE BREEDING PROGRAMMES

Advantages of MAS

For private breeding programmes, MAS has offered several attractive features, most

of which are related to time and resource allocations.

MABC clearly provides the information needed to reduce the number of generations of backcrossing, to combine (i.e. "stack") transgenes, "native" genes or QTL into one inbred or hybrid quickly, and to maximize the recovery of the recurrent parent's genome in the backcross-derived progeny. In several private breeding programmes, MABC has enabled the number of backcrossing generations needed to recover 99 percent of the recurrent parent genome to be reduced from six to three, reducing the time needed to develop a converted variety by one year (Crosbie *et al.,* 2006; Ragot *et al.,* 1995). As a line derived by MABC can be made to be very similar to the original non-converted line, most of its attributes, including agronomic performance, can be assumed to be equal or similar to those of the original line. Only limited phenotyping is therefore necessary to verify these assumptions, compared with the extensive multiyear phenotyping required when backcrossing is conducted without markers. One or two years can be saved with MABC during post-conversion phenotyping when compared with conventional backcrossing, resulting in an overall time advantage of MABC over conventional backcrossing of up to three years.

In many situations, the greatest advantages and profits are realized by those who are first to the market with their products. Also, for reasons related to the practices of seed production or legal aspects of crop registration procedures, it may be quite important to be able to produce near-isogenic versions of inbreds and hybrids; MABC provides such ability at a higher probability.

By contrast with MABC, SLS-MAS and MARS do not necessarily decrease the time needed to develop inbred lines. The use of MARS might actually increase it. The advantage of SLS-MAS and MARS resides in their ability to increase the rate of genetic gain (Eathington, 2005), which potentially results in higher performing lines and hybrids than can be developed through phenotypic selection only. Both SLS-MAS and MARS increase the frequency of favourable alleles in the population of selected individuals. The difference between the two approaches is that SLS-MAS operates on few loci while MARS operates on many. When SLS-MAS or MARS are used, the effective size of the population on which selection operates is increased, either directly for SLS-MAS or indirectly through several consecutive generations for MARS when compared with phenotypic pedigree selection. This increase in effective population size permits the application of a greater selection intensity and hence produces a higher genetic gain. SLS-MAS and MARS can also be seen as pre-selection steps if conducted prior to phenotypic selection and therefore improve the chances of evaluating genotypes with a higher frequency of favourable alleles phenotypically because the truly undesirable portion of the population may have been eliminated prior to phenotyping. Phenotypic selection can therefore be conducted with higher selection intensity than would be possible if no pre-selection had taken place, resulting potentially in additional genetic gain.

Alternatively, the resources used for phenotyping can be allocated differently based on whether individuals have been pre-selected or not with MAS. MAS schemes for forward breeding should enable breeding programmes to reallocate or focus resources for phenotypic evaluation in the target environment. For example, if DNA markers are linked to genes for resistance

to a disease or insect then it should be possible initially to select resistant progeny in the absence of the disease or insect by using the DNA data at continuous nurseries. The selected progeny could then be evaluated using relatively more expensive bioassays with the pest(s) in the target environment. This shift in resources is inherent to MARS schemes for complex traits (Edwards and Johnson, 1994; Johnson, 2004; Crosbie *et al.,* 2006). By enriching populations through rapid cycles of MARS at continuous nurseries, breeders should derive a higher frequency of progeny with favourable alleles and haplotypes that are then evaluated in the target environment. Without MARS, resources for evaluation in the target environment would be diluted by the inclusion of too many progeny with an undesirable genetic constitution.

Concerns about reduced genetic diversity among commercial maize hybrids and depletion of genetic diversity in gene pools used in breeding may be partially alleviated by successful implementations of MAS. MABC may revive interest in using essentially untapped maize exotic germplasm as a source of favourable alleles for improvement of elite varieties. Very small and targeted chromosomal segments of exotic origin can be introgressed into elite inbred lines with limited risk of carrying along undesirable characteristics. Such an approach could be beneficial in maize although no accounts of its implementation have been reported despite the many years as reports of its successful use in tomato (Tanksley *et al.,* 1996; Bernacchi *et al.,* 1998a, b; Robert *et al.,* 2001), rice (Xiao *et al.,* 1998), and soybean (Concibio *et al.,* 2003). MARS, in turn, may also contribute to increasing genetic diversity among commercial maize hybrids because, by focusing on selecting specific recombination events, it will result in the development of genuinely new genomic rearrangements. As QTL identified in any experiment represent only a fraction of the loci responsible for the phenotypes of complex traits, one can assume that breeding programmes in different private companies will conduct MARS based on their different genetic models and select for different genomic rearrangements. As a result, hybrids of similar and high performance might be developed that are based on different sets of favourable alleles at different loci, representing distinct "genetic solutions" and contributing to increased genetic diversity in farmers' fields.

An indirect but important advantage of MAS and its underlying information and technology relates to intellectual property. Some maize breeding programmes have created a form of wealth through their collection and knowledge of maize germplasm. Significant investments have been made in maize breeding as exemplified by the billions of United States dollars that were used to purchase a few private programmes between 1995 and 2005. Protecting and maximizing returns on such investments have always been important but are now of greater concern. Information from MAS should be advantageous for addressing issues concerning ownership and derivation of germplasm, relatedness among germplasm and for the formation of some claims in patents and similar documents.

Perhaps one of the greatest advantages of MAS is that, for the first time, maize breeders have the means of learning some of the genetic details about germplasm and the response to selection. Some maize programmes in the private sector have started this process (Niebur *et al.,* 2004). As real functions become associated with the many candidate genes and other DNA sequences,

the opportunities for learning about and understanding the response to selection will increase dramatically. It may then be possible to ameliorate some of the limitations of MAS and truly breed by design.

Limitations of MAS

While not truly an inherent limitation of the methods involved, one unavoidable limitation of MAS is the cost of assembling and integrating the necessary infrastructure and personnel. These can be substantial and beyond the means of many programmes. For such programmes, implementation of MAS could lead to a delusional or unbalanced reallocation of resources from vital activities such as high-quality phenotypic evaluation and selection in the target environment. Currently, only the largest maize breeding programmes in a given market or region have the scale of sales and diversity of products that can justify and support MAS and withstand some of the financial burdens of establishing and replacing components of the system (e.g. changes in the methods and platforms for detecting DNA polymorphisms).

Some inherent limitations to MAS are related to the estimates of QTL position and genetic effects and the rates of false positives and negatives. Confidence intervals for QTL are typically 10–15 cM; a genetic region that should not be a major barrier for implementing MAS although it could become a limitation to achieving genetic gain by preventing the selection of desired recombination events. The advent of association mapping and a growing pool of candidate genes should provide some resources needed to minimize problems related to the estimation of QTL position. The genetic effects of QTL are overestimated for many reasons, some of which are linked to experimental designs for phenotyping or population development while others are inherent to the process of QTL detection (Lee, 1995; Beavis, 1998; Melchinger, Utz and Schön, 1998; Holland, 2004). In addition, genetic effects related to epistasis are either poorly estimated or ignored by programmes in the private sector (Holland, 2001; Crosbie *et al.,* 2006). Such assessments of genetic effects will inflate predictions of genetic gain. The relative merit of MAS will depend on the nature of predictions, actual results and costs of alternative methods.

A possible limitation of MAS with maize is the structure and content of various gene pools. Examples of maize gene pools would include European flint and dent germplasm, United States dents and various heterotic groups within each of these and other larger pools. Surveys with DNA markers have established differences among such groups of germplasm (Smith and Smith, 1992; Niebur *et al.,* 2004). The relatively allele-rich maize gene pools coupled with genetic heterogeneity for many traits will hinder the ability to extrapolate information about genotype-phenotype relationships across gene pools. Such transfer of information is expected to be more successful in relatively homogeneous and less diverse maize gene pools (e.g. sweetcorn or popcorn) and with self-pollinated plant species (Lee, 1995). There have been undocumented reports of a few alleles at QTL that have relatively universal genetic effects across a relatively broad range of maize populations and target environments, but details of such genetic factors have not been publicly disclosed (Crosbie *et al.,* 2006). More resources will need to be devoted to discovering where genetic information cannot be easily extrapolated across gene pools or even populations within a gene pool compared with situations where

it could. Although this should not impact the economic efficiency of MABC or forward breeding, it could affect the overall cost efficiency of MAS.

Finally, the efficacy for MAS in relatively complex populations such as synthetics and open-pollinated varieties (OPVs) has not been investigated. Compared with the bi-allelic populations used in the private sector, such populations are likely to have more than two alleles at a given locus. Also, unlike the simple bi-allelic populations, allele frequency should be an important component of predictions with such populations. Therefore, there should be more genetic effects and interactions to consider when making predictions based on MAS with OPVs and synthetics.

In the future, successful implementation of MAS in maize may lead to more frequent problems related to limited genetic variation. The emphasis of aggressive private sector maize breeding programmes on crosses between elite, related inbred lines to create segregating source populations has led to concerns about the depletion of genetic diversity in such gene pools and the ability to enhance such gene pools with high quality genetic variation (Niebur *et al.*, 2004). Such concerns, which existed prior to the deployment of DNA markers and MAS in maize, are likely to increase as MAS becomes more prevalent. If MAS in forward breeding schemes is as effective as reported, then alleles and haplotypes may approach fixation more rapidly (Crosbie *et al.*, 2006). At that point, breeding programmes will need to repeat the process of calibrating genotype-phenotype relationships in a slightly different array of reference populations to start the next metacycle of MAS (Johnson, 2004).

There is much anticipation for the future of MAS as genic sequences become the marker loci, functional information is discovered for the many candidate genes and gene products are assessed for their potential as useful sources of information in breeding programmes (Varshney, Graner and Sorrels, 2005; Lee, 2006). Certainly, these huge sets of raw data will contribute to progress. Eventually, other sources of genetic variation unrelated to the primary DNA sequence such as DNA methylation will be evaluated for their influence on genotype-phenotype relationships. Currently, epigenetic variation is mostly ignored from that assessment although it is well known that much of the maize genome may be methylated (Kaeppler, 2004) and may be more dynamic than predicted by current genetic models and mechanisms (Fu and Dooner, 2002). Also, the influences of non-coding sequences such as small interfering RNA (siRNA), matrix attachment regions and long-distance regulatory sequences have yet to be considered for their effects on genetic variation and estimates of genetic values used in MAS (Lee, 2006).

Most of the early limitations of MAS, due to the availability or cost of genotypic data, have been overcome. However, the availability or cost of high-quality phenotypic information is becoming one of the major limitations of MAS. During the past 20 years, development of new technologies and automation and miniaturization of laboratory procedures have contributed to reducing the cost of marker data points as well as the time needed to produce them. Large-scale marker laboratories produce marker data points at less than a tenth of the cost of 20 years ago. By contrast, neither cost nor the time required to produce phenotypic data has changed much, if at all, in the same timeframe. As the establishment of marker-trait associations and ultimately

the success of MAS depends on access to high-quality phenotypic data, means will have to be found to decrease the cost of phenotypic information while maintaining or increasing its quality. Alternatively, a greater proportion of budgets needs to be devoted to collecting phenotypic information.

ACHIEVEMENTS OF MAIZE BREEDING PROGRAMMES WITH MAS

In some important ways, maize breeding has gradually changed since the mid 1990s with the advent of genomics. Genetic principles were always an important component of modern maize breeding and now genetic information of various types is seeping into breeding schemes. MAS is the connection between the growing pool of genetic information and actual plant breeding. Establishing and enhancing this connection have been important achievements.

For the simplest breeding scenario, programmes in the private sector have demonstrated that MABC is an effective and routine method to backcross one or more transgenes into established elite inbred lines, the direct parents of commercial hybrids. Hybrids with effective combinations of transgenes have been very successful in the market. Consequently, MAS has accelerated the delivery of some products to the market; an important achievement in competitive economies.

Programmes in the private sector have also demonstrated a sufficient degree of efficacy of MAS methods to secure protection of intellectual property in patents. Methods, ideas and linkage relationships have been included in claims of patents or patent applications related to MAS (e.g. US5 492 547 1996; US6 455 758B1 2002; US2005/0144664A1 2005; WO2005/000006A2 2005; WO2005/014858A2 2005), or the establishment of marker-trait associations (e.g. US5 746 023P 1998; US6 368 806B1 2002; US6 399 855B1 2002). Given the magnitude of the investments made in maize breeding by the private sector, receiving such a legal position may be a valuable achievement for the owner of the patent.

The efficacy of MAS for forward breeding of complex traits has yet to be firmly established. Positive results from calibration studies have been reported, but although accounts of MAS-derived commercial varieties have been made (Eathington, 2005), the impact on actual breeding and the development of new commercial hybrids has not been disclosed to a significant extent (Johnson, 2004; Niebur *et al.*, 2004; Crosbie *et al.*, 2006). At this point in time, it is therefore too early to make a definitive and databased assessment of this aspect of MAS.

The history and cost of the genetic gain achieved through MAS will certainly vary among target environments. In some regions of the world, such as the central United States, maize breeding achieved steady genetic gains in grain yield for several consecutive decades prior to the advent of MAS (Duvick, Smith and Cooper, 2004). Nevertheless, the cost per unit gain has increased as more resources are needed for phenotypic evaluation in more environments (Smith *et al.*, 1999). However, the advent of applied genomics and the discovery of many genes and gene functions, coupled with MAS, could reduce the dependence on costly phenotypic information for breeding. In regions where biotic and abiotic stress factors are more important than in the central Unites States, MAS may be very effective. Ultimately, the value and achievements of MAS will depend on the ecological and socio-economic context of the target environment.

COLLABORATION BETWEEN THE PRIVATE AND PUBLIC SECTORS IN MAS AND MAIZE IMPROVEMENT

The increased investments in maize breeding, expected returns on investment and concerns regarding intellectual property by the private sector have made it more difficult for corporations to collaborate with external parties of any kind. Such factors hinder the exchange of information and material that is common in collaborative projects. Nevertheless, around the world, the private and public sectors still manage to collaborate through various mechanisms and at different levels in the pursuit of maize improvement. Such collaboration involves interactions among multinational corporations, philanthropic foundations, national and subnational governments, universities and individuals. Major categories of collaboration include social programmes and institutions, research and development, and education.

In many regions of the world, private sector maize breeding would not have grown without some critical social programmes and institutions. For example, legislation related to intellectual property, transfer of capital and material, and regulatory approval of biotechnical innovations in maize improvement have been important components of legal systems that have encouraged financial investment in maize breeding. The stability of these systems and the rule of law have contributed to the long-term gains in selection. Also, long-term crop subsidy programmes in some regions have provided an element of security for investments in maize research and development by the private sector (Troyer, 2004; Crosbie *et al.,* 2006). In those same regions, MAS has been deployed initially and on the largest scale for maize breeding.

With respect to research and development, there is a long history of effective collaboration between the public and private sectors in maize breeding. While such interaction continues in the era of MAS, the nature of the collaboration has changed with the growth and development of the breeding programmes in the private sector. Initially, collaboration was absolutely vital for the private sector because breeding programmes in the public sector were important, or the sole, sources of the inbred lines used directly by the private sector to produce commercial hybrids or to source populations from which elite inbreds were derived. Also, the inbred lines from the public sector were usually provided on an unrestricted basis and without payments of royalties or licensing fees. Public breeding programmes continue to develop elite inbred lines, occasionally in collaboration with the private sector (e.g. the Germplasm Enhancement of Maize programme in the United States; Pollak, 2003). However, the direct impact of contemporary public germplasm varies greatly among regions and gradually, in many regions of the world, the private sector has become the primary source of elite maize inbred lines and commercial hybrids.

In addition to germplasm, most or all of the critical concepts, methods and basic technologies have their origins in the public sector (Niebur *et al.,* 2004; Troyer, 2004; Crosbie *et al.,* 2006). The private sector, with its unique ability to concentrate capital through various mechanisms (e.g. profits from products or licence fees, venture capital and financial markets), is in the best position to allocate resources quickly to assess, modify and apply new developments in MAS and ancillary areas of maize improvement across large geographical and political regions of a market zone. As

described in previous sections, cost-effective MAS requires several components of an integrated infrastructure, some features of which have had a relatively high rate of renovation and replacement (e.g. methods of detecting DNA polymorphism), and therefore required substantial financial resources. Competing corporations and the potential for profit provide the necessary motivation for such investments (Troyer, 2004; Crosbie *et al.,* 2006). To the authors' knowledge, such financial mechanisms either do not exist or are limited in the public sector.

Collaboration between the public and private sectors in MAS for maize may be strongest in basic genetics and genome annotation. In order for MAS to reach its full potential, it may be necessary to acquire a much better understanding of gene function and products. For any plant species, only a small percentage of genes and other DNA sequences have a function defined through direct experimentation (Lee, 2006). Discoveries in plant gene function will occur in many laboratories around the world and, ultimately, the development groups in the private sector will have the necessary concentration of resources and sense of purpose to assemble the relatively raw basic information into tools and products from MAS. The maize nuclear genome, with tens of thousands of genes and many other important DNA sequences, is mostly a "black box" with respect to understanding the role of these in mediating phenotypes in response to environmental cues. Such understanding, a potential key to MAS and maize improvement, can only occur through informal and formal collaboration between the public and private sectors investigating a broad array of plant species.

Examples of collaborative research between the public and private sectors relevant to MAS in maize include attempts to select for hybrid yield (Stromberg, Dudley and Rufener, 1994), QTL mapping and selecting of hybrid yield (Eathington, Dudley and Rufener, 1997) and grain quality (Laurie *et al.,* 2004), the development of the *IBM* population of recombinant inbred lines, and mapping genomic regions that include the *vgt1* locus in maize (Lee *et al.,* 2002; Salvi *et al.,* 2002). National collaborative research programmes such as Génoplante in France and GABI in Germany, as well as several projects within the European Commission-sponsored framework programmes, are additional examples of such collaboration. Certainly, other collaborative projects between the public and private sectors have been conducted in maize MAS but their proprietary nature prevents public disclosure.

Future collaborative research activities in maize MAS could assume many forms. In most regions of the world, the private sector has the obvious superiority in terms of infrastructure needed for genotyping, phenotyping and data analysis. These resources are mostly devoted to the direct pursuit of products and profits. That pursuit may also be the greatest disadvantage of the private sector because such a focus limits the attention devoted to many interesting yet seemingly ancillary observations of genotype-phenotype relations in MAS. Some components of that infrastructure could possibly be made accessible to the public sector as "in-kind" contributions to collaborative or service-related projects in regions that are unlikely to emerge as important markets for the private sector or for phenotypes and germplasm that are not of direct interest to the private sector.

Education and training are also important areas in which the public and

private sectors should collaborate. With the advent of MAS, there has been an obvious need for maize breeders in the private sector to become familiar with all aspects of the process, and the public sector has developed several new short courses and training sessions in MAS-related concepts (Niebur *et al.*, 2004; Crosbie *et al.*, 2006).

Such knowledge is now considered a standard component of recent graduate training. However, while new students may have an adequate grasp of the theoretical aspects of MAS, their lack of exposure to the private sector's advanced infrastructure represents a gap in their education. This situation is similar to that of students with a new degree in engineering who join advanced engineering and design groups in other industries: the private sector's capacity to concentrate and focus capital often leads to advanced infrastructure that does not exist in the public sector. In such situations, new students have to navigate a rather steep learning curve before they become productive members of their new group. To reduce the slope of the learning curve, the private sector could provide internships to graduate students or to professors who teach plant breeding courses. It is unlikely that the public sector will have the resources to duplicate or exceed some features of the infrastructure that has been developed for maize MAS in the private sector. Therefore, for some aspects of education, it will be to everyone's benefit to find ways to work together.

PRIVATE SECTOR PERSPECTIVES ON MAS FOR MAIZE IMPROVEMENT

The development of molecular markers in the 1980s provided the first tools to dissect the genetic basis of traits and select individuals based on their predicted genetic value. Back in these early days, the availability of genetic information was a limiting factor. Today's landscape is very different as advances in applied genomics and laboratory technology have provided the tools to generate genetic information for all traits of interest. Gene similarities and synteny across genomes mean that much of the information generated on any plant species has relevance to other plant species. The speed at which genetic information becomes available never ceases to increase. Rather than its availability, it is the ability to handle and utilize genetic information that is becoming the limiting factor for MAS. New and improved information technology and bioinformatics capabilities therefore need to be developed that connect the growing wealth of genetic information with maize breeding programmes where knowledge about the genetic basis of traits and allelic variation at these loci is translated into varieties.

QTL and gene mapping will remain key for the generation and use of genetic information. As sequencing of cereal genomes including maize progresses, physical mapping of cloned genes will become a powerful alternative to statistical approaches. Characterization of allelic diversity at loci of interest can proceed from analyses of bi-parental populations or association studies. An effective alternative is the use of sets of NILs, or introgression line (IL) libraries (Peleman and van der Voort, 2003). As NILs developed around a specific locus differ only by the allele at this locus, and because most traits of agronomic interest in maize are quantitative, phenotypic differences among such NILs are expected to be rather small. High precision phenotyping will not only be required but will be critical for the evaluation of such material (Peleman and van der Voort, 2003). Private corporations have realized the need

for such high precision phenotyping as can be seen from their active recruiting of trait-specific phenotyping scientists often located in targeted areas where the trait of interest can be more easily measured (e.g. positions dedicated to drought tolerance and located in arid regions of the world).

In order to further the implementation of MAS in breeding, increased numbers of marker data points will be required. Private corporations have established or are developing the capacity to produce hundreds of millions of data points per year in service laboratories, distinct from research units. Besides, smaller "biotech" companies are developing technologies that could reduce the cost of each marker data point to a mere few United States cents. Moving to marker systems that are not based on gels is permitting the automation of most laboratory steps. Data points are being produced around the clock with laboratory technicians working in shifts. Here again, private companies are actively recruiting highly qualified technology specialists as well as laboratory managers whose role is more to optimize the running of production plants than dwell on the science. Beyond laboratories, plant handling is becoming a bottleneck to high-throughput protocols. High-throughput facilities have to be established and equipped at continuous nursery sites potentially to handle millions of plants per year.

There is little doubt that the largest private maize breeding programmes are investing very heavily in the implementation of MAS. Unless regulatory issues change dramatically, MABC will remain the preferred means of delivering transgenes to the market. Faster MABC protocols will always represent a potential commercial advantage in an area where competition is fierce and a one-year advantage may mean much on the market. Most recent investments have been directed at implementing MARS in breeding. The size of the investment in this approach seems to suggest that private corporations have more insight into its benefits compared with conventional breeding than has been reported publicly. Genotype-driven breeding should also allow faster development of specialized varieties as the maize market becomes more and more fragmented based on end-use of the harvest: animal feed (silage or grain), ethanol, dry or wet milling. Favourable alleles for traits of interest are likely to be spread across more than two lines therefore requiring the assembly of alleles from many different sources in a single inbred line. Proposals have been made to achieve such goals (Peleman and van der Voort 2003), although software tools to determine the optimal breeding schemes are not yet available to generate these "ideal" genotypes.

Maize breeding is likely to change more in the coming 10 or 20 years than it has over the past 50. Developing new hybrids efficiently now requires integrating data from many sources, sometimes beyond maize, generating high-quality genotypic and phenotypic data needed for the construction of "ideal" genotypes, and finally selecting phenotypically the best individuals from populations of marker-assisted-derived materials. Many stakeholders beyond maize breeders now take an active part in the development of new varieties and therefore breeding will increasingly become the responsibility of groups of individuals with complementary skills than stand-alone breeders. Training of all to understand and challenge the contribution of others will be critical to operating multidisciplinary breeding teams efficiently.

MAS FOR MAIZE IMPROVEMENT IN DEVELOPING COUNTRIES

A rapid analysis of the implementation of MAS in private maize breeding programmes points to three elements as being of particular importance: availability of high-quality phenotypic data, access to low-cost molecular marker data points and access to reliable continuous nurseries. The importance of high-quality phenotypic analyses has been clearly recognized by groups in the private sector (Niebur *et al.*, 2004; Crosbie *et al.*, 2006). Implementation of MAS in maize breeding requires large amounts of marker data points to be generated. Private groups have spent much effort developing technologies and platforms to achieve cost-efficient genotyping. Simultaneously, highly efficient continuous nurseries have been established in tropical environments or local greenhouses.

By contrast, maize breeding for developing countries is rather fragmented. National agricultural research institutions and international centres of the CGIAR such as the International Maize and Wheat Improvement Center (CIMMYT) focus much of their efforts on poor farmers and underserved regions. Private maize breeding programmes are also established in a number of developing countries. Due to the large up-front costs of assembling infrastructure and personnel for genotyping, it is unlikely that individual national marker laboratories could produce data points in a cost-efficient manner. However, regional facilities serving the needs of several national programmes and supported by local laboratories that could process samples (processing samples could be as easy as taking and air-drying them) and provide information in a timely manner, would probably be very sustainable alternatives. Such a regional molecular service laboratory has been established recently in Nairobi, Kenya, in a joint effort by two CGIAR centres, CIMMYT and the International Livestock Research Institute (ILRI) and Kenya's Agricultural Research Institute (KARI), under the Canadian International Development Agency (CIDA)-funded Biosciences eastern and central Africa (BecA) platform, to provide technical access and training for African maize breeders (Delmer, 2005). Such a facility could be an excellent component of a comprehensive maize breeding effort if it is possible to establish and maintain high-quality personnel and facilities for all of the other aspects of maize breeding in key target environments. However, without high-quality capabilities in phenotypic evaluation and selection, molecular laboratories will be worthless. Research projects involving large-scale (transnational) phenotypic evaluations of key genetic material and focused on specific traits (tolerance to biotic or abiotic stresses) should provide genetic information that is both locally relevant and broadly applicable (geographically and in terms of germplasm). Such projects would also spread the cost of phenotyping across all participants but would only be successful with effective transnational coordination.

Private companies running MAS in maize could contribute to its implementation in developing countries in several ways. First, they could make some of their genetic information available, thereby adding to that already available in the public domain. Much information is being generated in the private sector on traits of importance to developing countries such as disease resistance (e.g. grey leaf spot, northern corn leaf blight, *Fusarium* stalk and ear rots), drought tolerance and nitrogen use efficiency. After validation of its relevance

to the germplasm and environments of target areas, this genetic information could be used to select efficiently for specific traits through MAS. Second, private companies could provide access to some of their genotyping or nursery platforms. Genotyping samples for MAS projects in developing countries would not substantially disrupt private companies' own research if conducted in periods of lower activity, and would provide these MAS projects with marker data points for as low a cost as possible. Third and probably most critically, private companies could train scientists from developing countries on the principles, mechanics and logistics of applying and implementing MAS in maize. Scientists in private maize breeding groups have already identified many of the pitfalls and overcome many of the hurdles linked with the implementation of MAS. Transfer of this knowledge to scientists from developing countries would help them immensely to design marker-based breeding schemes adapted to their sets of constraints.

Beyond their contribution to the implementation of MAS in maize in developing countries, private companies could, in very similar ways, contribute to MAS programmes in other species of importance to developing countries but remote from their core interests. Synteny and gene conservation across species should allow some of the maize genetic information to be transferable to other species. Technology platforms and breeding approaches developed for MAS in maize should be good models for other crops and some might be directly usable. Mechanisms or organizations need to be put in place for these transfers of knowledge and technologies to occur from private maize MAS programmes to other crops in developing countries. Private programmes will likely not drive these transfers but might be very willing to contribute or be directly involved in specific projects provided adequate frameworks exist.

Public–private partnerships will need to be established to manage intellectual property issues related to the transfers of information, material or technologies from private companies to developing countries (Naylor *et al.,* 2004). The African Agricultural Technology Foundation (AATF) is one initiative that has been established to deal with such issues. Several private corporations with major investments in MAS in maize have agreed to provide access to germplasm and knowledge for African countries (Naylor *et al.,* 2004; Delmer, 2005).

As with the private sector in Europe and North America, it will be necessary to provide regular and easy access to education and training in maize MAS as the phenotypes and population structures are likely to differ from those encountered by programmes in the private sector in relatively high-input production environments. Also, and in common with the changes in the private sector, some reorganization or restructuring of public sector programmes may be warranted with the advent of more specialized roles for some personnel.

Understanding the genetic basis of traits and cloning and sequencing the underlying genes will not have an impact on poor farmers unless translated into varieties through breeding. Implementing MAS requires significant investments in both people and infrastructures. Some of the most promising marker-based breeding schemes (e.g. MARS), take about as long as conventional breeding schemes to develop improved varieties and therefore require long-term funding commitments. Funding of practical crop improvement has declined for several years, particularly

in the international public sector (Knight, 2003), and as a result investments have favoured research at the expense of practical applications (Naylor *et al.,* 2004). Whether current funding mechanisms based on short-term (two to five years) grants are adequate to allow maize or any other breeding programmes in developing countries to benefit from the much needed advantages of MAS is questionable.

REFERENCES

Basten, C.J., Weir, B.S. & Zeng, Z-B. 1994. Zmap--a QTL cartographer. *In* C. Smith, J.S. Gavora, B. Benkel, J. Chesnais, W. Fairfull, J.P. Gibson, B.W. Kennedy & E.B. Burnside, eds. *Proc. 5th Wrld. Congr. Genet. Appl. Livest. Prodn.,* Vol. *22,* pp. 65–66. Guelph, Ontario, Canada, University of Guelph.

Beavis, W.D. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proc. 49th Ann. Corn and Sorghum Ind. Res. Conf.,* pp. 250–266. Washington, DC, USA, American Seed Trade Association.

Beavis, W.D. 1998. QTL analyses: power, precision and accuracy. *In* H. Paterson, ed. *Molecular dissection of complex traits*, pp. 143–162. Boca Raton, FL, USA, CRC Press.

Beavis, W.D. 1999. Integrating the central dogma with plant breeding. *Abstr. Plant & Anim. Genome* 7: 28.

Bernacchi, D., Beck-Bunn, T., Emmatty, D., Eshed, Y., Inai, S., Lopez, J., Petiard, V., Sayama, H., Uhlig, J., Zamir, D. & Tanksley, S. 1998a. Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor. Appl. Genet*. 97: 170–180.

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D. & Tanksley, S. 1998b. Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theor. Appl. Genet.* 97: 381–397.

Bernardo, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* 34: 20–25.

Bernardo, R. 2001. What if we knew all the genes for a quantitative trait in hybrid crops ? *Crop Sci.* 41: 1–4.

Bernardo, R. 2002. *Breeding for quantitative traits in plants*. Woodbury, MN, USA, Stemma Press.

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. 1980. Construction of genetic linkage map in human using restriction fragment length polymorphisms. *Am. J. Human Genet*. 32: 314–331.

Bouchez, A., Hospital, F., Causse, M., Gallais, A. & Charcosset, A. 2002. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* 162: 1945–1959.

Burr, B., Burr, F.A., Thompson, K.H., Albertsen, M.C. & Stuber, C.W. 1988. Gene mapping with recombinant inbreds in maize. *Genetics* 118: 519–526.

Castiglioni, P., Ajmone-Marsan, P., van Wijk, R. & Motto, M. 1999. AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. *Theor. Appl. Genet*. 99: 425–431.

Charmet, G., Robert, N., Perretant, M.R., Gay, G., Sourdille, P., Groos, C., Bernard, S. & Bernard, M. 1999. Marker-assisted recurrent selection for cumulating additive and interactive QTL's in recombinant inbred lines. *Theor. Appl. Genet*. 99: 1143–1148.

Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M. & Rafalski, A. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3: 19.

Coe, E., Hancock, D., Kowalewski, S. & Polacco, M. 1995. Gene list and working maps. *Maize Genet. Coop. Newsl.* 69: 191–267.

Concibio, V.C., La Vallee, B., McLaird, P., Pineda, N., Meyer, J., Hummer, L., Yang, J., Wu, K. & Delannay, X. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor. Appl. Genet.* 106: 575–582.

Crosbie, T.M., Eathington, S.R., Johnson, G.R., Edwards, M., Reiter, R., Stark, S., Mohanty, R.G., Oyervides, M., Buehler, R.E., Walker, A.K., Dobert, R., Delannay, X., Pershing, J.C., Hall, M.A. & Lamkey, K.R. 2006. Plant breeding: past, present, and future. *In* K.R. Lamkey & M. Lee, eds. *Plant Breeding: The Arnel R. Hallauer Internat. Symp.* Ames, IA, USA., Blackwell Publishing.

Davis, G.L., McMullen, M.D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S. & Coe Jr., E.H. 1999. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics* 152: 1137–1172.

Delmer, D.P. 2005. Agriculture in the developing world: connecting innovations in plant research to downstream applications. *Proc. Natl. Acad. Sci. USA* 102**:** 15739–15746.

Duvick, D.N., Smith, J.S.C. & Cooper, M. 2004. Long-term selection in a commercial hybrid maize breeding program. *In* J. Janick, ed. *Plant Breeding Rev.* 24: 109–152.

Eathington, S.R. 2005. Practical applications of molecular technology in the development of commercial maize hybrids*. Proc. 60th Ann. Corn and Sorghum Seed Res. Conf.* Washington, D.C., USA, American Seed Trade Association.

Eathington, S.R., Dudley, J.W. & Rufener, G.K. 1997. Usefulness of marker-QTL associations in early generation selection. *Crop Sci.* 37: 1686–1693.

Edwards, M. & Johnson, L. 1994. RFLPs for rapid recurrent selection. *Proc. Symp. Analysis of Molecular Marker Data,* pp. 33–40. Corvallis, OR, USA, American Society of Horticultural Science and Crop Science Society of America.

Edwards, M.D. & Page, N.J. 1994. Evaluation of marker-assisted selection through computer simulation. *Theor. Appl. Genet.* 88: 376–382.

Fehr, W.R. 1987. *Principles of cultivar development*. New York, USA, Macmillan.

Frisch, M. & Melchinger, A.E. 2001a. Marker-assisted backcrossing for introgression of a recessive gene. *Crop Sci.* 41: 1485–1494.

Frisch, M. & Melchinger, A.E. 2001b. Marker-assisted backcrossing for simultaneous introgression of two genes. *Crop Sci.* 41: 1716–1725.

Frisch, M., Bohn, M. & Melchinger, A.E. 1999a. Minimal sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. *Crop Sci.* 39: 967–975.

Frisch, M., Bohn, M. & Melchinger, A.E. 1999b.Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci.* 39: 1295–1301.

Fu, H. & Dooner, H.K. 2002. Intraspecific violation of genetic co-linearity and its implications in maize. *Proc. Natl. Acad. Sci. USA* 99: 9573–9578.

Gallais, A., Dillmann, C. & Hospital, F. 1997. An analytical approach of marker assisted selection with selection on markers only. *In* R. Krajewski & Z. Kaczmarek, eds. *Advances in biometrical genetics. Proc.10th. Mtg. EUCARPIA Section Biometrics in Plant Breeding,* pp. 111–116. Poznan, Institute of Plant Genetics, Polish Academy of Sciences.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2004. *Bayesian data analysis*, 2nd edition. Boca Raton, FL, USA, Chapman and Hall/CRC.

Gimelfarb, A. & Lande, R. 1994. Simulation of marker-assisted selection in hybrid populations. *Genet. Res*. 63: 39–47.

Goodman, M.M. & Carson, M.L. 2000. Reality vs. myth: corn breeding, exotics, and genetic engineering. *Proc. 55th Ann. Corn and Sorghum Res. Conf.,* pp. 149–172. Washington, D.C., USA, American Seed Trade Association.

Goodman, M.M., Stuber, C.W., Newton, K. & Weissinger, H.H. 1980. Linkage relationships of 19 enzyme loci in maize. *Genetics 96*: 697–710.

Hallauer, A.R. & Miranda, J.B. 1981. *Quantitative genetics in maize breeding*. Ames, IA, USA, Iowa State University Press.

Helentjaris, T., Slocum, M., Wright, S., Schasfer, A. & Nienhuis, J. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theor. Appl. Genet*. 72: 761–769.

Ho, J.C., McCouch, S.R. & Smith, M.E. 2002. Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. *Theor. Appl. Genet.* 105: 440–448.

Hoisington, D. 1989. Working linkage maps. *Maize Genet. Coop. Newsl.* 63: 141–151.

Holland, J.B. 2001. Epistasis and plant breeding. *In* J. Janick, ed. *Plant Breeding Rev.* 21: 29–32.

Holland, J.B. 2004. Implementation of molecular markers for quantitative traits in breeding programs – challenges and opportunities. *In* T. Fischer, N. Turner, J. Angus, L. McIntyre, M. Robertson, A. Borrell & D. Lloyd, eds. *New directions for a diverse planet. Proc. 4th Internat. Crop Sci. Congr.,* Brisbane, Australia (available at www.cropscience.org.au/icsc2004).

Hospital, F. 2001. Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* 158: 1363–1379.

Hospital, F. & Charcosset, A. 1997. Marker-assisted introgression of quantitative trait loci. *Genetics* 147: 1469–1485.

Hospital, F., Chevalet, C. & Mulsant, P. 1992. Using markers in gene introgression breeding programs. *Genetics* 132: 1199–1210.

Jansen, R.C. 1993. Interval mapping of multiple quantitative trait loci by using molecular markers. *Theor. Appl. Genet*. 85: 252–260.

Jansen, R.C., Jannink, J-L. & Beavis, W.D. 2003. Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Sci.* 43: 829–834.

Johnson, G.R. 2004. Marker-assisted selection. *In* J. Janick, ed. *Plant Breeding Rev.* 24: 293–310.

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volkaert, G., Rueda, J., Linacero, R., Vazquez, A. & Karp, A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed*. 3: 381–390.

Jourjon, M-F., Jasson, S., Marcel, J., Ngom, B. & Mangin, B. 2005. MCQTL: multi-allelic QTL mapping in multi-cross design. *Bioinformatics* 21: 128–130.

Kao, C-H., Zeng, Z-B. & Teasdale, R. 1999. Multiple interval mapping for quantitative trait loci. *Genetics* 152: 1203–1216.

Kaeppler, S. 2004. Biotechnology: new horizons. *In* C.W. Smith, J. Betrán, & E.C.A. Runge, eds. *Corn: origin, history, technology, and production*, pp. 399–425. Hoboken, NJ, USA, John Wiley & Sons.

Knapp, S.J. 1998. Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. *Crop Sci.* 38: 1164–1174.

Knight, J. 2003. Crop improvement: a dying breed. *Nature* 421: 568–570.

Lande, R. & Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743–756.

Lander, E.S. & Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Laurie, C.C., Chasalow, S.D., LeDeaux, J.R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T.R. &. Dudley, J.W. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168: 2141–2155.

Lee, M. 1995. DNA markers in plant breeding programs. *Adv. Agron.* 55: 265–344.

Lee, M. 2006. The phenotypic and genotypic eras of plant breeding. *In* K.R. Lamkey & M. Lee, eds. *Plant Breeding: The Arnel R. Hallauer Internat. Symp.* Ames, IA, USA, Blackwell Publishing.

Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D. & Hallauer, A.R. 2002. Expanding the genetic map of maize with the intermated B73 x Mo17 (*IBM*) population. *Plant Mol. Biol.* 48: 453–461.

Lindblad-Toh, K., Winchester, E., Daly, M., Wang, D., Hirschhorn, J.N., Laviolette, J.P., Ardie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J. & Lander, E.S. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet.* 24: 381–386.

Manly, K.F. & Olson, J.M. 1999. Overview of QTL mapping software and introduction to Map Manager QT. *Mamm. Genome* 10: 327–334.

Mazur, B.J. 1995. Commercializing the products of plant biotechnology. *Trends Biotech*. 13: 319–323.

Melchinger, A.E., Utz, H.F. & Schön, C.C. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149: 383–403.

Mohan, M., Nair, S., Krishna, T.G., Yano, M., Bhatia, C.R. & Sasaki, T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed*. 3: 87–103.

Moreau, L., Charcosset, A. & Gallais, A. 2004. Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica* 137: 111–118.

Moreau, L., Charcosset, A., Hospital, F. & Gallais, A. 1998. Marker-assisted selection efficiency in populations of finite size. *Genetics* 148: 1353–1365.

Moreau, L., Lemarié, S., Charcosset, A. & Gallais, A. 2000. Economic efficiency of one cycle of marker-assisted selection. *Crop Sci.* 40: 329–337.

Morris, M., Dreher, K., Ribaut, J-M. & Khairallah, M. 2003. Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol. Breed*. 11: 235–247.

Naylor, R.L., Falcon, W.P., Goodman, R.M., Jahn, M.M., Sengooba, T., Tefera, H. & Nelson, R.J. 2004. Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Policy* 29: 15–44.

Niebur, W.S., Rafalski, J.A., Smith, O.S. & Cooper, M. 2004. Applications of genomics technologies to enhance rate of genetic progress for yield of maize within a commercial breeding program. *In* T. Fischer, N. Turner, J. Angus, L. McIntyre, M. Robertson, A. Borrell & D. Lloyd, eds. *New directions for a diverse planet. Proc. 4th Internat. Crop Sci. Congr.,* Brisbane, Australia. (available at www.cropscience.org.au/icsc2004).

Openshaw, S. & Frascaroli, E. 1997. QTL detection and marker-assisted selection for complex traits in maize. *Proc. 52nd Ann. Corn and Sorghum Res. Conf.,* pp. 44–53. Washington, DC, American Seed Trade Association.

Parisseaux, B. & Bernardo, R. 2004. *In silico* mapping of quantitative trait loci in maize. *Theor. Appl. Genet*. 109: 508–514.

Peleman, J.D. & van der Voort, J.R. 2003. Breeding by design. *Trends Plant Sci.* 7: 330–334.

Penner, G.A., Bush, A., Wise, R., Kim, W., Domier L., Kasha, K., Laroche, A., Scoles, G., Molner, S.J. & Fedak, G. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Meth. Appl*. 2: 341–345.

Pollak, L.M. 2003. The history and success of the public-private project on germplasm enhancement of maize (GEM). *Adv. Agron.* 78: 45–87.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5: 94–100.

Ragot, M. & Hoisington, D.A. 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. *Theor. Appl. Genet*. 86: 975–984.

Ragot M., Biasiolli, M., Delbut, M.F., Dell'Orco, A., Malgarini L., Thevenin, P., Vernoy, J., Vivant, J., Zimmermann, R. & Gay, G. 1995. Marker-assisted backcrossing: a practical example. *In* A. Bervillé & M. Tersac, eds. *Les Colloques, No. 72, Techniques et utilisations des marqueurs moléculaires,* pp. 45–56. Paris, France, INRA.

Ragot, M., Gay, G., Muller, J-P. & Durovray, J. 2000. Efficient selection for the adaptation to the environment through QTL mapping and manipulation in maize. *In* J-M. Ribaut & D. Poland, eds. *Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments*, pp. 128–130. México, D.F., CIMMYT.

Ribaut, J-M. & Betrán, J. 1999. Single large-scale marker-assisted selection (SLS-MAS). *Mol. Breed*. 5: 531–541.

Ribaut, J-M., Jiang, C. & Hoisington, D. 2002. Efficiency of a gene introgression experiment by backcrossing. *Crop Sci.* 42: 557–565.

Ribaut, J-M., Edmeades, G., Perotti, E. & Hoisington, D. 2000. QTL analyses, MAS results, and perspectives for drought-tolerance improvement in tropical maize. *In* J-M. Ribaut & D. Poland, eds. *Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments,* pp. 131–136. México, D.F., CIMMYT.

Robert, V.J.M., West, M.A.L., Inai, S., Caines, A., Arntzen, L., Smith, J.K. & St-Clair, D.A. 2001. Marker-assisted introgression of blackmold resistance QTL alleles from wild *Lycopersicon chesmanii* to cultivated tomato (*L. esculentum*) and evaluation of QTL phenotypic effects. *Mol. Breed*. 8: 217–233.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R.H., Horn, G.T., Mullis, K.B. & Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487–491.

Salvi, S., Tuberosa, R., Chiapparino, E., Maccaferri, M., Veillet, S., van Beuningen, L., Isaac, P., Edwards, K. & Phillips, R.L. 2002. Towards positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol.* 48: 601–613.

Sax, K. 1923. The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8: 552–560.

Senior, M.L. & Heun, M. 1993. Mapping maize microsatellites and polymerase chain reaction confir-

mation of the target repeats using a CT primer. *Genome* 36: 884–889.

Senior, M.L., Chin, E.C.L., Lee, M., Smith, J.S.C. & Stuber, C.W. 1996. Simple sequence repeat markers developed from maize sequences found in the GENEBANK database: Map construction. *Crop Sci.* 36: 1676–1683.

Servin, B. & Hospital, F. 2002. Optimal positioning of markers to control genetic background in marker-assisted backcrossing. *J. Hered.* 93: 214–217.

Shattuck-Eidens, D.M., Bell, R.N., Neuhausen, S. & Helentjaris, T. 1990. DNA sequence variation within maize and melon: Observations from polymerase chain reaction and direct sequencing. *Genetics* 126: 207–217.

Skroch, P. & Nienhuis, J. 1995. Impact of scoring error and reproducibility of RAPD data on RAPD based estimates of genetic distance. *Theor. Appl. Genet*. 91: 1086–1091.

Smith, J.S.C. &. Smith, O.S. 1992. Fingerprinting crop varieties. *Adv. Agron*. 47: 85–140.

Smith, O.S., Hoard, K., Shaw, F. & Shaw, R. 1999. Prediction of single-cross performance. *In* J. G. Coors & S. Pandey, eds. *The genetics and exploitation of heterosis in crops,* pp. 277–285. Madison, WI, USA, American Society of Agronomy.

Sprague, G.F. &. Dudley, J.W. 1988. *Corn and corn improvement*, 3rd edition. Madison, Wisconsin, USA, American Society of Agronomy.

Stam, P. 1995. Marker-assisted breeding. *In* J.W. Van Ooijen & J. Jansen, eds. *Biometrics in plant breeding: applications of molecular markers*. *Proc. 9th Mtg. EUCARPIA Section on Biometrics in Plant Breeding,* pp. 32–44. Wageningen, Netherlands, CPRO-DLO.

Stromberg, L.D., Dudley, J.W. & Rufener, G.K. 1994. Comparing conventional early generation selection with molecular marker assisted selection in maize. *Crop Sci.* 34: 1221–1225.

Stuber, C.W. 1992. Biochemical and molecular markers in plant breeding. *In* J. Janick, ed. *Plant Breeding Rev.* pp 37–61.

Stuber, C.W. & Moll, R.H. 1972. Frequency changes of isozyme alleles in a selection experiment for grain yield in maize (*Zea mays* L.). *Crop Sci.* 12: 337–340.

Tanksley, S.D., Young, N.D., Patterson, A.H. & Bonierbale, M.W. 1989. RFLP mapping in plant breeding: New tools for an old science. *Biotech.* 7: 257–264.

Tanksley, S.D., Grandillo, S., Fulton, T.M., Zamir, D., Eshed, Y., Petiard, V., Lopez, J. & Beck-Bunn, T. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theor. Appl. Genet.* 92: 213–224.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA. *Nucl. Acids Res.* 17: 6463–6471.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E.S. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genet.* 28: 286–289.

Troyer, A.F. 2004. Persistent and popular germplasm in seventy centuries of corn evolution. *In* C.W. Smith, J. Betrán & E.C.A. Runge, eds. *Corn: origin, history, technology, and production,* pp. 133–232. Hoboken, NJ, USA, John Wiley & Sons.

Utz, H.F. & Melchinger, A.E. 1996. PLABQTL: a program for composite interval mapping of QTL. *J. Quant. Trait Loci* 2: 1.

van Berloo, R. & Stam, P. 1998. Marker-assisted selection in autogamous RIL populations: a simulation study. *Theor. Appl. Genet*. 96: 147–154.

van Berloo, R. & Stam, P. 2001. Simultaneous marker-assisted selection for multiple traits in autogamous crops. *Theor. Appl. Genet*. 102: 1107–1112.

van Ooijen, J.W. & Maliepaard, C. 1996. MapQTL version 3.0: software for the calculation of QTL positions on genetic maps. *Plant Genome IV,* San Diego, CA, USA (available at www.intl-pag.org/4/abstracts/p316).

Varshney, R.K., Graner, A. & Sorrels, M.E. 2005. Genomics-assisted breeding for crop improvement. *Trends Plant Sci*. 10: 621–630.

Visscher, P.M., Haley, C.S. & Thompson, R. 1996. Marker-assisted introgression in backcross breeding programs. *Genetics* 144: 1923–1932.

Vos P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res*. 23: 4407–4414.

Welsh, J. & McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* 18: 7213–7218.

Willcox, M.C., Kairallah, M.M., Bergvinson, D., Crossa, J., Deutsch, A., Edmeades, G.O., González-de-León, D., Jiang, C., Jewell, D.C., Mihm, A J., Williams, W.P. & Hoisington, D. 2002. Selection for resistance to southwestern corn borer using marker-assisted and conventional backcrossing. *Crop Sci*. 42: 1516–1528.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res*. 18: 6531–6535.

Xiao, J., Li, J., Grandillo, S., Nag Ahn, S., Yuan, L., Tanksley, S.D. & McCouch, S.R. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* 150: 899–909.

Xie, C. & Xu, S. 1998. Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity* 80: 489–498.

Young, N.D. 1999. A cautiously optimistic view for marker-assisted breeding. *Mol. Breed.* 5: 505–510.

Yousef, G.G. & Juvik, J.A. 2001. Comparison of phenotypic and marker-assisted selection form quantitative traits in sweet corn. *Crop Sci*. 41: 645–655.

Zeng, Z-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA* 90: 10972–10976.

Zeng, Z-B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.

CHAPTER 9

Molecular marker-assisted selection for resistance to pathogens in tomato

Amalia Barone and Luigi Frusciante

SUMMARY

Since the 1980s, the use of molecular markers has been suggested to improve the efficiency of releasing resistant varieties, thus overcoming difficulties met with classical breeding. For tomato, a high-density molecular map is available in which more than 40 resistance genes are localized. Markers linked to these genes can be used to speed up gene transfer and pyramiding. Suitable PCR markers targeting resistance genes were constructed directly on the sequences of resistance genes or on restriction fragment length polymorphisms (RFLPs) tightly linked to them, and used to select resistant genotypes in backcross schemes. In some cases, the BC_5 generation was reached, and genotypes that cumulated two homozygous resistant genes were also obtained. These results supported the feasibility of using marker-assisted selection (MAS) in tomato and reinforcing the potential of this approach for other genes, which is today also driven by the development of new techniques and increasing knowledge about the tomato genome.

INTRODUCTION

Tomato (*Solanum lycopersicum*, formerly *Lycopersicon esculentum*) is one of the most widely grown vegetable crops in the world. It is used as a fresh vegetable and can also be processed and canned as a paste, juice sauce, powder or as a whole. World volume has increased approximately 10 percent since 1985, reflecting a substantial increase in dietary use of the tomato. Nutritionally, tomato is a significant source of vitamins A and C. Furthermore, recent studies have shown the importance of lycopene, a major component of red tomatoes, which has antioxidant properties and may help to protect against cancer and heart disease (Rao and Agarwal, 2000).

One of the main constraints to tomato cultivation is damage caused by pathogens, including viruses, bacteria, nematodes and fungi, which cause severe losses in production. The control of pathogen spread mainly involves three strategies: husbandry practices, application of agrochemicals and use of resistant varieties. Husbandry techniques generally help to restrict the spread of pathogens and their vectors as well as to keep plants healthy, thus allowing pathogen attack to be limited. Chemical control gives good results for some pathogens, but poor results against others, such as bacteria, and has practically no effect on viruses. Moreover, reducing chemical treatments lowers the health risks to farmers and consumers. Therefore, in order to achieve sustainable agriculture and obtain high-quality, safe and healthy products, the use of resistant varieties is one of the principal tools to reduce pathogen damage.

Since the early part of the twentieth century, breeding for disease resistance has been a major method for controlling plant disease. Varieties that are resistant or tolerant to one or a number of specific pathogens are already available for many crops, and hybrids with multiple resistance to several pathogens exist and are currently used in vegetable production. In tomato, genetic control of pathogens is a very useful practice with most resistance being monogenic and dominant. Various sources of resistance have been used in traditional breeding programmes, and resistant breeding lines, varieties and F_1 hybrids have been developed with varying stability and levels of expression (Table 1) (Laterrot, 1996; Gardner and Shoemaker, 1999; Scott, 2005).

TABLE 1
List of pathogen resistances present in tomato breeding lines, varieties and F_1 hybrids obtained through conventional breeding

Virus
Beet curly top virus (BCTV)
Tobacco mosaic virus (TMV)
Tomato mosaic virus (ToMV)
Tomato yellow leaf curl virus (TYLCV)
Tomato spotted wilt virus (TSWV)
Bacteria
Corynebacterium michiganense
Pseudomonas solanacearum
Pseudomonas syringae pv. *tomato*
Nematodes
Meloidogyne spp.
Fungi
Alternaria alternata f. sp. *lycopersici*
Alternaria solani
Cladosporium fulvum
Fusarium oxysporum f. sp. *lycopersici*
Fusarium oxysporum f. sp. *radicis-lycopersici*
Phytophthora infestans
Pyrenochaeta lycopersici
Stemphylium solani
Verticillium dahliae

Modified from Laterrot (1996) and updated as reported in the text.

MARKER-ASSISTED BREEDING FOR PATHOGEN RESISTANCE

Although conventional plant breeding has had a significant impact on improving tomato for resistance to important diseases, the time-consuming process of

TABLE 2
Resistance genes mapped on the tomato molecular map

Pathogen	Gene[1]	Chromosomal location	Reference
Virus			
Alfalfa mosaic virus (AMV)	*Am*	6	Parrella *et al*., 2004
Cucumber mosaic virus (CMV)	*Cmr*	12	Stamova and Chetelat, 2000
Potato virus Y (PVY)	*pot-1*	3	Parrella *et al*., 2002
Tomato mottle virus (ToMoV)	2 genes	6	Griffiths and Scott, 2001
Tobacco mosaic virus (TMV)	*Tm-1, Tm2a*	2, 9	Young and Tanksley, 1988; Levesque *et al*., 1990
Tomato spotted wilt virus (TSWV)	*Sw5*	9	Stevens, Lamb & Rhoads, 1995
Tomato yellow leaf curl virus (TYLCV)	*Ty-1* (Q), *Ty-2*	6, 11	Zamir *et al*., 1994; Chagué *et a*l., 1997; Hanson *et al*., 2000
Bacteria			
Clavibacter michiganensis	*Cm1.1- Cm 10.1* (Q)	1, 6, 7, 8, 9, 10	Sandbrink *et al*., 1995
	QTLs	5, 7, 9	van Heusden *et al*., 1999
	Rcm2.0 (Q), *Rcm5.1* (Q)	2, 5	Kabelka, Franchino & Francio, 2002
Pseudomonas syringae pv. *tomato*	*Prf* *Pto*	6 6	Salmeron *et a*l., 1996 Martin *et al*., 1993
Ralstonia solanacearum	*Bw 1, Bw 3, Bw 4, Bw 5* (Q)	6, 10, 4, 6	Danesh *et al*., 2004; Thoquet *et al*., 1996
Xanthomonas campestris pv *vesicatoria*	*Bs4*	5	Ballvora *et al*., 2001
	rx-1, rx-2, rx-3	1	Yu *et al*., 1995
Nematodes			
Globodera rostochiensis	*Hero*	4	Ganal *et al*., 1995
Meloidogyne spp.	*Mi, Mi-3, Mi-9*	6, 12, 6	Williamson *et al*., 1994; Yaghoobi *et al*., 1995; Ammiraju *et al*., 2003
Fungi			
Alternaria alternata f. sp. *lycopersici*	*Asc*	3	van der Biezen, Glagotlkaya & Overduin, 1995
	QTLs 2a, 2c, 3, 9, 12	2, 3, 9, 12	Robert *et al*., 2001
Alternaria solani	*EBR*-QTLs	All	Foolad *et al*., 2002; Zhang *et al*., 2003
Cladosporium fulvum	*Cf-1, Cf-2, Cf-4, Cf-5, Cf-9*	1, 6, 1, 6, 1	Balint-Kurti *et al*., 1994; Jones *et al*., 1993
Fusarium oxysporum f. sp. *radicis- lycopersici*	*Frl*	9	Vakalounakis *et al*., 1997
Fusarium oxysporum f. sp. *lycopersici*	*I1, I2, I3*	7, 11, 7	Bournival, Vallejos and Scott, 1990; Sarfatti *et al*., 1991; Tanksley and Costello, 1991; Ori *et al*., 1997
Leveillula taurica	*Lv*	12	Chunwongse *et al*., 1994
Oidium lycopersicon	*Ol-1, ol-2, Ol-3, Ol-4*	6, 4, 6, 6	Huang *et al*., 2000; Bai *et al*., 2004; De Giovanni *et al*., 2004
	Ol-qtl1, Ol-qtl2, Ol-qtl3	6, 12	Bai *et al*., 2003
Phytophthora infestans	lb1-lb12 (Q)	All	Brouwer, Jones and St. Clair, 2004
	Ph-1, Ph-2, Ph-3	7, 10, 9	Moreau *et al*., 1998; Chunwongse *et al*., 2002
Pyrenochaeta lycopersici	*py-1*	3	Doganlar *et al.,* 1998
Stemphylium spp.	*Sm*	11	Behare *et al*., 1991
Verticillium dahliae	*Ve1, Ve2*	9	Diwan *et al*., 1999; Kawchuck *et al*., 2001

[1] Q in parenthesis, QTL and qtl indicate quantitative trait loci. Recessive resistance genes are reported with small letters.

making crosses and backcrosses, and the selection of the desired resistant progeny, make it difficult to respond adequately to the evolution of new virulent pathogens. Moreover, several interesting resistances are difficult to use because the diagnostic tests often cannot be developed due to the challenge posed by inoculum production and maintenance. In addition, where symptoms are detectable only on adult plants and/or fruits, diagnostic tests can be particularly expensive and difficult to perform.

Since the 1980s, the use of molecular markers has been suggested as a tool for breeding many crops, including tomato. In the last two decades, molecular markers have been employed to map and tag major genes and quantitative trait loci (QTL) involved in monogenic and polygenic resistance control, known respectively as vertical and horizontal resistance. To date, more than 40 genes (including many single genes and QTL) that confer resistance to all major classes of plant pathogens have been mapped on the tomato molecular map (Table 2) and/or cloned from *Solanaceous* species, as reported by Grube, Radwanski and Jahn (2000). Since then, other resistance genes together with resistance gene analogues (RGAs), which are structurally related sequences based on the protein domain shared among cloned R genes (Leister *et al*., 1996), have been added to the map. A molecular linkage map of tomato based on RGAs has also been constructed in which 29 RGAs were located on nine of the 12 tomato chromosomes (Foolad *et al*., 2002; Zhang *et al.,* 2002). Several RGA loci were found in clusters and their locations coincided with those of several known tomato R genes or QTL. This map provides a basis for further identifying and mapping genes and QTL for disease resistance and will be useful for MAS.

In fact, independently of the type of marker used for selection, by making it possible to follow the gene under selection through generations rather than waiting for phenotypic expression of the resistance gene, markers tightly linked to resistance genes can greatly aid disease resistance programmes. In particular, genetic mapping of disease resistance genes has greatly improved the efficiency of plant breeding and also led to a better understanding of the molecular basis of resistance.

DNA marker technology has been used in commercial plant breeding programmes since the early 1990s, and has proved helpful for the rapid and efficient transfer of useful traits into agronomically desirable varieties and hybrids (Tanksley *et al*., 1989; Lefebvre and Chèvre, 1995). Markers linked to disease resistance loci can now be used for MAS programmes, thus also allowing several resistance genes to be cumulated in the same genotype ("pyramiding" of resistance genes), and they may be also useful for cloning and sequencing the genes. In tomato, several resistance genes have been sequenced to date, among them *Cf-2*, *Cf-4*, *Cf-5*, *Cf-9*, *Pto*, *Mi*, *I2*, and *Sw5*. These cloned R genes now provide new tools for tomato breeders to improve the efficiency of breeding strategies, via MAS. Although MAS is still not used routinely for improving disease resistance in many important crops (Michelmore, 2003), it is being used by seed companies for improving simple traits in tomato (Foolad and Sharma, 2005). Furthermore, while the deep knowledge of the tomato genome and the availability of a high-density molecular map for this species (Pillen *et al.,* 1996) should provide further opportunities to accelerate breeding through MAS, the time-consuming and expensive process of developing markers associated with genes of inter-

est and the high cost of genotyping large populations has and will continue to limit the use of MAS in most tomato breeding programmes.

The potential of MAS to speed up the breeding of tomato using molecular markers linked to various resistance genes has been examined in the authors' laboratory. The two main goals of the research were to find the most suitable markers, and to test the feasibility of MAS for pyramiding resistance genes in tomato "elite" lines selected for their good processing qualities.

STRATEGIES FOR GENE TRANSFER AND PYRAMIDING

Six tomato genotypes carrying various resistance genes (Table 3) were crossed with tomato "elite" lines previously selected for yield and quality but lacking resistance traits. Each resistant genotype was crossed initially with each "elite" tomato line and various backcross schemes were then carried out starting from different F_1 hybrids. At each backcross generation the screening of resistant genotypes was performed using molecular markers linked to the resistance genes and DNA extracted from young leaves at seedling stage. Only resistant plants were then transplanted and grown in the greenhouse. At flowering, crosses were made with the recurrent parent to obtain the subsequent generations.

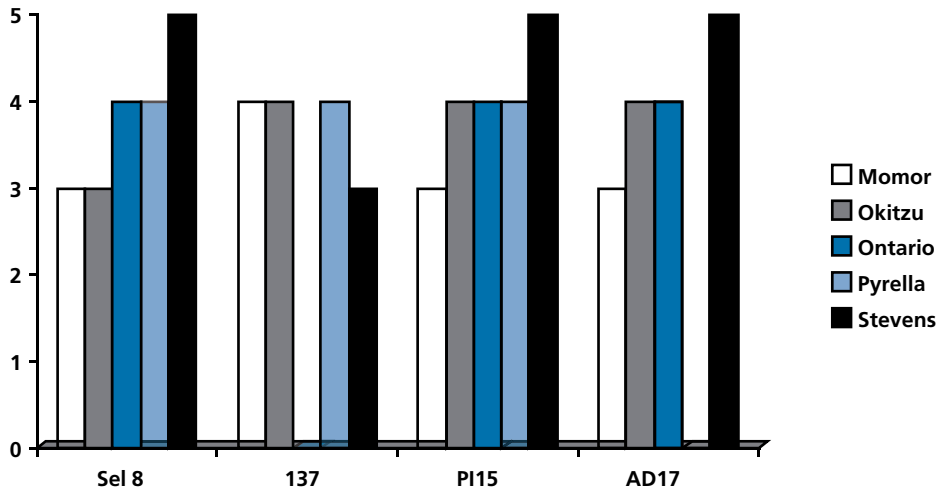
As the efficiency of MAS depends on the availability of polymerase chain reaction (PCR)-based markers highly linked to the resistance gene to be selected, for each resistance gene the most suitable marker system was investigated. For this purpose, three different strategies were undertaken. The first involved searching PCR markers already available in the literature and verifying their usefulness on the genetic material used. The second consisted of designing PCR primers from the sequence of cloned genes reported in the GeneBank database of the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/Genbank), while the third involved designing PCR primers from RFLP markers tightly linked to resistance genes. This last strategy was made possible by the availability of sequences of various mapped tomato RFLPs in the SolGenes database (www.sgn.cornell.edu).

In most cases, the results were obtained using cleaved amplified polymorphic sequences (CAPS; Konieczyn and Ausubel, 1993), which require one PCR reaction followed by restriction digest of the amplified fragment. In three cases (markers linked to genes *Mi*, *Sw5* and *Tm2a*), the primers and enzymes used were those reported in the literature (Williamson *et al*., 1994; Folkertsma *et al*., 1999; Sobir *et al*., 2000). In the case of gene *py-1*, the procedure reported in the literature (Doganlar *et al*., 1998) was

TABLE 3
Tomato genotypes used as resistant parents in the backcross breeding schemes. For each genotype resistant genes are reported

Genotype	Resistance gene	Pathogen
Momor	*Frl, Tm2a, Ve*	*Fusarium oxysporum* f. sp. *radicis-lycopersici,* TMV, *Verticillium dahliae*
Motelle	*I2, Mi, Ve*	*Fusarium oxysporum* f. sp *lycopersici, Meloidogyne* spp., *Verticillium dahliae*
Okitzu	*I2, Tm2a*	*Fusarium oxysporum* f. sp *lycopersici*, TMV
Ontario	*Pto*	*Pseudomonas syringae*
Pyrella	*py-1*	*Pyrenochaeta lycopersici*
Stevens	*Sw5*	TSWV

FIGURE 1
Number of generations reached in backcross schemes carried out between four susceptible recurrent genotypes and five resistant donor genotypes



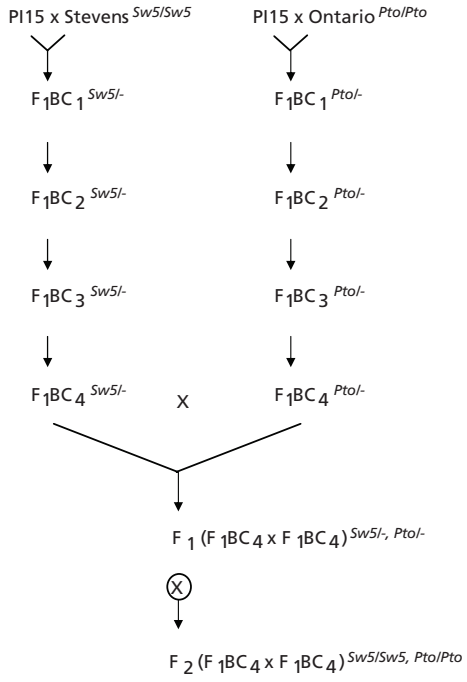
simplified, enabling a faster and cheaper marker system, i.e. a sequence characterized amplified region (SCAR; Kawchuk, Hachey and Lynch, 1998) marker, which only requires one PCR reaction to detect polymorphism between the resistant and the susceptible genotypes, to be set up. (Barone *et al*., 2004).

The second strategy was followed to design primers and enzymes suitable for targeting three resistance genes (*I2, Pto* and *Ve2*). This strategy allowed gene-assisted selection to be achieved through the simple PCR procedure. Finally, the third strategy was applied in the case of one CAPS marker targeting the resistance gene *Frl;* it was derived from one RFLP tomato marker (TG101) linked to the gene (Fazio, Stevens and Scott, 1999).

The markers found were used to select resistant genotypes in backcross breeding schemes, while the process itself allowed three generations to be screened annually. At present, for some cross combinations, the BC_5 generation has been reached, for others the BC_2-BC_3 (Figure 1). Where a BC_5 generation was already available, the breeding programme continued by selfing BC_5 resistant genotypes. In all other cases the backcross programme will continue up to the fifth backcross generation. At the end of each backcross scheme, the resistant BC_5F_3 genotypes, selected through molecular marker analysis, will also be tested directly for resistance by inoculating the pathogen and monitoring signs of disease. This will allow verification that no linkage breakage and loss of resistance gene occurred.

This procedure was already adopted in the case of one backcross scheme aimed at transferring a resistance gene to tomato spotted wilt virus (TSWV) to the susceptible genotype AD17 (Langella *et al*., 2004). The *in vivo* test performed on F_1BC_5, F_2BC_5 and, F_3BC_5 generations confirmed the introgression of the resistance trait and revealed that the resistance gene *Sw5* was

FIGURE 2
Breeding scheme used for the pyramiding of two resistant homozygous genes (*Sw5* and *Pto*) in the susceptible genotype PI15



fixed at the homozygous stage at the F_3BC_5 generation.

Finally, besides the transfer of one resistance gene to each susceptible genotype, a crossing scheme was undertaken to accumulate two or three resistance genes in the same genotype. In this case, the decision was made to stop the backcross scheme at the BC_3 or BC_4 generation as both parental lines were cultivated varieties and hence genetically very similar, and therefore the recovery of the recurrent genome could be satisfactory. F_1BC_4 hybrids carrying the same genetic background in the recurrent parent have been intercrossed, following the breeding scheme shown in Figure 2. At the end of each F_1BC_4 x F_1BC_4 cross and after selecting the genotypes carrying all the resistant alleles at the heterozygous level, one or two selfing generations will be carried out to fix all the resistant genes at the homozygous level.

This strategy has already started in some cases and the first homozygous multiresistant genotypes have been obtained. Also available are two F_2 genotypes out of 52 analysed plants, obtained by intercrossing the F_1BC_4 progeny from PI15 x Stevens with the F_1BC_4 progeny from PI15 x Ontario (Table 4). This F_2 generation exhibited two genotypes carrying both resistant genes *Sw5* and *Pto* at the homozygous level as well as 29 genotypes carrying both genes at the heterozygous level.

The work reported here on transferring resistance genes among tomato genotypes demonstrates the usefulness of MAS for improving traditional breeding strategies. The contribution of molecular markers linked to resistance genes was very efficient in reducing the time and space necessary for selection, enabling both early screening for resistance and reduced numbers of genotypes to be transplanted. The most challenging work was the search for suitable markers, which often required both considerable time and financial resources. Different strategies were used successfully to find the most suitable markers to perform MAS for transferring eight resistance genes into superior tomato genotypes; such strategies could be repeated in tomato for many other genes due to advanced molecular knowledge of the genome of this species.

PERSPECTIVES

The availability of PCR-based markers for many resistance genes allows MAS for biotic resistance in tomato to be applied successfully in any laboratory without the need for highly sophisticated techniques.

TABLE 4
Resistant heterozygous plants obtained in some cross combinations realized for pyramiding of resistance genes

Cross	Pyramided genes	Generation	Analysed plant (number)	Resistant plant (number)
(137 x Momor) F_1BC_3 x (137 x Ontario) F_1BC_3	*Tm2a – Frl* *Ve – Pto*	F_1	50	7
(AD17 x Okitzu) F_1BC_4 x (AD17 x Stevens) F_1BC_4	*Tm2a – Frl* *Sw5*	F_1	24	5
(AD17 x Ontario) F_1BC_4 x (AD17 x Stevens) F_1BC_4	*Pto – Sw5*	F_1	24	6
(PI15 x Stevens) F_1BC_4 x (PI15 x Ontario) F_1BC_4	*Sw5 - Pto*	F_2	52	29

Indeed, once a marker has been set up, its use on large populations for resistance screening is then routine. Technical facilities are today available for screening many samples simultaneously and also costs for equipment are decreasing. In addition, the rapid development of new molecular techniques, combined with the ever-increasing knowledge about the structure and function of resistance genes (Hulbert *et al.*, 2001), will help to identify new molecular markers for MAS, such as single nucleotide polymorphisms (SNPs). Moreover, thanks to the International *Solanaceae* Genome Project (SOL), sequencing of the tomato genome is in progress, and in a few years this will enhance information on resistance genes. This in turn will facilitate the development of molecular markers from transcribed regions of the genome, thereby allowing large-scale gene-assisted selection (GAS) to be achieved.

Over the coming years, selection for pathogen resistance in tomato will be underpinned by research aimed at: mapping other resistance genes for new pathogens; developing PCR-based functional markers (Andersen and Ludderstedt, 2003); designing the most suitable breeding schemes (Peleman and van der Voort, 2003), especially for transferring QTL resistances; large-scale screening through automation; allele-specific diagnostics (Yang *et al.*, 2004); and DNA microarrays (Borevitz *et al.*, 2003). In effect, the combination of new knowledge and new tools will lead to changes in the strategies used for breeding by exploiting the potential of integrating "omics" disciplines with plant physiology and conventional plant breeding, a process that will drive the evolution of MAS into genomics-assisted breeding (Morgante and Salamini, 2003; Varshney, Graner and Sorrells, 2005).

ACKNOWLEDGEMENTS

The authors wish to thank Mr Mark Walters for editing the manuscript and Angela Cozzolino for her technical assistance. This research was partially financed by the Ministry of Agriculture and Forestry of Italy in the framework of project PROMAR (Plant Protection by the Use of Molecular Markers). Contribution no. 119 from the Department of Soil, Plant and Environmental Sciences.

REFERENCES

Ammiraju, J.S.S., Veremis, J.C., Huang, X., Roberts, P.A. & Kaloshian, I. 2003. The heat-stable root-nematode resistance gene Mi-9 from *Lycopersicon peruvianum* is localized on the short arm of chromosome 6. *Theor. Appl. Genet.* 106: 478–484.

Andersen, J.R. & Lubberstedt, T. 2003. Functional markers in plants. *Trends Plant Sci.* 8: 554–560.

Bai, Y., Huang, C-C., van der Hulst, R., Meijer-Dekens, F., Bonnema, G. & Lindhout, P. 2003. QTLs for tomato powdery mildew resistance (*Oidium lycopersici*) in *Lycopersicon parviflorum* G1.1601 co-localize with two qualitative powdery mildew resistance genes. *Mol. Plant - Microbe Interact*. 16: 169–176.

Bai, Y., van der Hulst, R., Huang, C-C., Wei, L., Stam, P. & Lindhout, P. 2004. Mapping Ol-4, a gene conferring resistance to *Oidium neolycopersici* and originating from *Lycopersicon peruvianum* LA2172, requires multi-allelic, single locus markers. *Theor. Appl. Genet.* 109: 1215–1223.

Balint-Kurti, P.J., Dixon, M.S., Jnes, D.A., Norcott, K.A. & Jones, J.D.G. 1994. RFLP linkage analysis of the Cf-4 and Cf-9 genes for resistance to *Cladosporium fulvum* in tomato. *Theor. Appl. Genet.* 88: 691–700.

Ballvora, A., Pierre, M., van den Ackerveken, G., Chornack, S., Rossier, O., Ganal, M., Lahaye, T. & Bonas, U. 2001 Genetic mapping and functional analysis of the tomato *Bs4* locus governing recognition of the *Xanthomonas campestris* pv.*vesicatoria* AvrBs4 protein. *Mol. Plant-Microbe Interact.* 14: 629–638.

Barone, A., Langella, R., Ercolano, M.R., Di Matteo, A. & Monti, L. 2004 Strategies for pathogen resistance gene transfer in tomato with the aid of molecular markers. *In* S.G. Pandalai, ed. *Recent research developments in genetics and breeding*, pp. 211–222. Kerala, India, Research Signpost.

Behare, J., Laterrot, H., Sarfatti, M. & Zamir, D. 1991. RFLP mapping of the *Stemphylium* resistance gene in tomato. *Mol. Plant-Microbe Interact.* 4: 489–492.

Borevitz, J.O., Liang, D., Plouffe, D., Chang, H-S., Zhu, T., Weigel, D., Berry, C.C., Winzeler, E. & Chory, J. 2003. Large-scale identification of single-feaqture polymorphisms in complex genomes. *Genome Res.* 13: 513–523.

Bournival, B.L., Vallejos, C.E. & Scott, J.W. 1990. Genetic analysis of resistances to race 1 and 2 of *Fusarium oxysporum* f. sp. *lycopersici* from the wild tomato *Lycopersicon pennellii*. *Theor. Appl. Genet.* 79: 641–645.

Brouwer, D.J., Jones, E.S. & St Clair, D.A. 2004. QTL analysis of quantitative resistance to *Phytophthora infestans* (late blight) in tomato and comparisons to potato. *Genome* 47: 475–492.

Chagué, V., Mercier, J.C., Guénard, M., de Courcel, A.G.L. & Vedel, F. 1997. Identification of RAPD markers linked to a locus involved in quantitative resistance to TYLCV in tomato by bulked segregant analysis. *Theor. Appl. Genet.* 95: 671–677.

Chunwongse, J., Bunn, T.B., Crossman, C., Jing, J. & Tanksley, S.D. 1994. Chromosomal localization and molecular-marker tagging of the powdery mildew resistance gene (Lv) in tomato. *Theor. Appl. Genet.* 89: 76–79.

Chunwongse, J., Chunwongse, C., Black, L. & Hanson, P. 2002. Molecular mapping of the *Ph-3* gene for late blight resistance in tomato. *J. Hort. Sci. & Biotech.* 77: 281–286.

Danesh, D., Aarons, S., McGill, G.E. & Young, N.D. 1994. Genetic dissection of oligenic resistance to bacterial wilt in tomato. *Mol. Plant-Microbe Interact.* 7: 464–471.

De Giovanni, C., Dell'Orco, P., Bruno, A., Ciccarese, F., Lotti, C. & Ricciardi, L. 2004. Identification of PCR-based markers (RAPD, AFLP) linked to a novel powdery mildew resistance gene (ol-2) in tomato. *Plant Sci.* 166: 41–48.

Diwan, N., Fluhr, R., Eshed, Y., Zamir, D. & Tanksley, S.D. 1999. Mapping of Ve in tomato: a gene conferring resistance to broad-spectrum pathogen, *Verticillium dahliae* race 1. *Theor. Appl. Genet.* 98: 315–319.

Doganlar, S., Dodson, J., Gabor, B., Beck-Bunn, T., Crossman, C. & Tanksley, S.D. 1998 Molecular mapping of the py-1 gene for resistance to corky root rot (*Pyranochaeta lycopersici*) in tomato. *Theor. Appl. Genet*. 97: 784–788.

Fazio, G., Stevens, M.R., & Scott, J.W. 1999. Identification of RAPD markers linked to fusarium crown and root rot resistance (*Frl*) in tomato. *Euphytica* 105: 205–210.

Folkertsma, R.T., Spassova, M.I., Prins, M., Stevens, M.R., Hille, J. & Goldbach, R.W. 1999. Construction of a bacterial artificial chromosome (BAC) library of *Lycopersicon esculentum* cv. Stevens and its application to physically map the *Sw5* locus. *Mol. Breed.* 5: 197–207.

Foolad, M.R. & Sharma, A. 2005. Molecular markers as selection tools in tomato breeding. *Acta Hort.* 695: 225–240.

Foolad, M.R., Zhang, L.P., Khan, A.A., Nino-Liu, D. & Lin, G.Y. 2002. Identification of QTLs for early blight (*Alternaria solani*) resistance in tomato using backcross populations of a *Lycopersicon esculentum* x *L. hirsutum* cross. *Theor. Appl. Genet.* 104: 945–958.

Ganal, M.W., Simon, R., Brommonschenkel, S., Arndt, M., Tanksley, S.D. & Kumar, A. 1995. Genetic mapping of a wide spectrum nematode resistance gene (*Hero*) against *Globodera rostochiensis* in tomato. *Mol. Plant-Microbe Interact*. 8: 886–891.

Gardner, R.G. & Shoemaker, P.B. 1999. "Mountain Supreme" early blight-resistant hybrid tomato and its parents NC EBR-3 and NC EBR-4 tomato breeding lines. *Hort. Sci.* 34: 745–746.

Griffiths, P.D. & Scott, J.W. 2001. Inheritance and linkage of tomato mottle virus resistance genes derived from *Lycopersicon chilense* accession LA 1932. *J. Am. Soc. Hort. Sci.* 126: 462–467.

Grube, R.C., Radwanski, E.R. & Jahn, M. 2000. Comparative genetics of disease resistance within the *Solanaceae Genetics* 155: 873–887.

Hanson, P.M., Bernacchi, D., Green, S., Tanksley, S.D. & Muniy-Appa, V. 2000. Mapping a wild tomato introgression associated with tomato leaf curl virus resistance in a cultivated tomato line. *J. Am. Soc. Hort. Sci.* 125: 15–20.

Huang, C.C., van de Putte, P.M., Haanstra-van der Meer, J.G., Mejer-Dekens, F. & Lindhout, P. 2000. Characterization and mapping of resistance to Oidium lycopersicum in two *Lycopersicon hirsutum* accessions: evidence for close linkage of two Ol-genes on chromosome 6. *Heredity* 85: 511–520.

Hulbert, S.H., Webb, C.A., Smith, S.M. & Sun, Q. 2001. Resistance gene complexes: evolution and utilization. *Ann. Rev. Phytopathol*. 39: 285–312.

Jones, D.A., Dickinson, M.J., Balint-Kurt, P.J., Dixon, M.S. & Jones, J.D.G. 1993. Two complex resistance loci revealed in tomato by classical and RFLP mapping of *Cf-2*, *Cf-4*, *Cf-5*, and *Cf-9* genes for resistance to *Cladosporium fulvum*. *Mol. Plant-Microbe Interact*. 6: 348–357.

Kabelka, E., Franchino, B. & Francio, D.M. 2002. Two loci from Lycopersicon hirsutum LA407 confer resistance to strains of *Clavibacter michiganense* subsp. *michiganensis*. *Phytopath*. 92: 504–510.

Kawchuck, L.M., Hachey, J. & Lynch, D.R. 1998. Development of sequence characterized DNA markers linked to a dominant *Verticillium* wilt resistance gene in tomato. *Genome* 41: 91–95.

Kawchuk, L.M., Hachey, J., Lynch, D.R., Kulcsar, F., van Rooijen, G., Waterer, D.R., Robertson, A., Kokko, E., Byers, R., Howar, R.J., Fischer, R. & Prufer, D. 2001. Tomato *Ve* disease resistance genes encode cell surface-like receptors. *Proc. Nat. Acad. Sci. USA* 98: 6511–6515.

Konieczyn, A. & Ausubel, F.M. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4: 403–410.

Langella, R., Ercolano, M.R., Monti, L., Frusciante, L. & Barone, A. 2004 Molecular marker assisted transfer of resistance to TSWV in tomato elite lines. *J. Hort. Sci. & Biotech*. 79: 806–810

Laterrot, H. 1996. Breeding strategies for disease resistance in tomatoes with emphasis to the tropics: current status and research challenges, pp.126–132. *Proc.1st Internat. Symp. Trop. Tomato Dis*. Recife, Brazil.

Lefebvre, V. & Chèvre, A.M. 1995. Tools for marking plant disease and pest resistance genes: a review. *Agronomie* 15: 3–19.

Leister, D., Ballvora, A., Salamini, F. & Gebhardt, C. 1996. PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genet*. 14: 421–429.

Levesque, H., Vedel, F., Mathieu, C. & de Courcel, A.G.L. 1990. Identification of a short rDNA spacer sequence highly specific of a tomato line containing Tm-1 gene introgressed from *Lycopersicon hirsutum. Theor. Appl. Genet.* 80: 602–608.

Martin, G.B., Brommonschenkel, S., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D. & Tanksley, S.D. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262: 1432–1436.

Michelmore, R.W. 2003. The impact zone: genomics and breeding for durable disease resistance. *Curr. Opin. Plant Biol*. 6397–6404.

Moreau, P., Thoquet, P., Olivier J., Laterrot, H. & Grimsley, N.H. 1998. Genetic mapping of Ph-2, a single locus controlling partial resistance to Phytophthora infestans in tomato. *Mol. Plant-Microbe Interact.* 11: 259–268.

Morgante, M. & Salamini, F. 2003. From plant genomics to breeding practice. *Curr. Opin. Biotechnol*. 14: 214–219.

Ori, N., Eshed, Y., Paran, I., Prestino, G., Aviv, D., Tanksley, S.D., Zamir, D. & Fluhr, R. 1997. The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Plant Cell* 9: 521–532.

Parrella, G., Ruffel, S., Moretti, A., Morel, C., Palloix, A. & Caranta, C. 2002. Recessive resistance genes against potyviruses are localized in colinear genomic regions of the tomato (*Lycopersicon* spp.) and pepper (*Capsicum* spp.) genomes. *Theor. Appl. Genet*. 105: 855–861.

Parrella, G., Moretti, A., Gognalons, P., Lesage, M-L., Marchoux, G., Gebre-Selassie, K. & Carata, C. 2004. The Am gene controlling resistance to alfalfa mosaic virus in tomato is located in the cluster of dominant resistance genes on chromosome 6. *Phytopath*. 94: 345–350.

Peleman, J.D. & van der Voort, J.R. 2003. Breeding by design. *Trends Plant Sci.* 8: 330–334.

Pillen, K., Pineda, O., Lewis, C.B. & Tanksley, S.D. 1996. Status of genome mapping tools in the taxon *Solanaceae. In* A. Paterson, ed. *Genome mapping in plants*, pp. 281–308. Austin, TX, USA, Landes.

Rao, A.V. & Agarwal, S. 2000. Role of antioxidant lycopene in cancer and heart disease. *J. Am. Coll. Nutr*. 19: 563–569.

Robert, V.J.M., West, M.A.L., Inai, S., Caines, A., Arntzen, L., Smith, J.K. & St. Clair, D.A. 2001. Marker-assisted introgression of blackmold resistance QTL alleles from wild *Lycopersicon cheesmanii* to cultivated tomato (*L. esculentum*) and evaluation of QTL phenotypic effects. *Mol. Breed.* 8: 217–233.

Salmeron, J.M., Oldroyd, G.E.D., Rommens, C.M.T., Scofield, S.R., Kim, H.S., Lavello, D.T., Dahlbeck, D. & Staskawicz, B.J. 1996. Tomato Prf is a member of the leucine-rich repeat class of

plant disease resistance genes and lies embedded within the tomato Pto kinase gene cluster. *Cell* 86: 123–133.

Sandbrink, J.M., van-Ooijen, J.W., Purimahua, C.C., Vrielink, M., Verkerk, R., Zabel, P. & Lindhout, P. 1995. Localization of genes for bacterial canker resistance in *Lycopersicon peruvianum* using RFLPs. *Theor. Appl. Genet.* 90: 444–450.

Sarfatti, M., Abu-Abied, M., Katan, J. & Zamir, D. 1991. RFLP mapping of I1, a new locus in tomato conferring resistance to *Fusarium oxysporum* f. sp. *lycopersici* race 1. *Theor. Appl. Genet.* 80: 22–26.

Scott, J.W. 2005. Perspectives on tomato disease resistance breeding: past, present, and future. *Acta Hort*. 695: 217–224.

Sobir, Ohmori T., Murata, M. & Motoyoshi, F. 2000. Molecular characterization of the SCAR markers tightly linked to the *Tm-2* locus of the genus *Lycopersicon*. *Theor. Appl. Genet.* 101: 64–69.

Stamova, B.S. & Chetelat, R.T. 2000. Inheritance and genetic mapping of cucumber mosaic virus resistance introgressed from *Lycopersicon chilense* into tomato. *Theor. Appl. Genet.* 101: 527–537.

Stevens, M.R., Lamb, E.M. & Rhoads, D.D. 1995. Mapping the Sw-5 locus for tomato spotted wild virus resistance in tomatoes using RAPD and RFLP analyses. *Theor. Appl. Genet.* 90: 451–456.

Tanksley, S.D. & Costello, W. 1991. The size of *L. pennellii* chromosome 7 segment containing the I-3 gene in tomato breeding lines measured by RFLP probing. *Rep. Tomato Genet. Coop.* 41: 60.

Tanksley, S.D., Young, N.D., Paterson, A.H. & Bonierbale, M.W. 1989. RFLP mapping in plant breeding: new tools for an old science. *BioTech.* 7: 257–264.

Thoquet, P., Olivier, J., Sperisen, C., Rogowsky, P., Laterrot, H. & Grimsley, N.H. 1996. Quantitative trait loci determining resistance to bacterial wilt in tomato cultivar Hawaii 7996. *Mol. Plant-Microbe Interact*. 9: 826–836.

Vakalounakis, D.J., Laterrot, H., Moretti, A., Ligoxigakis, E.K. & Smardas, K. 1997. Linkage between *Fr-1* (*Fusarium oxysporum* f.sp. *radici-lycopersicy* resistance) and *Tm-2* (tobacco mosaic virus resistance-2) loci in tomato (*Lycopersicon esculentum*). *Ann. Appl. Biol*. 130: 319–323.

van der Biezen, E.A., Glagotlkaya, T. & Overduin, B. 1995. Inheritance and genetic mapping of resistance to *Alternaria alternata* f. sp. l*ycopersici* in *Lycopersicon pennellii*. *Mol. Gen. Genet.* 247: 453–461.

van Heusden, A.W., Koornneef, M., Voorrips, R.E., Bruggemann, W., Pet, G., Vrielink-van Ginkel, R., Chen, X. & Lindhout, P. 1999. Three QTLs from *Lycopersicon peruvianum* confer a high level of resistance to *Clavibacter michiganensis* subsp. *michiganensis*. *Theor. Appl. Genet.* 99: 1068–1074.

Varshney, R.K., Graner, A. & Sorrells, M.E. 2005. Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10: 621–630.

Williamson, V.M., Ho, J-Y., Wu, F.F., Miller, N. & Kaloshian, I. 1994. A PCR-based marker tightly linked to the nematode resistance, *Mi*, in tomato. *Theor. Appl. Genet.* 87: 757–763.

Yaghoobi, J., Kaloshian, I., Wen, Y. & Willimson, V.M. 1995. Mapping of a new nematode resistance locus in *Lycopersicon* peruvianum. *Theor. Appl. Genet.* 91: 457–464.

Yang, W., Bai, X., Kabelka, E., Eaton, C., Kamoun, S., van der Knapp, E. & Francis, D. 2004. Discovery of single nucleotide polymorphisms in *Lycopersicon esculetum* by computer aided analysis of expressed sequence tags. *Mol. Breed.* 14: 21–34.

Young, N.D. & Tanksley, S.D. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the Tm-2 gene in tomato. *Genetics* 120: 579–589.

Yu, Y.G., Wang, J.F., Stall, R.E. & Allejos, C.E. 1995. Genomic localization of tomato genes that control a hypersensitivity reaction to *Xanthomonas campestris* pv *vesicatoria* (Doidge) dye. *Genetics* 141: 675–682.

Zamir, D., Ekstein-Michelson, I., Zakay, U., Navot, N., Zeidan, M., Sarfatti, M., Eshed, Y., Harel, E., Pleban, T., van-Oss, H., Kedar, N., Rabinowitch, H.D. & Czosnek, H. 1994. Mapping and introgression of a tomato yellow leaf curl virus tolerance gene, Ty-1. *Theor. Appl. Genet.* 88: 141–146.

Zhang, L.P., Khan, A., Nino-Liu, D. & Foolad, M.R. 2002. A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum* x *Lycopersicon hirsutum* cross. *Genome* 45: 133–146.

Zhang, L.P., Lin, G.Y., Nino-Liu, D. & Foolad, M.R. 2003. Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* x *L. hirsutum* cross by selective genotyping. *Mol. Breed.* 12: 3–19.

SECTION III

Marker-assisted selection in livestock – case studies

CHAPTER 10

Strategies, limitations and opportunities for marker-assisted selection in livestock

Jack C.M. Dekkers and Julius H.J. van der Werf

SUMMARY

This chapter reviews the principles, opportunities and limitations for detection of quantitative trait loci (QTL) in livestock and for their use in genetic improvement programmes. Alternate strategies for QTL detection are discussed, as are methods for inclusion of marker and QTL information in genetic evaluation. Practical issues regarding implementation of marker-assisted selection (MAS) for selection in breed crosses and for selection within breeds are described, along with likely routes towards achieving that goal. Opportunities and challenges are also discussed for the use of molecular information for genetic improvement of livestock in developing countries.

INTRODUCTION

Since the 1970s, the discovery of technology that enables identification and genotyping of large numbers of genetic markers, and research that demonstrated how this technology could be used to identify genomic regions that control variation in quantitative traits and how the resulting QTL could be used to enhance selection, have raised high expectations for the application of gene- (GAS) or marker-assisted selection (MAS) in livestock. Yet, to date, the application of GAS or MAS in livestock has been limited (see e.g. review by Dekkers, 2004 and the case study chapters that follow). However, recent further advances in technology, combined with a substantial reduction in the cost of genotyping, have stimulated renewed interest in the large-scale application of MAS in livestock.

Successful application of MAS in breeding programmes requires advances in the following five areas:

- *Gene mapping*: identification and mapping of genes and genetic polymorphisms.
- *Marker genotyping*: genotyping of large numbers of individuals for large numbers of markers at a reasonable cost for both QTL detection and routine application for MAS.
- *QTL detection*: detection and estimation of associations of identified genes and genetic markers with economic traits.
- *Genetic evaluation*: integration of phenotypic and genotypic data in statistical methods to estimate breeding values of individuals in a breeding population.
- *MAS*: development of breeding strategies and programmes for the use of molecular genetic information in selection and mating programmes.

This chapter outlines the main strategies for the application of MAS in livestock and identifies and discusses the limitations and opportunities for successful MAS in commercial breeding programmes. It concludes by discussing limitations and opportunities for applying MAS in developing countries.

MARKERS AND LINKAGE DISEQUILIBRIUM

Over the past decades, a substantial number of alternate types of genetic markers have become available to study the genetic architecture of traits and for their use in MAS, including restriction fragment length polymorphisms (RFLPs), microsatellites, amplified fragment length polymorphisms (AFLPs) and single nucleotide polymorphisms (SNPs). Detailed information on these markers can be found elsewhere in this publication. Although alternate marker types have their own advantages and disadvantages, depending on their abundance in the genome, degree of polymorphism, and ease and cost of genotyping, what is crucial for their use for both QTL detection and MAS is the extent of linkage disequilibrium (LD) that they have in the population with loci that contribute to genetic variation for the trait. Linkage disequilibrium relates to dependence of alleles at different loci and is central to both QTL detection and MAS. Thus, a thorough understanding of LD and of the factors that affect the presence and extent of LD in populations is essential for a discussion of both QTL detection and MAS.

Linkage disequilibrium

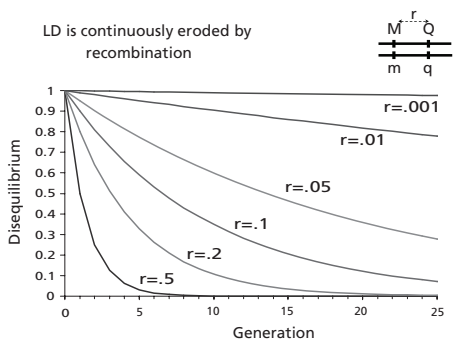
Consider a marker locus with alleles M and m and a QTL with alleles Q and q that is on the same chromosome as the marker, i.e. the marker and the QTL are linked. An individual that is heterozygous for both loci would have genotype $MmQq$. Alleles at the two loci are arranged in *haplotypes* on the two chromosomes of a homologous

pair that each individual carries. An individual with genotype *MmQq* could have the following two haplotypes: *MQ*/*mq*, where the / separates the two homologous chromosomes. Alternatively, it could carry the haplotypes *Mq*/*mQ*. This alternative arrangement of linked alleles on homologous chromosomes is referred to as the marker-QTL *linkage phase*. The arrangement of alleles in haplotypes is important because progeny inherit one of the two haplotypes that a parent carries, barring recombination.

The presence of linkage equilibrium (LE) or disequilibrium relates to the relative frequencies of alternative haplotypes in the population. In a population that is in linkage *equilibrium*, alleles at two loci are randomly assorted into haplotypes. In other words, chromosomes or haplotypes that carry marker allele *M* are no more likely to carry QTL allele *Q* than chromosomes that carry marker allele *m*. In technical terms, the frequency of the *MQ* haplotypes is equal to the product of the population allele frequency of *M* and the frequency of *Q*. Thus, if a marker and QTL are in linkage *equilibrium*, there is no value in knowing an individual's marker genotype because it provides no information on QTL genotype. If the marker and QTL are in linkage *dis*equilibrium, however, there will be a difference in the probability of carrying *Q* between chromosomes that carry *M* and *m* marker alleles and, therefore, a difference in mean phenotype between marker genotypes would also be expected.

The main factors that create LD in a population are mutation, selection, drift (inbreeding), and migration or crossing. See Goddard and Meuwissen (2005) for further background on these topics. The main factor that breaks down LD is recombination, which can rearrange haplotypes that

FIGURE 1
Break-up of LD over generations

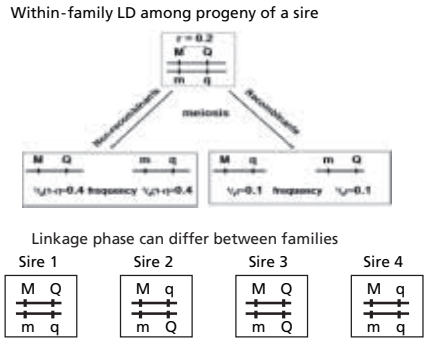


exist within a parent in every generation. Figure 1 shows the effect of recombination (r) on the decay of LD over generations. The rate of decay depends on the rate of recombination between the loci. For tightly linked loci, any LD that has been created will persist over many generations but, for loosely linked loci (r > 0.1), LD will decline rapidly over generations.

Population-wide versus within-family LD

Although a marker and a linked QTL may be in LE across the population, LD will always exist *within* a family, even between loosely linked loci. Consider a double heterozygous sire with haplotypes *MQ*/*mq* (Figure 2). The genotype of this sire is identical to that of an F_1 cross between inbred lines. This sire will produce four types of gametes: non-recombinants *MQ* and *mq* and recombinants *Mq* and *mQ*. As non-recombinants will have higher frequency, depending on the recombination rate between the marker and QTL, this sire will produce gametes that will be in LD. Furthermore, this LD will extend over a larger distance (Figure 1), because it has undergone only one generation of recombination. This specific type of LD,

FIGURE 2
Within-family marker-QTL LD



however, only exists within this family; progeny from another sire, e.g. an *Mq*/*mQ* sire, will also show LD, but the LD is in the opposite direction because of the different marker-QTL linkage phase in the sire (Figure 2). On the other hand, *MQ*/*mQ* and *Mq*/*mq* sire families will not be in LD because the QTL does not segregate in these families. When pooled across families these four types of LD will cancel each other out, resulting in linkage *equilibrium* across the population. Nevertheless, the within-family LD can be used to detect QTL and for MAS provided the differences in linkage phase are taken into account, as will be demonstrated later.

QTL DETECTION AND TYPES OF MARKERS FOR MAS

Application of molecular genetics for genetic improvement relies on the ability to genotype individuals for specific genetic loci. For these purposes, three types of observable genetic loci can be distinguished, as described by Dekkers, 2004:

- direct markers: loci for which the functional polymorphism can be genotyped;
- LD-markers: loci in population-wide LD with the functional mutation;
- LE-markers: loci in population-wide linkage equilibrium with the functional mutation but which can be used for QTL detection and MAS based on within-family LD.

For these alternate types of markers, different strategies are appropriate to detect QTL in livestock populations. These are summarized in Table 1 and will be described in more detail. Strategies for QTL detection in livestock differ from those used in plants because of the lack of inbred lines.

QTL detection using LD markers within crosses

Crossing two breeds that differ in allele and, therefore, haplotype frequencies, creates extensive LD in the crossbred population. This LD extends over large distances

TABLE 1
Summary of strategies for QTL detection in livestock

Type of population	Within crosses		Outbred population			
	F2/Backcross	Advanced intercross	Half- or full-sib families	Extended pedigree	Non-pedigreed population sample	
Type of markers	LD markers		LE markers		LD markers	
Genome coverage	Genome-wide		Genome-wide		Candidate gene regions	Genome-wide
Marker density	Sparse	Denser	Sparse	More dense	Few loci	Dense
Type of LD used	Population-wide LD		Within-family LD		Population-wide LD	
Number of generations of recombination used for mapping	1	>1	1	>1	>>1	
Extent of LD around QTL	Long	Smaller	Long	Smaller	Small	
Map resolution	Poor	Better	Poor	Better	High	

because it has undergone only one generation of recombination in the F_2 (Figure 1). Thus, although these markers may be in LE with QTL within the parental breeds, they will be in partial LD with the QTL in the crossbred population if the marker and QTL differ in frequency between the breeds. This population-wide LD enables detection of QTL that differ between the parental breeds based on a genome scan with only a limited number of markers spread over the genome (~ every 15 to 20 cM). This approach has formed the basis for the extensive use of F_2 or backcrosses between breeds or lines for QTL detection, in particular in pigs, poultry and beef cattle (see Andersson, 2001 for a review). The extensive LD enables detection of QTL that are some distance from the markers but also limits the accuracy (map resolution) with which the position of the QTL can be determined.

More extensive population-wide LD is also expected to exist in synthetic lines, i.e. lines that were created from a cross in recent history. These can be set up on an experimental basis through advanced intercross lines (Darvasi and Soller, 1995) or be available as commercial breeding lines. Depending on the number of generations since the cross, the extent of LD will have eroded over generations and will, therefore, span shorter distances than in F_2 populations (Figure 1). This will require a more dense marker map to scan the genome with equivalent power as in an F_2 but will enable more precise positioning of the QTL.

QTL detection using LE markers in outbred populations

As linkage phases between the marker and QTL can differ from family to family, use of within-family LD for QTL detection requires QTL effects to be fitted on a within-family basis, rather than across the population. Similar to F_2 or backcrosses, the extent of within-family LD is extensive and, thus, genome-wide coverage is provided by a limited number of markers but significant markers may be some distance from the QTL, resulting in poor map resolution. Thus, LE markers can be readily detected on a genome-wide basis using large half-sib families, requiring only sparse marker maps (~15 to 20 cM spacing). Many examples of successful applications of this methodology for detection of QTL regions are available in the literature, in particular for dairy cattle, utilizing the large paternal half-sib structures that are available through extensive use of artificial insemination (see Weller, Chapter 12).

QTL detection using LE markers can also be applied to extended pedigrees by modelling the co-segregation of markers and QTL (Fernando and Grossman, 1989). These approaches use statistical models that are described further in the section on genetic evaluation using LE markers. Depending on the number of generations with phenotypes and marker genotypes that are included in the analysis, map resolution will be better than with analysis of half-sib families because multiple rounds of recombination are included in the data set.

QTL detection using LD markers in outbred populations

The amount and extent of LD that exists in the populations that are used for genetic improvement are the net result of all forces that create and break down LD and are, therefore, the result of the breeding and selection history of each population, along with random sampling. On this basis, populations that have been closed for many generations are expected to be in linkage *equilibrium*, except for closely linked loci.

Thus, in those populations, only markers that are tightly linked to QTL may show an association with phenotype (Figure 1), and even then there is no guarantee because of the chance effects of random sampling.

There are two strategies to find markers that are in population-wide LD with QTL (see Table 1):

- evaluating markers that are in, or close to, genes that are thought to be associated with the trait of interest (candidate genes);
- a genome scan using a high-density marker map, with a marker every 0.5 to 2 cM.

The success of both approaches obviously depends on the extent of LD in the population. Studies in human populations have generally found that LD extends over less than 1 cM. Thus, many markers are needed to obtain sufficient marker coverage in human populations to enable detection of QTL based on population-wide LD. Opportunities to utilize population-wide LD to detect QTL in livestock populations may be considerably greater because of the effects of selection and inbreeding. Indeed, Farnir *et al.* (2000) identified substantial LD in the Dutch Holstein population, which extended over 5 cM. Similar results have been observed in other livestock species (e.g. in poultry, Heifetz *et al.*, 2005). The presence of extensive LD in livestock populations is advantageous for QTL detection, but disadvantageous for identifying the causative mutations of these QTL; with extensive LD, markers that are some distance from the causative mutation can show an association with phenotype.

The candidate gene approach utilizes knowledge from species that are rich in genome information (e.g. human, mouse), effects of mutations in other species, previously identified QTL regions, and/or knowledge of the physiological basis of traits, to identify genes that are thought to play a role in the physiology of the trait. Following mapping and identification of polymorphisms within the gene, associations of genotype at the candidate gene with phenotype can be estimated (Rothschild and Plastow, 1999).

Whereas the candidate gene approach focuses on LD within chosen regions of the genome, recent advances in genome technology have enabled sequencing of entire genomes, including of several livestock species; the genomes of the chicken and cattle have been sequenced and public sequencing of the genome of the pig is under way. In addition, sequencing has been used to identify large numbers of positions in the genome that include SNPs, i.e. DNA base positions that show variation. For example, in the chicken, over 2.8 million SNPs were identified by comparing the sequence of the Red Jungle Fowl with that of three domesticated breeds (International Chicken Polymorphism Map Consortium, 2004). This, combined with reducing costs of genotyping, now enables detection of QTL using LD-mapping with high-density marker maps.

QTL detection using combined LD and linkage analysis in outbred populations

As markers may not be in complete LD with the QTL, both population-wide associations of markers with QTL and co-segregation of markers and QTL within families can be used to detect QTL. Using these combined properties of being both LD and LE markers, methods have been developed to combine LD and linkage information. These methods are further explored under genetic evaluation models in what follows.

INCORPORATING MARKER INFORMATION IN GENETIC EVALUATION PROGRAMMES

The value of genotypic information for predicting the genetic merit of animals is dependent on the predictive ability of the marker genotypes. The three types of molecular loci described previously differ not only in methods of detection but also in methods of their incorporation in genetic evaluation procedures. Whereas direct and, to a lesser degree, LD markers, allow selection on genotype across the population, use of LE markers must allow for different linkage phases between markers and QTL from family to family, i.e. LE markers are family specific and family specific information must be derived. As discussed later in this chapter, this makes LE markers a lot less attractive for use in breeding programmes. In this section, the different types of models that have been proposed for genetic evaluation based on marker information are described and this is followed by a brief description of some practical issues regarding implementation of such methods and the likely routes towards achieving that goal.

Modelling QTL effects in genetic evaluation

By using QTL information in genetic evaluation, in principle, part of the assumed polygenic variation is substituted by a separate effect due to a genetic polymorphism at a known locus. This has the immediate effect of having a much better handle on the Mendelian sampling process, as phenotypic co-variance can be evaluated based on specific genetic similarity rather than on an average relationship. For example, on average two full sibs share 50 percent of their alleles, but at a specific locus it is now possible to know whether these full sibs carry exactly the same complete genotype (both paternal and maternal alleles are in common), or actually have a completely different genotype. The actual degree of similarity of full sibs at a QTL can thus vary between 0 and 1. This additional information helps to better evaluate the genetic merit due to specific QTL, and to better predict offspring that do not yet have phenotypic measurements.

A number of different approaches have been described to accommodate marker information in genetic evaluation. Roughly, these methods can be distinguished through their modelling of the QTL effect and through the type of genetic marker information used. The QTL effect can be modelled as random or fixed, while the molecular information comes from LE, LD or direct markers.

With a fixed QTL model, regression on genotype probabilities would be used in genetic evaluation to account for the effect of QTL polymorphisms. In the simplest additive QTL model, suitable for estimating breeding values, simple regressions could be included on the probability of carrying the favourable mutation. Regression can be on known genotypes (class variables), or probabilities can be derived for ungenotyped animals in a general complex pedigree (Kinghorn, 1999). A fixed QTL model is sensible if few alleles are known to be segregating, and where dominance and/or epistasis are important. The model also assumes effects being the same across families. The effects of various genotypes could be fitted separately, giving power to account for dominance and epistasis in case of multiple QTL. For selection purposes, a fixed QTL effect, if additive, would be added to the polygenic estimated breeding values (EBVs), similar to breed effects in across-breed evaluations. The advantage of a fixed QTL model is the limited number of effects that need to be fitted.

Alternatively, QTL effects could be modelled as random effects, with each individual having a different QTL effect. Co-variances are based on the probability of QTL alleles being identical by descent rather than on numerator relationships as in the usual animal model with polygenic effects. With full knowledge about segregation, this would effectively fit all founder alleles as different effects. The random QTL model was first described by Fernando and Grossman (1989), where for each animal both the paternal and the maternal allele were fitted. Without loss of information, these effects can be collapsed into one genotypic effect for each animal (Pong-Wong *et al.,* 2001). The random QTL model makes no assumptions about number of alleles at a QTL and it automatically accommodates possible interaction effects of QTL with genetic background (families or lines). Therefore, the random QTL model is less reliant on assumptions about homogeneity of QTL effects. The random QTL model is a natural extension to the usual mixed model and seems therefore a logical way to incorporate genotype information into an overall genetic evaluation system. These models result in EBVs for QTL effects along with a polygenic EBV. The total EBV is the simple sum of these estimates. One of the main computational limitations of this method, however, is the large number of equations that must be solved, which increases by two per animal for each QTL that is fitted. Thus, the number of QTL regions that can be incorporated is limited.

Genetic evaluation using direct markers

When the genotype of an actual functional mutation is available, no pedigree information is needed to predict the genotypic effect, as QTL genotypes are measured directly. When there is only a small number of alleles, the number of specific genotypes is limited. In genetic evaluation, it would seem appropriate to treat the genotype effect as a fixed effect, i.e. the assumption is that genotype differences are the same in different families and herds or flocks. Such assumptions might be reasonable for a bi-allelic QTL model in a relatively homogeneous population. Alternatively, random QTL models could be used with different effects for different founder alleles, or even QTL by environment interactions. In both fixed and random QTL models, genotype probabilities can be derived for individuals with missing genotypes.

Genetic evaluation using LE markers

When the genotype test is not for the gene itself, but for a linked marker, QTL probabilities derived from marker genotypes will be affected by the recombination rate between marker and QTL and by the extent of LD between the QTL and marker across the population. If LD between the QTL and a linked marker only exists within families, marker effects or, at a minimum, marker-QTL linkage phase must be determined separately for each family. This requires marker genotypes and phenotypes on family members. If linkage between the marker and QTL is loose, phenotypic records must be from close relatives of the selection candidate because associations will erode quickly through recombination. With progeny data, marker-QTL effects or linkage phases can be determined based on simple statistical tests that contrast the mean phenotype of progeny that inherited alternate marker alleles from the common parent. A more comprehensive approach is based on Fernando and Grossman's (1989) random QTL model, where marker information from complex pedigrees can be used

to derive co-variances between QTL effects, yielding best linear unbiased prediction (BLUP) of breeding value for both polygenic and QTL effects. Random effects of paternal and maternal QTL alleles are added to the standard animal model with random polygenic breeding values. The variance-co-variance structure of the random QTL effects, also known as the gametic relationship matrix (GRM), is based on probabilities of identity by descent (IBD), and is now derived from co-segregation of markers and QTL within a family. Probabilities of IBD derived from pedigree and marker data link QTL allele effects that are expected to be equal or similar, therefore using data from relatives to estimate an individual's QTL effects. For example, if two paternal half-sibs i and j have inherited the same paternal allele for markers that flank the QTL (with recombination rate r), they are likely IBD for the paternal QTL allele and the correlation between the effects of their paternal QTL alleles will be $(1\text{-}r)^2$. The method is appealing, but computationally demanding for large-scale evaluations, especially when not all animals are genotyped and complex procedures must be applied to derive IBD probabilities.

Genetic evaluation using LD markers

Most QTL projects have moved towards fine mapping where the final result is a marker or marker haplotype in LD with the QTL, if not the direct mutation. A haplotype of marker alleles close enough to the putative QTL is likely to be in LD with QTL alleles. Such a marker test provides information about QTL genotype across families, and is in a sense not very different from a direct marker. The most convenient way to include genotypic information from marker haplotypes in genetic evaluation systems is through the random QTL model. In their original paper, Fernando and Grossman (1989) derived IBD from genotype data on single markers and recombination rates between marker and QTL. However, the random QTL model is more versatile, and co-variances based on IBD probabilities can also use information beyond pedigree, based on LD. The latter can be derived from marker or haplotype similarity, e.g. based on a number of marker genotypes surrounding a putative QTL. Meuwissen and Goddard (2001) proposed using both linkage and LD information to derive IBD-based co-variances (termed LDL analysis). Lee and van der Werf (2005) showed that with denser markers, the value of linkage information, and therefore pedigree, reduces. Hence, when QTL positions become more accurately defined, genetic information from close markers (within a few cM) can be used increasingly to derive LD-based IBD probabilities, thereby defining co-variances between random QTL effects without the need for a family structure or information through pedigree.

Lee and van der Werf (2006) have shown that LD information results in a very dense GRM. Genetic evaluation, which is usually based on mixed model equations that are relatively sparse, is currently not feasible computationally for the LDL method for a large number of individuals and alternative models are needed. One approach is to model population-wide LD by simply including the marker genotype or haplotype as a fixed effect in the animal model evaluation, as suggested by Fernando (2004). An advantage of modelling population-wide LD effects as fixed rather than random is that fewer assumptions about population history are needed. A disadvantage is that estimates are not "BLUPed", i.e. regressed towards a mean depending on

the amount of information that is available to estimate their effects. This will be important if some of the genotype or haplotype effects cannot be estimated with substantial accuracy because the number of individuals with that genotype or haplotype is limited. Haplotype effects could also be fitted as random, but more development is needed in this area.

Whole genome approach for genetic evaluation using high-density LD markers

With more and more QTL being discovered, the polygenic component will slowly be replaced by multiple QTL effects, the inheritance of each being followed by marker brackets or more generally by information on haplotypes. Nejati-Javaremi, Smith and Gibson (1997) presented the concept of the total allelic relationship, where the co-variance between two individuals was derived from allelic identity by descent, or by state (based on molecular marker information), with each location weighted by the variance explained by that region. This approach contrasts with the average relationships derived from pedigree that are used in the numerator relationship matrix. Nejati-Javaremi, Smith and Gibson (1997) showed that using total allelic relationship resulted in a higher selection response than pedigree based relationships, because it more accurately accounts for the variation in the additive genetic relationships between individuals. Therefore, the gain of following inheritance at specific genome locations contributes to more accurate genetic evaluation, and is able to deal more specifically with within and between loci interactions and with specific modes of inheritance at different QTL.

When large-scale marker genotyping becomes cheap and available to breeders at low cost, this approach could even be used for non-detected QTL and genetic evaluation could be based on a "whole genome approach" (Meuwissen, Hayes and Goddard, 2001). In this approach, marker haplotypes are fitted as independent random effects for each, e.g. 1 cM region of the genome. In the work by Meuwissen, Hayes and Goddard (2001), variances associated with each haplotype were either assumed to be equal for each chromosomal region or estimated from the data using Bayesian procedures with alternate prior distributions. In essence, this procedure estimates breeding values for each haplotype, and EBVs of individuals are computed by simply summing EBVs for the haplotypes that they contain.

Using this procedure, Meuwissen, Hayes and Goddard (2001) demonstrated through simulation, that for populations with an effective population size of 100 and a spacing of 1 or 2 cM between informative markers across the genome, sufficient LD was present to predict genetic values with substantial accuracy for several generations based on associations of marker haplotypes with phenotype on as few as 500 individuals. It should be noted that, in the approach proposed by these authors, no polygenic effect is included since all regions of the genome are included in the model. It may, however, be useful to include a polygenic effect because LD between markers and QTL will not be complete for all regions. In addition, this model assumes that haplotype effects are independent within and across regions. Incorporating IBD probabilities to model co-variances between haplotypes within a region as in Meuwissen and Goddard (2000), and by incorporating co-variances between adjacent regions caused by LD between regions, could lead to further improvements but would also lead to increasing computational demands.

In general, for the purpose of increased genetic change of economically important quantitative traits, and in the context of well recorded and efficient breeding programmes, there is no need to have knowledge of functional mutations since nearby markers will have a high predictive value about genetic merit. Moreover, the benefit from the extra investment and time spent on finding functional mutations might be superseded by the genetic change that can be made in the breeding programme in the meantime.

Implementation of marker-assisted genetic evaluation

It is important to note that, for most of the gene marker tests currently on the market, integration with existing systems for genetic evaluation is not obvious. This is because the gene testing is either for a Mendelian characteristic, or it predicts phenotypic differences for traits that are not the same as those in current genetic evaluation. Moreover, breeders would not only be interested in more accurate EBVs based on gene markers, but they would also want to know the actual QTL genotypes for their breeding animals. This information on individual genotype will become less relevant if more gene tests become available and if testing becomes cheaper and more widespread. This might still take some years. Thus, as gene marker testing is gradually introduced, it is more likely to create additional selection criteria to consider and it will take some time before QTL information is seamlessly and optimally integrated in existing genetic evaluation programmes. In particular, if genetic evaluation is based on information from many different breeding units, such as in cattle or sheep, genotyping information will initially be available for only a small proportion of the breeding animals, possibly not justifying a total overhaul of the system for genetic evaluation. Simple ad hoc procedures where QTL effects are estimated and presented separately as additional effects are initially a more likely route to implementation.

Solutions for fixed QTL genotype effects, along with genotype probabilities as outputs of genetic evaluation, might be interesting to breeders and, compared with random QTL effects, may be more likely to be presented and used separately from polygenic EBVs. This would also be the case for genotypic information on Mendelian characters, where there is no polygenic component.

INCORPORATING MAS IN SELECTION PROGRAMMES

Molecular information can be used to enhance both the processes of integrating superior qualities of different breeds and within-breed selection. These strategies are further described below.

Between-breed selection

Crossing breeds results in extensive LD, which can be capitalized upon using MAS in a number of ways. If a large proportion of breed differences in the trait(s) of interest are due to a small number of genes, gene introgression strategies can be used. If a larger number of genes is involved, MAS within a synthetic line is the preferred method of improvement.

Marker-assisted introgression

Introgression of the desirable allele at a target gene from a donor to a recipient breed is accomplished by multiple backcrosses to the recipient, followed by one or more generations of intercrossing. The aim of the backcross generations is to produce

individuals that carry one copy of the donor QTL allele but that are similar to the recipient breed for the rest of the genome. The aim of the intercrossing phase is to fix the donor allele at the QTL. Marker information can enhance the effectiveness of the backcrossing phase of gene introgression strategies by: (i) identifying carriers of the target gene(s) (foreground selection); and (ii) enhancing recovery of the recipient genetic background (background selection). The effectiveness of the intercrossing phase can also be enhanced through foreground selection on the target gene(s). If the target gene cannot be genotyped directly, carrier individuals can be identified based on markers that flank the QTL at <10 cM, because of the extensive LD in crosses. The markers must have breed-specific alleles in order to identify line origin. For the introgression of multiple target genes, gene pyramiding strategies can be used during the backcrossing phase to reduce the number of individuals required (Hospital and Charcosset, 1997; Koudandé *et al.,* 2000). For background selection, markers are used that are spread over the genome at <20 cM intervals, such that most genes that affect the trait will be within 10 cM from a marker. Combining foreground and background selection, selection will be for the donor breed segment around the target locus but for recipient breed segments in the rest of the genome. Foreground selection will result in selection for both the target locus and for donor breed loci that are linked to this locus, some of which could have an unfavourable effect on performance. To reduce this so-called linkage drag around the target locus, in the molecular score used for background selection greater emphasis can be given to markers that are in the neighbourhood of the target locus (apart from the flanking markers, which are used in foreground selection).

Most studies have considered marker-assisted introgression (MAI) of single QTL (e.g. Hospital and Charcosset, 1997) but often several QTL must be introgressed simultaneously. Koudandé *et al.* (2000) showed that large populations are needed to obtain sufficient individuals that are heterozygous for all QTL in the backcrossing phase. This would make MAI not feasible in livestock breeding programmes. In many cases, however, immediate fixation of introgressed QTL alleles may not be required. Instead, the objective of the backcrossing phase can be to enrich the recipient breed with the favourable donor QTL alleles at sufficiently high frequency for selection following backcrossing. The effectiveness of such strategies was demonstrated by Chaiwong *et al.* (2002).

Marker-assisted improvement of synthetic lines

In MAI studies it is usually assumed that the aim is to recover the recipient breed genotype, except for the donor QTL. An alternative objective could be to aim simply for individuals with highest merit. Selection would then be for QTL genotype as well as EBV, estimated across breeds or lines. This EBV selection would replace background selection, as recovery of the recipient genotype is achieved through selection on genetic merit rather than through selecting for breed of origin. This strategy would be more competitive if the original breeds overlap in merit, and indeed, as was shown by Dominik *et al.* (2007), background selection based on anonymous markers would be less profitable.

Strategies for using markers to select within a hybrid population were first proposed by Lande and Thompson (1990). These strategies capitalize on population-wide LD that initially exists in crosses

between lines or breeds. Thus, marker-QTL associations identified in the F_2 generation can be selected for several generations, until the QTL or markers are fixed or the disequilibrium disappears. Zhang and Smith (1992) evaluated the use of markers in such a situation with selection on BLUP EBV. Although both studies considered the ideal situation of a cross with inbred lines, there will be opportunities to utilize a limited number of markers to select for favourable QTL regions that are detected in crosses between breeds, thereby enhancing the development of superior synthetics. Pyasatian, Fernando and Dekkers (2006) investigated use of the whole genome approach of Meuwissen, Hayes and Goddard (2001) for MAS in a cross by including all markers as random effects in the model for genetic evaluation. They showed that this resulted in substantially greater responses to selection than selection on identified QTL regions only. Due to the much greater LD, whole genome selection in a cross can be accomplished with a much smaller number of markers compared with the number required for whole genome selection in an outbred population.

Within-breed selection

The procedures described previously for incorporating markers in genetic evaluation result in estimates of breeding values associated for QTL, together with estimates of polygenic breeding values. Alternatively, if molecular data are not incorporated into genetic evaluations, as will be the case for more ad hoc approaches and for gene tests for Mendelian characteristics, separate selection criteria will be available that capture the molecular information. The following three selection strategies can then be distinguished (Dekkers, 2004):

- select on the QTL information alone;
- tandem selection, with selection on QTL followed by selection on polygenic EBV;
- selection on the sum of the QTL and polygenic EBV.

Selection on QTL or marker information alone ignores information that is available on all other genes (polygenes) that affect the trait and is expected to result in the lowest response to selection unless all genes that affect the trait are included in the QTL EBV. This strategy does not, however, require additional phenotypes other than those that are needed to estimate marker effects, and can be attractive when phenotype is difficult or expensive to record (e.g. disease traits, meat quality, etc.). Selection on the sum of the QTL and polygenic EBV is expected to result in maximum response in the short term, but may be suboptimal in the longer term because of losses in polygenic response (Gibson, 1994). Indexes of QTL and polygenic EBV can be derived that maximize longer-term response (Dekkers and van Arendonk, 1998) or a combination of short- and longer-term responses (Dekkers and Chakraborty, 2001). However, if selection is on multiple QTL and emphasis is on maximizing shorter-term response, selection on the sum of QTL and polygenic EBV is expected to be close to optimal. Optimizing selection on a number of EBVs, indexes and genotypes, while also considering inbreeding rate and other practical considerations is not a trivial task. Kinghorn, Meszaros and Vagg (2002) have proposed a mate selection approach that could be used to handle such problems, and it can be expected that with more widespread use of genotypic information for a larger number of regions, specific knowledge about individual QTL becomes less interesting and will simply contribute to prediction of whole EBV or whole genotype.

Meuwissen and Goddard (1996) published a simulation study that looked at the main characteristics determining efficiency of MAS using LE markers. They found that MAS could improve the rate of genetic improvement up to 64 percent by selecting on the sum of QTL and polygenic EBV. Their work also demonstrated that MAS is mainly useful for traits where phenotypic measurement is less valuable because of: (i) low heritability; (ii) sex-limited expression; (iii) availability only after sexual maturity; and (iv) necessity to sacrifice the animal (e.g. slaughter traits). Selection of animals based on (most probable) QTL genotype will allow earlier and more accurate selection, increasing the short- and medium-term selection response.

Most simulation studies have assumed complete marker genotype information but in practice only a limited number of individuals will be genotyped. However, in an advanced breeding programme with complete information on phenotype and pedigree information, marker and QTL genotype probabilities could be derived for un-genotyped animals and genotyping strategies could be optimized to achieve a high value for the investments made. Marshall, Henshall and van der Werf (2002) looked at strategies to minimize genotyping cost in a sheep breeding programme. Close to maximal gain could be achieved when genotyping was undertaken only for high ranking males and animals whose marker genotype probability could not be derived with enough certainty based on information on relatives. Marshall, van der Werf and Henshall (2004) also looked at progeny testing of sires to determine family-specific marker-QTL phase within a breeding nucleus. Again, testing of a limited number of males provided a lot of information about phase for several generations of breeding animals, as progeny tested sires have relationships with descendants. However, in breeding programmes for more extensive production systems (beef, sheep), pedigree recording is often incomplete and only a small proportion of animals are genotyped. Moreover, these genotyped animals are not necessarily the key breeding animals. The utility of linked markers will be even more limited if pedigree relationships cannot be used to resolve genotype probabilities and marker-QTL phase of un-genotyped individuals.

A second point of caution is that many studies on MAS have taken a single-trait approach and shown that genetic markers could have a large impact on responses for traits that are difficult to improve by phenotypic selection. However, within the context of a multitrait breeding objective, the overall impact of such markers on the breeding goal may be less because a greater response for one trait often appears at the expense of another. For example, genetic markers for carcass traits improve the ability to select (i.e. earlier, with higher accuracy) for such traits, but selection emphasis for other traits is reduced. Therefore, the overall effect of MAS on the breeding programme will generally be much smaller than predicted for single trait MAS-favourable cases. The main effects of MAS would be to shift the selection response in favour of the marked traits, rather than achieving much additional overall response. Hence, while it will be easier to select for carcass and disease resistance, further improvement for these traits will be at the expense of genetic change for production traits (growth, milk).

The impact of MAS on the rate of genetic gain may be limited in conventional breeding programmes (ranging up to perhaps 10 percent extra gain) unless

the variation in profitability is dominated by traits that are hard to measure. However, new technologies often lead to other breeding programme designs being closer to optimal. Genotypic information has extra value in the case of early selection and where within-family variance can be exploited, which is particularly the case in programmes where reproductive technologies are used. Reproductive technologies usually lead to early selection and more emphasis on between-family selection. DNA marker technology and reproductive technologies are therefore highly synergistic and complementary (van der Werf and Marshall, 2005) and gene markers have much more value in such programmes. Gene marker information is also clearly valuable in introgression programmes, as demonstrated by simulation (Chaiwong *et al.*, 2002; Dominik *et al.*, 2006) as well as in practice (Nimbkar, Pardeshi and Ghalsasi, 2005). Yet, although these examples are favourable to the value of gene marker information, the added value of MAS still relies heavily on a high degree of trait and pedigree recording.

OPPORTUNITIES FOR MAS IN DEVELOPING COUNTRIES

Complete phenotypic and pedigree information is often only available in intensive breeding units. Therefore, in the context of low input production systems, some questions can be raised concerning the validity and practicality of the simulation studies described above, and it would be more difficult to realize the value of marker information. It would be harder and more expensive to determine the linkage phase in the case of using linked markers. Moreover, even if the genetic marker were a direct or LD marker, its effect on phenotype would have to be estimated for the population and the environment in which it is used. This would require phenotypes and genotypes on a sample of a rather homogeneous population to avoid spurious associations that could result from unknown population stratification. Therefore, a gene marker for a QTL is likely to be most successful in an environment with intensive pedigree and performance recording. Nevertheless, in low input environments, direct and LD markers will be more useful than LE markers because the latter require routine recording of phenotypes and genotypes to estimate QTL effects within families.

In addition to MAS within local breeds, several other strategies for breed improvement could be pursued in developing countries, including gene introgression and MAS within synthetic breeds. This would be most advantageous for introducing specific disease resistance alleles into breeds with improved production characteristics to make them more tolerant to the environments encountered in developing countries. Gene introgression is, however, a long and expensive process and only worthwhile for genes with large effects. MAS within synthetic breeds, e.g. a cross between local and improved temperate climate breeds, can allow development of a breed that is based on the best of both breeds (e.g. Zhang and Smith, 1992). Because of the extensive LD within the cross, a limited number of markers would be needed. Care should, however, be taken to avoid the impact of genotype x environment interactions if MAS is implemented in a more controlled environment.

REFERENCES

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nature Revs. Genet*. 2: 130–138.

Chaiwong, N., Dekkers, J.C.M., Fernando, R.L. & Rothschild, M.F. 2002. Introgressing multiple QTL in backcross breeding programs of limited size. *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prodn*. Electronic Communication No. 22: 08. Montpellier, France.

Darvasi, A. & Soller, M. 1995. Advanced intercross lines: an experimental population for fine genetic mapping. *Genetics* 141: 1199–1207.

Dekkers, J.C.M. & van Arendonk, J.A.M. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations, *Genet. Res*. 71: 257–275.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci*. 82: E313–E328.

Dekkers J.C.M. & Chakraborty, R. 2001. Potential gain from optimizing multi-generation selection on an identified quantitative trait locus. *J. Anim. Sci.* 79: 2975–2990.

Dominik, S., Henshall, J., O'Grady, J. & Marshall, K.J. 2007. Factors influencing the efficiency of a marker assisted introgression program in merino sheep. *Genet. Sel. Evol*. In press.

Farnir, F., Coppieters, W., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res*. 10: 220–227.

Fernando, R.L. 2004. Incorporating molecular markers into genetic evaluation. Session G6.1. *Proc. 55th Meeting of the European Association of Animal Production*. 5–9 September 2004, Bled, Slovenia.

Fernando, R.L. & Grossman, M. 1989. Marker-assisted selection using best linear unbiased prediction. *Genet. Sel. Evol*. 21: 467–477.

Gibson, J.P. 1994. Short-term gain at the expense of long-term response with selection of identified loci, *Proc. 5th Wld. Cong. Genet. Appl. Livest. Prodn.* CD-ROM Communication No. 21: 201–204. University of Guelph, Canada.

Goddard, M.E. & Meuwissen, T.H.E. 2005. The use of linkage disequilibrium to map quantitative trait loci. *Austr. J. Exp. Agric*. 45: 837–845.

Heifetz, E.M., Fulton, J. E., O'Sullivan, N., Zhao, H., Dekkers, J.C.M. & Soller, M. 2005. Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* 171: 1173–1181.

Hospital, F. & Charcosset, A. 1997. Marker-assisted introgression of quantitative trait loci. *Genetics* 147: 1469–1485.

International Chicken Polymorphism Map Consortium. 2004. A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms. *Nature* 432: 717–722.

Kinghorn, B.P. 1999. Use of segregation analysis to reduce genotyping costs. *J. Anim. Breed. Genet.* 116: 175–180.

Kinghorn, B.P., Meszaros, S.A. & Vagg, R.D. 2002. Dynamic tactical decision systems for animal breeding, *Proc. 7th Wld. Congr. Genet. Appl. Livest. Prodn*. Communication No. 23-07. Montpellier, France.

Koudandé, O.D., Iraqi, F., Thomson, P.C., Teale, A.J. & van Arendonk, J.A.M. 2000. Strategies to optimize marker-assisted introgression of multiple unlinked QTL. *Mammal. Genome* 11: 145–150.

Lande, R. & Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743–56.

Lee, S.H. & van der Werf, J.H.J. 2005. The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree. *Genetics* 169: 455–466.

Lee, S.H. & van der Werf, J.H.J. 2006. An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genet. Sel. Evol*. 38: 25–43.

Marshall, K., Henshall, J. & van der Werf, J.H.J. 2002. Response from marker-assisted selection when various proportions of animals are marker typed: a multiple trait simulation study relevant to the sheep meat industry. *Anim. Sci.* 74: 223–232.

Marshall, K.J., van der Werf J.H.J. & Henshall, J. 2004. Exploring major gene - marker phase typing strategies in marker assisted selection schemes. *Anim. Sci.* 78: 213–227.

Meuwissen, T.H.E. & Goddard, M.E. 1996. The use of marker haplotypes in animal breeding schemes. *Genet. Sel. Evol*. 28: 161–176.

Meuwissen, T.H.E. & Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155: 421–430.

Meuwissen, T.H.E. & Goddard, M.E. 2001. Prediction of identity by descent probabilities from marker haplotypes. *Genet. Sel. Evol*. 33: 605–634.

Meuwissen, T.H.E., Hayes, B. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

Nejati-Javaremi, A., Smith, C. & Gibson, J.P. 1997. Effect of total allelic relationship on accuracy and response to selection. *J. Anim. Sci*. 75: 1738–1745.

Nimbkar, C., Pardeshi, V. & Ghalsasi, P. 2005. Evaluation of the utility of the FecB gene to improve the productivity of Deccani sheep in Maharashtra, India. pp. 145–154. *In* H.P.S. Makkar & G.J. Viljoen, eds. *Applications of gene-based technologies for improving animal production and health in developing countries.* Netherlands, Springer.

Pong-Wong, R., George, A.W., Woolliams, J.A. & Haley, C.S. 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genet. Sel. Evol*. 33: 453–471.

Pyasatian, N., Fernando, R.L. & Dekkers, J.C.M. 2006. Genomic selection for composite line development using low density marker maps. In *Proc. 8th Wrld. Congr. Genet. Appl. Livest. Prodn.*, Paper 22–65. Belo Horizonte, Brazil.

Rothschild, M.F. & Plastow, G.S. 1999. Advances in pig genomics and industry applications. *AgBiotechNet.* 1: 1–7.

van der Werf, J.H.J. & Marshall, K. 2005. Combining gene-based methods and reproductive technologies to enhance genetic improvement of livestock in developing countries, pp. 131–144. *In* H.P.S Makkar & G.J. Viljoen, eds. *Applications of gene-based technologies for improving animal production and health in developing countries*. Netherlands, Springer.

Zhang, W. & Smith, C. 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor. Appl. Genet*. 83: 813–820.

CHAPTER 11

Marker-assisted selection in poultry

Dirk-Jan de Koning and Paul M. Hocking

SUMMARY

Among livestock species, chicken has the most extensive genomics toolbox available for detection of quantitative trait loci (QTL) and marker-assisted selection (MAS). The uptake of MAS is therefore not limited by technical resources but mostly by the priorities and financial constraints of the few remaining poultry breeding companies. With the cost of genotyping decreasing rapidly, an increase in the use of direct trait- single nucleotide polymorphism (SNP)-associations in MAS can be predicted.

CURRENT STATUS OF CHICKEN BREEDING PROGRAMMES

Poultry production has been the fastest growing livestock industry over the last decades especially in middle- and low-income countries (Taha, 2003). In 2001, poultry production accounted for 70 million tonnes of poultry meat and 47 million tonnes of eggs (Arthur and Albers, 2003). Among poultry, chicken account for 85 percent of meat production and 96 percent of egg production (Bilgili, 2001; Arthur and Albers, 2003; Taha, 2003). While chickens have been domesticated and selected for thousands of years, modern poultry breeding started during the 1950s. One of the most notable features is the diversification between chickens bred for meat production (broilers) and those bred for table egg production (layers). This is a result of the negative genetic correlation in chicken between growth and reproductive traits. Within breeds, there is a separation into male and female lines that are crossed to produce commercial hybrids. In broilers, male lines are selected for growth and carcass quality whereas in female lines less emphasis is placed on growth and more on reproductive traits such as egg production and hatchability. In table egg-laying chickens, male lines are selected for high egg production and high egg weight whereas in female lines selection may emphasize rate of lay with less attention to egg size. In both broiler and layer lines the primary selection goal is the improvement of feed efficiency and economic gain.

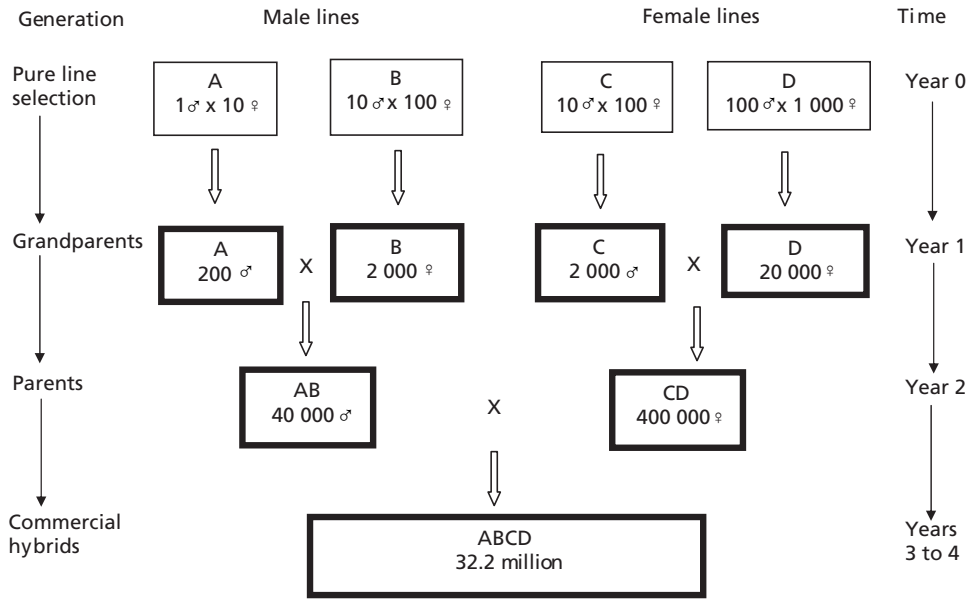
Significant heterosis for fitness traits in poultry is well established and all commercial poultry (chickens, turkeys and ducks) are hybrids that are produced in a selection and multiplication pyramid that is illustrated in Figure 1. Crossing male and female lines maximizes heterosis at the grandparent and parent levels of the hierarchy, and allows traits that have been genetically improved in different lines to be combined in the commercial birds. The power of this structure to deliver large economic gains in chickens is a result of their high reproductive rate and short generation interval and is clearly illustrated by this example of an egg-laying improvement programme. Even greater numerical efficiency is possible in broilers: a single pen containing ten females and one male at the nucleus level might produce 150 great-grandparents after selection (line D of Figure 1); these will produce 50 female offspring each or 7 500 grandparents in a year and these grandparents will generate 375 000 female parent stock during the succeeding year. These hybrid parent females will each produce over 130 male and female offspring and generate nearly 50 million commercial broilers or 70 000 tonnes of meat. The figure illustrates the rapidity with which genetic improvement at the nucleus level can be disseminated to commercial flocks and the fact that relatively few pure-line birds are needed to produce very large numbers of commercial layers.

The existence of this breeding structure results in rapid transmission of genetic change to commercial flocks (about four years), including traits that might be improved by MAS. Conversely, undesirable genetic change can also be disseminated very quickly to a very large number of birds. In practice, far more birds are kept at the nucleus level than shown in Figure 1 where the numbers presented are purely for illustrative purposes.

STATUS OF FUNCTIONAL GENOMICS IN CHICKEN

Among the various livestock species, chicken has the most comprehensive genomic tool-

FIGURE 1
Multiplication pyramid for a four-strain commercial hybrid layer



The boxes in bold are the multiplication (crossing) phase. The top line of the boxes indicates the lowest level of the pure-line (nucleus) selection stage. The numbers in the boxes refer to the minimal numbers of birds in line A, and corresponding numbers in lines B, C and D, that are required to generate hybrid laying hens at the commercial level and the numbers that result from this process. The time scale is indicated on the right hand side of the figure assuming a generation interval of one year.

Adapted from Bowman, 1974.

box. The chicken genome consists of 39 pairs of chromosomes: eight cytologically distinct macrochromosomes, the sex chromosomes Z and W and 30 pairs of cytologically indistinguishable microchromosomes. Linkage maps were developed initially using three separate mapping populations (Bumstead and Palyga, 1992; Crittenden *et al.*, 1993; Groenen *et al.*, 1998) that were later merged to provide a consensus map with 1 889 markers (Groenen *et al.*, 2000). A good overview of the consensus linkage map and the cytogenetic map can be found in the *First Report of Chicken Genes and Chromosomes 2000* and its successor in 2005 (Schmid *et al.*, 2000, 2005). All chicken maps can be viewed at www.thearkdb.org.

More recently, the chicken genome became the first livestock genome to be sequenced with a six-fold coverage (six full genome equivalents) (Hillier *et al.*, 2004). The chicken genome sequence can be browsed via a number of Web sites, which are summarized at www.chicken-genome.org/resources/databases.html. The genome sequence effort was accompanied by partial sequencing of three distinct poultry breeds (a broiler, a layer and a Chinese Silky), to identify SNPs between and among these and the reference sequence of the Red Jungle Fowl. This resulted in an SNP map consisting of about 2.8 million SNPs (Wong *et al.*, 2004). The chicken polymorphism database (ChickVD) can be browsed at: http://chicken.genomics.org.cn/index.jsp

(Wang *et al.*, 2005). The SNP map will facilitate the development of genome-wide SNP assays, containing between 5 000 and 20 000 SNPs per assay.

For the study of gene expression, there are various complementary DNA (cDNA) microarrays available, varying from targeted arrays (immune, neuroendocrine, embryo) to whole genome generic arrays. Recently, a whole-genome Affymetrix chip was developed in collaboration with the chicken genomics community (www.affymetrix.com and www.chicken-genome.org/resources/affymetrix-faq1.htm). Altogether, this provides a very comprehensive toolbox to study the functional genomics of chicken, whether this be an individual gene or the entire genome.

CURRENT UPTAKE OF MAS IN CHICKEN

Implementation of MAS requires knowledge of marker-trait associations based on QTL and candidate gene studies, and ideally from studies of the underlying genetic mechanisms. There have been a large number of QTL studies in chicken covering a wide range of traits including growth, meat quality, egg production, disease resistance (both infectious diseases and production diseases) and behaviour. These studies have recently been reviewed (Hocking, 2005). A total of 27 papers reported 114 genome-wide significant QTL from experimental crosses largely involving White Leghorn and broiler lines. A summary of the QTL that have been detected is presented in Table 1. While the abundance of QTL would indicate ample opportunity for MAS in chicken, it must be noted that nearly all studies were carried out in experimental crosses and hence the results do not reflect QTL within selected populations. However, these results do provide a good starting point to search for QTL within commercial populations, as demonstrated for growth and carcass traits where many published QTL also explained variation within a broiler dam line (de Koning *et al.*, 2003; de Koning *et al.*, 2004). To the authors' knowledge, there are no other QTL studies within commercial lines of poultry in the public domain. Of the QTL from experimental crosses, only a small number has been followed up by fine mapping analyses and the responsible gene mutation has only been described for some disease resistance QTL (Liu *et al.*, 2001a, b; Liu *et al.*, 2003).

A good example of how QTL mapping combined with functional studies can identify functional variants is for Marek's disease. Marek's disease (MD) is an infec-

TABLE 1
Quantitative traits and chromosomal locations in experimental chicken crosses

Trait	Chromosome (number of QTL)	Total QTL	Number of papers
Behaviour/fear	1,2(3),3,4(2),7,10,27, E22	11	5
Body fat	1(2),3,5,7(2),15,28	8	3
Body weight	1(7),2(4),3(4),4(5),5,8(2),11,12,13(2),27(3)Z(2)	32	9
Carcass quality	1(2),2,3,4(2).5(2),6(2),7(3),8(2),9,13(2),27,Z(2)	21	1
Disease resistance	1(4),2(2),3(2),4(2),5(5),6(2),7,8,14,18,27,Z	23	10
Egg number	8,Z(2)	3	1
Egg quality	2,11,Z	3	2
Egg weight	1,2,3,4(3),14,23,Z	9	3
Feed intake	1,4	2	2
Sexual maturity	Z(2)	2	2

Source: Hocking, 2005.

tious viral disease caused by a member of the herpes virus family and costs the poultry industry about US$1 000 million per annum. An F_2 cross between resistant and susceptible lines was challenged experimentally and genotyped, providing the data for a QTL analysis that resulted in a total of seven QTL for susceptibility to MD (Vallejo *et al.*, 1998; Yonash *et al.*, 1999). Subsequently, the founder lines of the F_2 cross were used for a micro-array study to identify genes that were differentially expressed between the two lines following artificial infection. Fifteen of these genes were mapped onto the chicken genome and two of them mapped to a QTL region for resistance to MD (Liu *et al.*, 2001a). At the same time, protein interaction studies between a viral protein (SORF2) and a chicken splenic cDNA library revealed an interaction with the chicken growth hormone (GH) (Liu *et al.*, 2001b). This led to the detection of a polymorphism in the GH gene that was associated with differences in the number of tumours between the susceptible and the resistant line (Liu *et al.*, 2001b). GH coincided with a QTL for resistance and was differentially expressed between founder lines (Liu *et al.*, 2001a).

Alongside the various genome scans for QTL, a large number of candidate gene studies have been carried out. The majority of studies summarized in Table 2 have been conducted on White Leghorn strains and have utilized restriction fragment length polymorphisms (RFLPs), SNPs or single strand conformation polymorphisms (SSCPs). These techniques require both that the gene is known and that the experimenter is able to sequence part of the gene to detect polymorphisms that distinguish the experimental lines.

Candidate gene studies have been used in two ways. First, candidate genes may be used merely as a marker for a trait (typically disease) based on prior knowledge and, second, and much less often, to search for the mutation within a gene that is associated with phenotypic variation in a trait. Currently, potential (candidate) genes for a QTL may be obtained from a knowledge of physiology (Dunn *et al.*, 2004) or comparative linkage maps (i.e. locating genes that are in the location of the QTL based on common areas of the gene-rich genomes of different species, usually human and mouse). There are likely to be many more of the second type of candidate gene studies as information from large-scale gene expression and proteomic experiments begin to suggest novel gene candidates for traits of commercial and biological importance. It should also be noted that there is good evidence that genetic variation is not limited to genomic DNA: associations between polymorphisms in mitochondrial genes and MD resistance, body weight and egg shell quality were reported by Li *et al.* (1998a, b).

Despite great enthusiasm for breeding companies to be involved in functional genomics research in poultry, there are very few applications of MAS in commercial poultry breeding. One existing example is the use of blood group markers to improve resistance to MD where selection of haplotypes B^{21} and B^{12} based on conventional serological tests has been widely used (McKay, 1998). In discussions with the industry it is clear that most interest is in QTL or candidate genes for resistance to diseases like MD or ascites, a genetic condition associated with pulmonary hypertension, leading to mortality in fast growing birds. There is also considerable interest among breeders of layer lines for egg quality, especially egg shell quality because of its importance for food

TABLE 2
Association of candidate genes with quantitative traits in poultry

Trait	Chromosomes[1]	Gene symbols	References
Age at first egg	1,2,3	GH, NPY, ODC	Feng *et al.,* 1997; Dunn *et al.,* 2004; Parsanejad *et al.,* 2004
Disease resistance (*E. coli*)	16	MHC1, MHC4, TAP2	Yonash *et al.,* 1999
Disease resistance (MD[2])	1,NK	GH, LY6E	Kuhnlein *et al.,* 1997; Liu *et al.,* 2001a,b and 2003
Disease resistance (Sal[3])	4,6,7,16,19, 1,17,NK	TNC, PSAP, NRAMP1[4], MHC1,CASP1, IAP1, TLR4, TLR5	Hu *et al.,* 1997; Lamont *et al.,* 2002; Leveque *et al.* 2003; Liu and Lamont, 2003; Iqbal *et al.,* 2005
Double yolked eggs	10	GNRHR	Dunn *et al.,* 2004
Egg production	Z,1,20	GHR, GH, PEPCK	Feng *et al.,* 1997; Kuhnlein *et al.,* 1997; Parsanejad *et al.,* 2003
Egg weight	1	IGF1	Nagaraja *et al.,* 2000
Eggshell quality	1,3,20	IGF1, ODC, PEPCK	Nagaraja *et al.,* 2000; Parsanejad *et al.,* 2003, 2004
Body fat	1,1,5,Z	GH, IGF1, TGFβ3, GHR	Feng et al., 1998; Fotouhi *et al.,* 1993; Li *et al.,* 2003; Zhou *et al.,* 2005
Feed efficiency	3,20	ODC, PEPCK	Parsanejad et al., 2003 and 2004
Body weight/carcass quality	1,3,5,Z,1,1,1	IGF1, ODC, TGFβ3, GHR, APOA2, PIT1	Feng *et al.,* 1998; Li *et al.,* 2003; Jiang *et al.,* 2004; Parsanejad *et al.,* 2004; Li *et al.,* 2005; Zhou *et al.,* 2005
Organ weight (spleen)	3,5,32	TGFβ2, TGFβ3, TGFβ4[5]	Li *et al.,* 2003
Skeletal traits	1,3,5,32	IGF1, TGFβ2, TGFβ3, TGFβ4[5]	Li *et al.,* 2003; Zhou *et al.,* 2005

[1] NK = gene has not yet been assigned to a chromosome.
[2] Marek's Disease.
[3] Salmonellosis.
[4] Now known as Slc11a1.
[5] TGFβ4 in the paper is now known to be TGFβ1.

safety. For production traits such as growth and egg numbers, breeders make sufficient progress using traditional selection methods, and they expect little improvement from MAS for such traits unless markers can be used to increase the accuracy of selection. Nonetheless, among breeders of broiler stock there is interest in markers for traits that are difficult to measure such as feed efficiency and meat quality in addition to disease resistance.

POTENTIAL FOR MAS IN CHICKEN

The technical aspects and potential implications of implementing MAS in livestock are discussed in Chapter 10 and Dekkers (2004), and van der Beek and van Arendonk (1996) evaluated the technical aspects of MAS in poultry breeding. A review of the potential of MAS in poultry is provided by Muir (2003) but this includes many of the technical issues that are common across livestock species. This chapter therefore focuses on poultry-specific issues, and readers are referred to Chapter 10 or Muir (2003) for a more comprehensive overview of applications and limitations of MAS.

Muir (2003) identified two cases where MAS could increase the selection intensity in poultry breeding: (i) traits that are measured later in life or are costly to measure (such as egg production and feed efficiency for broiler breeders); and (ii) selection within full-sib families for sex-limited traits (e.g. male chicks for egg production). Accuracy of selection can also be improved via MAS when selecting between full-sib families for sex-limited traits and traits that cannot be measured directly on one or both

sexes and/or have a low heritability (e.g. egg production, disease resistance, carcass quality and welfare traits).

Limiting factors for application of MAS (Muir, 2003) include biological factors (reproductive capacity) and many theoretical considerations related to the effectiveness of MAS (e.g. diverting selection pressure from polygenes to a single marked gene), which are generally applicable to MAS in livestock (Dekkers, 2004; Chapter 10). One of the concerns of Muir (2003) is the expected lack of major QTL for traits that have been under selection for many generations (following simulation results). However, recent QTL studies within commercial lines of pigs (Evans *et al.,* 2003; Nagamine *et al.,* 2003) and poultry (de Koning *et al.,* 2003, 2004) have demonstrated that many sizeable QTL are still segregating in commercial populations despite decades of selection.

There is strong academic interest in chicken genomics outside agriculture from, among others, developmental biologists and evolutionary geneticists, and this has contributed greatly to the development of the current functional genomics toolbox available for chicken. Among livestock species, chickens are best placed to pioneer new approaches where QTL studies are complemented by gene expression studies (Liu *et al.,* 2001a) or where they become fully integrated within "genetical genomics" (de Koning, Carlborg and Haley, 2005; de Koning and Haley, 2005).

If poultry breeders decide to embrace MAS, one of the main questions is whether they are prepared to re-structure their breeding programmes around MAS or implement these around their current breeding strategies. Adopting the terminology of Dekkers (2004), there are three levels of MAS: gene-assisted selection (GAS) where the functional mutation and its effects are known; linkage disequilibrium MAS (LD-MAS) where a marker (or marker haplotypes) is in population-wide disequilibrium with a QTL; and linkage equilibrium MAS (LE-MAS) where markers are in Hardy-Weinberg equilibrium with the QTL at the population level, but linkage disequilibrium exists within families. A fourth type of MAS that was recently proposed is "genome-wide MAS" (GW-MAS), where dense markers (i.e. SNPs) across the genome are used to predict the genetic merit of an individual without targeting any individual QTL or measuring (expensive) phenotypes on every generation (Meuwissen, Hayes, and Goddard, 2001). Integrating current evaluations with MAS is most straightforward for GAS and LD-MAS because the QTL effect can be included in routine evaluations as a fixed effect (Chapter 10). LE-MAS, on the other hand, requires extensive genotyping and fairly complicated statistical procedures (Wang, Fernando and Grossman, 1998), while GW-MAS reduces the genome to a "black-box" but does not require selection of QTL using arbitrary thresholds. Furthermore, the dense marker information required for GW-MAS may dispense with often faulty pedigree records because all pedigree information is encoded in the genome-wide genotypes.

In terms of quantitative genetic theory, there are ongoing developments in the tools required to detect and evaluate QTL in arbitrary pedigrees, moving away from strictly additive-dominance models to epistasis and parent-of-origin effects (Liu, Jansen and Lin, 2002; Shete and Amos, 2002). At the same time, the technology to analyse more than 10 000 SNPs in a single assay is available, and a cost of as little as US$0.02 per genotype is likely for chicken

SNPs in the near future. However, the fine mapping and characterization of identified QTL remain costly and time-consuming processes and are often restricted to the most promising QTL, resulting in hundreds of QTL that will never make it past the stage of mapping to a 30 cM confidence interval.

While current research and developments in poultry functional genomics are relevant to all four possible applications of MAS to livestock, poultry breeders need to decide at what level they want to exploit molecular information and for which traits.

The emerging picture is that breeders are more comfortable with known gene mutations as this provides an easy route to implementation as well as knowledge about the underlying biology. Furthermore, there is concern that the marker-trait linkage will break down over a relatively few generations of selection in large commercial flocks. While candidate gene studies would provide the quickest route to implementation, fine mapping and characterization of QTL (e.g. using expression studies) may reveal gene variants that are not obvious candidate genes for quantitative traits.

POTENTIAL FOR MAS IN POULTRY IN DEVELOPING COUNTRIES

Owing to the relatively low value of single animals, the high reproductive rate in poultry and good portability of eggs or day-old hatchlings, the concentration of resources is very high in the poultry breeding industry and all poultry breeding is privately owned. Fifty years ago there were many primary breeders in each and every industrialized country, but not so long ago there were only 20 breeding companies worldwide. Today, three groups of primary breeders dominate the international layer market. Equally, in the chicken meat industry, there are four major players in broiler breeding worldwide (Flock and Preisinger, 2002). The concentration process is probably now complete, and the present players are sufficient to meet the global supply for 700 000 million eggs as final products. A similar trend is expected in the pig industry, where international breeding companies of hybrid products are increasing their market share (Preisinger, 2004). For large-scale farming of broilers and layers in developing countries there are additional challenges with regard to heat stress and potential disease pressure. With increasing poultry production in developing countries, breeding companies may give priority to using breeding and molecular tools to address these additional challenges. While chickens are very efficient in converting grain into valuable meat and egg protein, and smallholder chicken production can be valuable for sustaining the livelihoods of farmers in the developing world, this type of poultry production would require robust dual-purpose (meat and egg) birds, rather than specialized broiler and layer lines. It is unlikely that the commercial breeders will develop such lines but there may be scope for national or international research organizations to do so. Any MAS would have to be done at the institutional level where the line is developed and would necessitate prior knowledge of trait-marker associations at the farm level. The implementation of whole genome SNP approaches to farm level recording might facilitate progress in this area but the challenges, both practical and theoretical, are formidable.

CONCLUDING REMARKS

Among livestock species, chicken have by far the most comprehensive genomic toolbox. However, uptake of MAS will

depend strongly on whether the industry wishes to supplement its current selection programme with a known gene variant or whether it is prepared to restructure breeding programmes around MAS. Compared with, for instance, the dairy cattle industry, the poultry breeding community may be slower to embrace emerging complex approaches to MAS. This is somewhat surprising because the closed structure of the poultry breeding pyramid offers much better protection of intellectual property than the dairy cattle industry where semen from highest ranking bulls is available for all. On the other hand, the fact that blood groups have been used to select for resistance to MD suggests that poultry breeders have some experience and skills in this type of selection. Poultry breeding companies contribute significantly to poultry genomics research but may not be fully convinced about the economic feasibility of MAS. To implement MAS successfully, a company must tackle the problems of identifying the traits to select and their economic significance, the lack of current knowledge of the genes or markers associated with these traits, and their association with other economic selection criteria. The current "toolbox" provides the means to answer some of these questions but there are obvious concerns about human and capital resources and the potential loss of gains in other traits in a competitive market. Coupled with these reservations must be the very evident success of current breeding programmes in achieving many desirable commercial goals.

ACKNOWLEDGEMENTS

The Roslin Institute is supported by a core strategic grant from the Biotechnology and Biological Sciences Research Council (BBSRC). The authors gratefully acknowledge information on breeding objectives from representatives of the poultry breeding industry.

REFERENCES

Arthur, J.A. & Albers, G.A.A. 2003. Industrial perspective on problems and issues associated with poultry breeding. *In* W. M. Muir & S. E. Aggrey, eds. *Poultry genetics, breeding and biotechnology,* pp. 1–12. Wallingford, UK, CABI Publishing.

Bilgili, S.F. 2001. Poultry products and processing in the international market place. *Proc. Internat. Anim. Agric. & Food Sci. Conf*. Indianapolis, IN, USA (available at www.fass.org/fass01/pdfs/Bilgili.pdf).

Bowman, J.C. 1974. *An introduction to animal breeding*. The Institute of Biology's Studies in Biology No 46. London, Edward Arnold.

Bumstead, N. & Palyga, J. 1992. A preliminary linkage map of the chicken genome. *Genomics* 13: 690–697.

Crittenden, L.B., Provencher, L., Sanatangello, L., Levin, I., Abplanalp, H., Briles, R.W., Briles, W.E. & Dodgson, J.B. 1993. Characterization of a Red Jungle Fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. *Poultry Sci.* 72: 334–348.

de Koning, D.J., Carlborg, O. & Haley. C.S. 2005. The genetic dissection of immune response using gene-expression studies and genome mapping. *Vet. Immun. & Immunopath.* 105: 343–352.

de Koning, D.J. & Haley, C.S. 2005. Genetical genomics in humans and model organisms. *Trends Genetics* 21: 377–381.

de Koning, D.J., Haley, C.S., Windsor, D., Hocking, P.M., Griffin, H., Morris, A., Vincent, J. & Burt, D.W. 2004. Segregation of QTL for production traits in commercial meat-type chickens. *Genetic. Res.* 83: 211–220.

de Koning, D.J., Windsor, D., Hocking, P.M., Burt, D.W., Law, A., Haley, C.S., Morris, A., Vincent, J. & Griffin, H. 2003. Quantitative trait locus detection in commercial broiler lines using candidate regions. *J. Anim. Sci.* 81:1158–1165.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sc.* 82 E-Suppl: E313–E328.

Dunn, I.C., Miao, Y.W., Morris, A., Romanov, M.N., Wilson, P.W. & Waddington, D. 2004. A study of association between genetic markers in candidate genes and reproductive traits in one generation of a commercial broiler breeder hen population. *Heredity* 92: 128–134.

Evans, G.J., Giuffra, E., Sanchez, A., Kerje, S., Davalos, G., Vidal, O., Illan, S., Noguera, J.L., Varona, L., Velander, I., Southwood, O.I., de Koning, D.J., Haley, C.S., Plastow, G.S. & Andersson, L. 2003. Identification of quantitative trait loci for production traits in commercial pig populations. *Genetics* 164: 621–627.

Feng, X.P., Kuhnlein, U., Aggrey, S.E., Gavora, J.S. & Zadworny, D. 1997. Trait association of genetic markers in the growth hormone and the growth hormone receptor gene in a White Leghorn strain. *Poultry Science* 76: 1770–1775.

Feng, X.P., Kuhnlein, U., Fairfull, R.W., Aggrey, S.E., Yao, J. & Zadworny, D. 1998. A genetic marker in the growth hormone receptor gene associated with body weight in chickens. *J. Heredity* 89: 355–359.

Flock, D.K. & Preisinger, R. 2002. Breeding plans for poultry with emphasis on sustainability. *7th World Congress on Genetics Applied to Livestock Production*, 19–23 August, 2002, Montpellier, France, Session 24: Sustainable breeding plans in developed countries. Communication 24-02.

Fotouhi, N., Karatzas, C.N., Kuhnlein, U. & Zadworny, D. 1993. Identification of growth-hormone DNA polymorphisms which respond to divergent selection for abdominal fat-content in chickens. *Theor. Appl. Genet.* 85: 931–936.

Groenen, M.A., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L.B., Dodgson, J., Hillel, J., Lamont, S., de Leon, A.P., Soller, M., Takahashi, H. & Vignal, A. 2000. A consensus linkage map of the chicken genome. *Genome Research* 10: 137–147.

Groenen, M.A., Crooijmans, R.P., Veenendaal, A., Cheng, H.H., Siwek, M. & van der Poel, J.J. 1998. A comprehensive microsatellite linkage map of the chicken genome. *Genomics* 49: 265–274.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E., Dodgson, .J.B., Chinwalla, A.T., Cliften, P.F., Clifton, S.W., Delehaunty, K.D., Fronick, C., Fulton, R.S., Graves, T.A., Kremitzki, C., Layman, D., Magrini, V., McPherson, J.D., Miner, T.L., Minx, P., Nash, W.E., Nhan, M.N., Nelson, J.O., Oddy, L.G., Pohl, C.S., Randall-Maher, J., Smith, S.M., Wallis, J.W., Yang, S.P., Romanov, M.N., Rondelli, C.M., Paton, B., Smith, J., Morrice, D., Daniels, L., Tempest, H.G., Robertson, L., Masabanda, J.S., Griffin, D.K., Vignal, A., Fillon, V., Jacobbson, L., Kerje, S., Andersson, L., Crooijmans, R.P., Aerts, J., van der Poel, J.J., Ellegren, H.. Caldwell, R.B., Hubbard, S.J., Grafham, D.V., Kierzek, A.M., McLaren, S.R., Overton, I.M., Arakawa, H., Beattie, K.J., Bezzubov, Y., Boardman, P.E., Bonfield, J.K., Croning, M.D., Davies, R.M., Francis, M.D., Humphray, S.J., Scott, C.E., Taylor, R.G., Tickle, C., Brown, W.R., Rogers, J., Buerstedde, J.M., Wilson, S.A., Stubbs, L., Ovcharenko, I., Gordon, L., Lucas, S., Miller, M.M., Inoko, H., Shiina, T., Kaufman, J., Salomonsen, J., Skjoedt, K., Wong, G.K., Wang, J., Liu, B., Wang, J., Yu, J., Yang, H., Nefedov, M., Koriabine, M., Dejong, P.J., Goodstadt, L., Webber, C., Dickens, N.J., Letunic, I., Suyama, M., Torrents, D., Von, M.C., Zdobnov, E.M., Makova, K., Nekrutenko, A.,

Elnitski, L., Eswara, P., King, D.C., Yang, S., Tyekucheva, S., Radakrishnan, A., Harris, R.S., Chiaromonte, F., Taylor, J., He, J., Rijnkels, M., Griffiths-Jones, S., Ureta-Vidal A., Hoffman, M.M., Severin, J., Searle, S.M., Law, A.S., Speed, D., Waddington, D., Cheng, Z., Tuzun, E., Eichler, E., Bao, Z., Flicek, P., Shteynberg, D.D., Brent, M.R., Bye, J.M., Huckle, E.J., Chatterji, S., Dewey, C., Pachter, L., Kouranov, A., Mourelatos, Z., Hatzigeorgiou, A.G., Paterson, A.H., Ivarie, R., Brandstrom, M., Axelsson, E., Backstrom, N., Berlin, S.,Webster, M.T., Pourquie, O., Reymond, A., Ucla, C., Antonarakis, S.E., Long, M.J.J., Emerson, J.J Betran, E., Dupanloup, I., Kaessmann, H., Hinrichs, A.S., Bejerano, G., Furey, T.S., Harte, R.A., Raney, B., Siepel, A., Kent, W.J., Haussler, D., Eyras, E., Castelo, R., Abril, J.F., Castellano, S., Camara, F., Parra, G., Guigo, R., Bourque, G., Tesler, G., Pevzner, P.A., Smit, A., Fulton, L.A., Mardis, E.R. & Wilson, R.K. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.

Hocking, P.M. 2005. Review of QTL results in chicken. *Wld. Poultry Sci. J.* 61: 215–226.

Hu, J.X., Bumstead, N., Barrow, P., Sebastein, G., Olien, L., Morgan, K. & Malo, D. 1997. Resistance to salmonellosis in the chicken is linked to NRAMP1 and TNC. *Genome Res.* 7: 693–704.

Iqbal, M., Philbin, V.J., Withanage, G.S.K., Wigley, P., Beal, R.K., Goodchild, M.J., Barrow, P., McConnell, I., Maskell, D.J., Young, J., Bumstead, N., Boyd, Y. & Smith A.L. 2005. Identification and functional characterization of chicken Toll-like receptor 5 reveals a fundamental role in the biology of infection with *Salmonella enterica* serovar *typhimurium*. *Infect. & Immunity* 73: 2344–2350.

Jiang, R., Li, J.,Qu, Li, H. & Yang, N. 2004. A new single nucleotide polymorphism in the chicken pituitary- specific transcription factor (POULF1) gene associated with growth rate. *Anim. Genet.* 35: 344–346.

Kuhnlein, U., Ni, L., Weigend, S., Gavora, J.S., Fairfull, W. & Zadworny, D. 1997. DNA polymorphisms in the chicken growth hormone gene: response to selection for disease resistance and association with egg production. *Anim. Genet.* 28: 116–123.

Lamont, S.J., Kaiser, M.G. & Liu, W. 2002. Candidate genes for resistance to *Salmonella enteritidis* colonization in chickens as detected in a novel genetic cross. *Vet. Immun. & Immunopath.* 87: 423–428.

Leveque, G., Forgetta, V., Morroll, S., Smith, A.L., Bumstead, N., Barrow, P., Loredo-Osti, J.C. , Morgan, K. & Malo, D. 2003. Allelic variation in TLR4 is linked to susceptibility to *Salmonella enterica* serovar *typhimurium* infection in chickens. *Infect. & Immunity* 71: 1116–1124.

Li, S., Aggrey, S.E., Zadworny, D., Fairfull, W. & Kuhnlein, U. 1998a. Evidence for a genetic variation in the mitochondrial genome affecting traits in White Leghorn chickens. *J. Hered.* 89: 222–226.

Li, S., Zadworny, D., Aggrey, S.E. & Kuhnlein, U. 1998b. Mitochondrial pepck: A highly polymorphic gene with alleles co- selected with Marek's disease resistance in chickens. *Anim. Genet.* 29: 395–397.

Li, H., Deeb, N., Zhou, H., Mitchell, A.D., Ashwell, C.M. & Lamont, S.J. 2003. Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-beta genes. *Poult. Sci.* 82: 347–356.

Li, H., Deeb, N., Zhou, H., Ashwell, C.M. & Lamont, S.J. 2005. Chicken quantitative trait loci for growth and body composition associated with the very low density apolipoprotein-II gene. *Poult. Sci.* 84: 697–703.

Liu, W. & Lamont, S.J. 2003. Candidate gene approach: Potential association of caspase-1, inhibitor of apoptosis protein-1, and Prosaposin gene polymorphisms with response to *Salmonella enteritidis* challenge or vaccination in young chicks. *Anim. Biotech.* 14: 61–76.

Liu, H.C., Cheng, H.H., Tirunagaru, V., Sofer, L. & Burnside, J. 2001a. A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim. Genet.* 32: 351–359.

Liu, H.C., Kung, H.J., Fulton, J.E., Morgan, R.W. & Cheng. H.H. 2001b. Growth hormone interacts with the Marek's disease virus SORF2 protein and is associated with disease resistance in chicken. *Proc. Natl. Acad. Sci. USA* 98: 9203–9208.

Liu, H.C., Niikura, M., Fulton, J.E. & Cheng. H.H. 2003. Identification of chicken lymphocyte antigen 6 complex, locus E (LY6E, alias SCA2) as a putative Marek's disease resistance gene via a virus-host protein interaction screen. *Cytogenet. & Genome Res.* 102: 304–308.

Liu, Y., Jansen, G.B. &. Lin, C.Y. 2002. The covariance between relatives conditional on genetic markers. *Genet. Select. Evoln.* 34: 657–678.

McKay, J.C. 1998. A poultry breeder's approach to avian neoplasia. *Avian Pathology* 27: S74–S77.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

Muir, W.M. 2003. Incorporating molecular information in breeding programmes: applications and limitations. *In* W.M. Muir & S.E. Aggrey, eds. *Poultry Genetics, Breeding and Biotechnology*, pp. 549–562. Wallingford, UK, CABI Publishing.

Nagamine, Y., Haley, C.S., Sewalem, A. & Visscher, P.M. 2003. Quantitative trait loci variation for growth and obesity between and within lines of pigs (*Sus scrofa*). *Genetics* 164: 629–635.

Nagaraja, S.C., Aggrey, S.E., Yao, J., Zadworny, D., Fairfull R.W. & Kuhnlein, U. 2000. Trait association of a genetic marker near the IGF-I gene in egg-laying chickens. J. *Heredity* 91: 150–156.

Parsanejad, R., Torkamanzehi, A., Zadworny, D. & Kuhnlein, U. 2003. Alleles of cytosolic phosphoenolpyruvate carboxykinase (PEPCK): trait association and interaction with mitochondrial PEPCK in a strain of white leghorn chickens. *Poultry Sci.* 82: 1708–1715.

Parsanejad, R., Praslickova, D., Zadworny, D. & Kuhnlein, U. 2004. Ornithine decarboxylase: haplotype structure and trait associations in White Leghorn chickens. *Poultry Sci.* 83: 1518–1523.

Preisinger, R. 2004. Internationale Tendenzen der Tierzüchtung und die Rolle der Zuchtunternehmen. *Züchtungskunde* 76(6): 395–402.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J.M., Wimmers, K., Burt, D.W., Smith, J., A'Hara, S., Law, A., Griffin, D.K., Bumstead, N., Kaufman, J., Thomson, P.A., Burke, T., Groenen, M.A., Crooijmans, R.P., Vignal, A., Fillon, V., Morisson, M., Pitel, F.,Tixier-Boichard, M.T., Ladjali-Mohammedi, K., Hillel, J., Maki-Tanila, A., Cheng, H.H., Delany, M.E., Burnside, J. & Mizuno, S. 2000. First report on chicken genes and chromosomes 2000. *Cytogenet. & Cell Genet.* 90: 169–218.

Schmid, M., Nanda, I., Hoehn, H., Schartl, M., Haaf, T., Buerstedde, J.M., Arakawa, H., Caldwell, R.B., Weigend, S., Burt, D.W., Smith, J., Griffin, D.K., Masabanda, J.S., Groenen, M.A., Crooijmans, R.P., Vignal, A., Fillon, V., Morisson, M., Pitel, F., Vignoles, M., Garrigues, A., Gellin, J., Rodionov, A.V., Galkina, S.A., Lukina, N.A., Ben-Ari, G., Blum, S., Hillel, J., Twito, T., Lavi, U., David, L., Feldman, M.W., Delany, M.E., Conley, C.A., Fowler, V.M., Hedges, S.B., Godbout, R., Katyal, S., Smith, C., Hudson, Q., Sinclair, A. & Mizuno, S. 2005. Second report on chicken genes and chromosomes 2005. *Cytogen.& Genome Res.* 109: 415–479.

Shete, S. & Amos, C.I. 2002. Testing for genetic linkage in families by a variance components approach in the presence of genomic imprinting. *Am. J. Human Genet.* 70: 751–757.

Taha, F.A. 2003. *The poultry sector in middle income countries and its feed requirements: the case of Egypt.* Outlook Report No. WRS03-02, pp. 3–14. Washington, DC, Economic Research Service, USDA.

Vallejo, R.L., Bacon, L.D., Liu, H.C., Witter, R.L., Groenen, M.A., Hillel, J. & Cheng, H.H. 1998. Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens. *Genetics* 148: 349–360.

van der Beek, S. &. van Arendonk, J.A.M. 1996. Marker-assisted selection in an outbred poultry nucleus. *Anim. Sc.* 62: 171–180.

Wang, T., Fernando, R.L. & Grossman, M. 1998. Genetic evaluation by best linear unbiased prediction using marker and trait information in a multibreed population. *Genetics* 148: 507–515.

Wang, J., He, X., Ruan, J., Dai, M., Chen, J., Zhang, Y., Hu, Y., Ye, C., Li, S., Cong, L., Fang, L., Liu, B., Li, S., Wang, J., Burt, D.W., Wong, G.K., Yu, J., Yang, H. & Wang, J. 2005. ChickVD: a sequence variation database for the chicken genome. *Nucl. Acid Res.* 33: D438–D441.

Wong, G.K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruane, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L.W., Yang, S.P., Warren, W.C., Wilson, R.K., Brandstrom, M., Ellegren, H., Crooijmans, R.P., van der Poel, J.J., Bovenhuis, H., Groenen, M.A., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J.R., Rogers, S., Walker, B.A., van, H.A., Kaufman, J., Bumstead, N., Lamont, S.J., Zhou, H., Hocking, P.M., Morrice, D., de Koning, D.J., Law, A., Bartley, N., Burt, D.W., Hunt, H., Cheng, H.H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M.T., Andersson, B., Webber, C., Ponting, C.P., Overton, I.M., Boardman, P.E., Tang, H., Hubbard, S.J., Wilson, S.A., Yu, J., Wang, J. & Yang, H. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432: 717–722.

Yonash, N., Bacon, L.D. Witter, R.L. & Cheng. H.H. 1999. High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. *Anim. Genet.* 30: 126–135.

Zhou, H., Mitchell, A.D., McMurtry, J.P., Ashwell, C.M. & Lamont, S.J. 2005. Insulin-like growth factor-1 gene polymorphism associations with growth, body composition, skeleton integrity, and metabolic traits in chickens. *Poult. Sci.* 84: 212–219.

CHAPTER 12

Marker-assisted selection in dairy cattle

Joel Ira Weller

SUMMARY

Considering the long generation interval, the high value of each individual, the very limited female fertility and the fact that nearly all economic traits are expressed only in females, it would seem that cattle should be a nearly ideal species for application of marker-assisted selection (MAS). As genetic gains are cumulative and eternal, application of new technologies that increase rates of genetic gain can be profitable even if the nominal annual costs are several times the value of the nominal additional annual genetic gain. Complete genome scans for quantitative trait loci (QTL) based on the granddaughter design have been completed for most commercial dairy cattle populations, and significant across-study effects for economic traits have been found on chromosomes 1, 3, 6, 9, 10, 14 and 20. Quantitative trait loci associated with trypanotolerance have been detected in a cross between the African N'Dama and the Boran breeds as the first step in the introgression of these genes into breeds susceptible to trypanosomosis. In dairy cattle, the actual DNA polymorphism has been determined twice, for QTL on BTA 6 and BTA 14. In both cases the polymorphism caused a non-conservative amino acid change, and both QTL chiefly affect fat and protein concentration. Most theoretical studies have estimated the expected gains that can be obtained by MAS to be in the range of a 5 to 20 percent increase in the rates of genetic gain obtained by traditional selection programmes. Applied MAS programmes have commenced for French and German Holsteins. In both programmes genetic evaluations including QTL effects are computed by variants of marker-assisted best linear unbiased prediction (MA-BLUP).

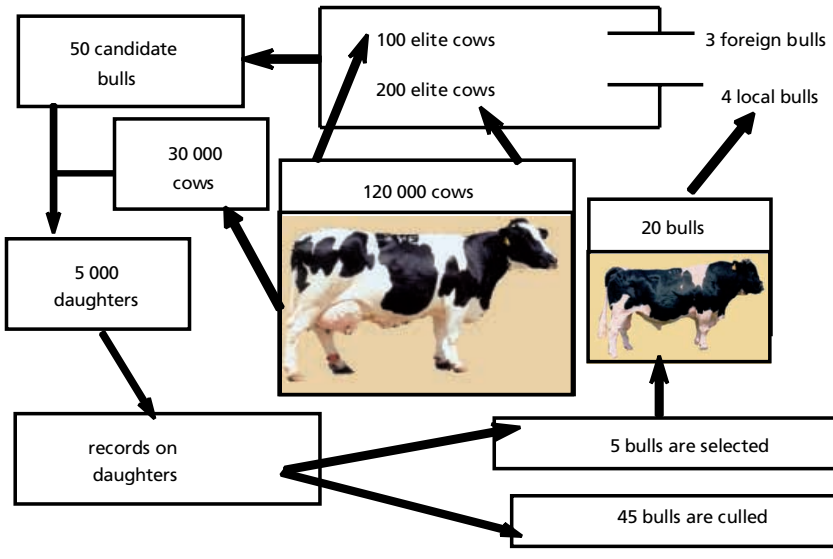
INTRODUCTION

Compared with other agricultural species, dairy cattle are unique in terms of the value of each animal, their long generation interval and the very limited fertility of females. Thus unlike plant and poultry breeding, most dairy cattle breeding programmes are based on selection within the commercial population. Similarly, detection of quantitative trait loci (QTL) and marker-assisted selection (MAS) programmes are generally based on analysis of existing populations. The specific requirements of dairy cattle breeding have led to the generation of very large data banks in most developed countries, which are available for analysis. In this chapter, dairy cattle breeding programmes in the developed and developing countries are reviewed and compared. The important issues in the application of MAS are then outlined. These include economic considerations based on phenotypic selection, the current status of cattle marker maps, methods to detect QTL and to estimate QTL effects and location suitable for dairy cattle, the current state of QTL detection in dairy cattle, methods to incorporate information from genetic markers in genetic evaluation systems, methods to identify the actual polymorphisms responsible for observed QTL and description of the reported results, methods and theory for MAS in dairy cattle, the current status of MAS and, finally, the future prospects for MAS in dairy cattle.

DAIRY CATTLE BREEDING PROGRAMMES IN DEVELOPED COUNTRIES

In most developed countries, dairy cattle breeding programmes are based on the "progeny test" (PT) design. The PT is the design of choice for moderate to large dairy cattle populations, including the United States Holsteins, which include over ten million animals. An example of the Israeli PT design is given in Figure 1.

FIGURE 1
The Israeli Holstein breeding programme



This population consists of approximately 120 000 cows of which 90 percent are milk recorded. Approximately 20 bulls are used for general service. Each year about 300 elite cows are selected as bull dams. These are mated to the two to four best local bulls and an equal number of foreign bulls to produce approximately 50 bull calves for progeny testing. At the age of one year, the bull calves reach sexual maturity, and approximately 1 000 semen samples are collected from each young bull. These bulls are mated to approximately 30 000 first parity cows to produce about 5 000 daughters, or 100 daughters per young bull. Gestation length for cattle is nine months. Thus the young bulls are approximately two years old when their daughters are born, and are close to four when their daughters calve and begin their first lactation. At the completion of their daughters' first lactations, most of the young bulls are culled. Only four to five are returned to general service, and a similar number of the old proven sires are culled. By this time the selected bulls are approximately five years old.

Various studies have shown that rates of genetic gain by a PT scheme are about 0.1 to 0.2 genetic standard deviations of the selection index per year (Nicholas and Smith, 1983; Israel and Weller, 2000). The PT was devised to take advantage of the nearly unlimited fertility of males. However, compared with breeding schemes for other species, the PT has several major weaknesses. First, for a PT system to be effective, the population must include at least several tens of thousands of animals with recording on production traits and paternity. Inaccurate recording can significantly reduce rates of genetic gain (Israel and Weller, 2000). Second, generation intervals, especially along the sire-to-dam and sire-to-sire paths, are much longer than the biological requirements. The increase in generation interval reduces genetic gain per year. As artificial insemination (AI) institutes generally pay a premium price for male calves of elite cows, these cows are often given preferential treatment in order to increase their genetic evaluations (Powell and Norman, 1988). The small number of bulls actually used for general service, and the even smaller number of bulls used as bull sires, tends to reduce the effective population size, which increases inbreeding and decreases genetic variance in the population. The effective population size of the United States Holstein population with ten million cows has been estimated at about 100 (Farnir *et al.*, 2000). Finally, there is virtually no selection along the dam-to-dam path. Generally, 70-80 percent of healthy female calves produced are used as replacements.

Various studies have suggested that selection intensities along the dam-to-dam path could be increased by application of multiple ovulation and embryo transfer (MOET) and sexed semen. Costs of both technologies are still prohibitively high to be applied to the entire population, as shown below. To overcome this problem for MOET, Nicholas and Smith (1983) proposed a "nucleus" breeding scheme. In nucleus schemes, the selection population consists of several hundred individuals, and bulls are not progeny tested. Instead, bulls are selected based on the genetic evaluations of their dams and sisters, which shortens the generation interval on the sire-to-dam and sire-to-sire paths, but reduces the reliabilities of the genetic evaluations. Dams of bulls and cows are selected based chiefly on their own production records, and MOET is applied to increase the number of progeny per dam. As the selection population consists of only several hundred individuals, MOET costs

are manageable if costs are spread over the entire national dairy industry. Rates of genetic gain within the nucleus are thus higher than can be obtained by a national PT design. This gain is transferred to the general population through the use of bulls from the nucleus population. In addition to the greater overall rate of genetic gain, the nucleus scheme has the advantage that it is necessary to collect data on a much smaller population, which should reduce costs and increase accuracy. The disadvantages of MOET are that overall costs and rates of increase of inbreeding will be greater unless steps are taken to reduce inbreeding. However, these steps will also slightly decrease rates of genetic gain. In practice, no country has replaced its standard PT scheme with a nucleus breeding programme.

DAIRY CATTLE BREEDING IN DEVELOPING COUNTRIES

The genus *Bos* includes five to seven species, of which *Bos taurus* and *Bos indicus* are the most widespread and economically important. *B. taurus* is the main dairy cattle species, and is found generally in temperate climates. Several tropical and subtropical cattle breeds are the result of crosses between B. *taurus* and B. *indicus*, which interbreed freely. In the tropics, cows need at least some degree of tolerance to environmental stress due to poor nutrition, heat and disease challenge to sustain relatively high production levels (Cunningham, 1989). Tropical breeds are adapted to these stresses but have low milk yield, whereas more productive temperate breeds cannot withstand the harsh tropical conditions, to the point of not being able to sustain their numbers (de Vaccaro, 1990). Furthermore, most tropical countries are developing countries, which lack systematic large-scale milk and pedigree recording.

A number of studies have been conducted on crosses between imported and local breeds in the tropics. Generally, the F_1 *B. taurus* x *B. indicus* crosses are economically superior to either of the purebred strains (FAO, 1987). The heterosis effect of the F_1 cross is due to genes for disease resistance from the local parent, and genes for milk production from the imported strain (Smith, 1988; Cunningham, 1989). However, this heterosis is lost in future generations if the F_1 is backcrossed to either parental strain. Madalena (1993) presented an F_1 continuous replacement scheme to capitalize on its superiority. Recently, Kosgey, Kahi and van Arendonk (2005) proposed a closed adult nucleus MOET scheme to increase milk production in tropical crossbred cattle.

ECONOMIC CONSIDERATIONS IN APPLYING MAS TO DAIRY CATTLE

For any new technique to be economically viable, overall gains must be greater than overall costs. This also applies to using MAS within a dairy cattle breeding programme. However, unlike investment in new equipment, genetic gains never "wear out", i.e. breeding is unique in that genetic gains are cumulative and eternal. Thus, as shown by Weller (1994, 2001) investments in MAS or other techniques that enhance breeding programmes are economically viable even if "nominal" costs are greater than "nominal" gains.

For example, consider an ongoing breeding programme with a constant rate of genetic gain per year. Assume that the annual rate of genetic gain has a nominal economical value of V. The cumulative discounted returns to year T, R_v, will be a function of the nominal annual returns, the discount rate, d, the profit horizon, T, and the number of years from the beginning of the programme until

first returns are realized, t. R_v is computed as follows (Hill, 1971):

$$R_v = V\frac{r_d^t - r_d^{T+1}}{(1-r_d)^2} - \frac{(T-t+1)r_d^{T+1}}{1-r_d} \quad \{1\}$$

where r_d= 1/(1+d). For example, with d = 0.08, T = 20 years, and t = 5 years, R_v = 32.58V. That is, the cumulative returns are equal to nearly 33 times the nominal annual returns. For an infinite profit horizon, Equation {1} reduces to:

$$R_v = \frac{Vr^t}{(1-r_d)^2} = \frac{V}{d^2(1+d)^{t-2}} \quad \{2\}$$

and R_v = 124.04V.

The value of nominal annual genetic gain will now be compared with the annual costs of a breeding programme, assuming a fixed nominal cost per year. Costs, unlike genetic gain, only have an effect in the year they occur. Assuming that annual costs are equal during the length of the breeding programme, and that first costs occur in the year after the base year, C_T, the net present value of the total costs of the breeding programme is computed as follows:

$$C_T = \frac{C_c r_d (1-r_d^T)}{1-r_d} \quad \{3\}$$

where C_c = annual costs of the breeding programme. Using the same values for T and d, C_T = 9.82C_c. Thus, with a profit horizon of 20 years, cumulative profit is positive if V > 0.31C_c. For an infinite profit horizon, C_T = 12.5C_c, and profit will be positive if V > 0.1C_c.

Therefore, a breeding programme can be profitable even if the nominal annual costs are several times the value of the nominal annual genetic gain. For example, consider the United States of America dairy cattle population, which consists of about ten million cows. Annual genetic gain is about 100 kg milk per year. The value of a 1 kg gain in milk production has been estimated at US$0.1 (Weller, 1994). Thus, the nominal annual value of a 10 percent increase in the rate of genetic gain (10 kg per year) is:

V = (10 kg per cow per year)
(US$0.1 per kg)(10 000 000 cows) = {4}
US$10 000 000 per year

The cumulative value with a profit horizon of 20 years and an 8 percent discount rate would be US$326 million, and break-even annual costs for a technology that increases annual genetic gain by 10 percent are US$32 million per year. Thus, it would be profitable to spend quite a lot for a relatively small genetic gain.

The value of genetic gain to a specific breeding enterprise will generally be less than the gain to the general economy. This is because most of the gains obtained by breeding will be passed on to the consumers. Brascamp, van Arendonk and Groen (1993) considered the economic value of MAS based on changes in returns from semen sales for a breeding organization operating in a competitive market. In this case, a breeding firm that adopts a MAS programme can increase its returns either by increasing its market share or increasing the mean price of a semen dose. Although the value of genetic gain will be less, relatively small changes in genetic merit can result in large changes in market share.

CURRENT STATUS OF MARKER MAPS IN CATTLE

Cattle have 29 pairs of autosomes and one pair of sex chromosomes. All the autosomes are acrocentric, and map units are

scored from the centromere. Chromosomes are denoted with the prefix "BTA" (*B. taurus*). Similar to other mammals, the bovine DNA includes 3×10^9 base pairs (bp), and the map length is approximately 3 000 cM. The human genome is estimated to encode 20 000–25 000 protein-coding genes (International Human Genome Sequencing Consortium, 2004), and it can be assumed that the number of genes in other mammals, including cattle, should be quite similar. Thus, a single map unit, on average, includes approximately eight genes and one million bp.

As in other animal species, microsatellites are still the marker of choice for map construction due to their prevalence and high polymorphism. Although single nucleotide polymorphisms (SNPs) are much more prevalent, genetic maps based on SNPs are still in the future. More than 50 000 SNPs have been identified in humans, but only several thousand have been validated in cattle (www.afns.ualberta.ca/Hosted/Bovine%20Genomics/), and rates of polymorphism are generally unknown. With the completion of the six-fold coverage of the bovine genome by the Bovine Genome Sequencing Project at Baylor College of Medicine (www.hgsc.bcm.tmc.edu/projects/bovine/) many more SNPs will be identified.

Several genetic maps are available on the internet. The United States Meat Animal Research Center (MARC) (www.marc.usda.gov/) includes thousands of markers, chiefly microsatellites. The ArkDB database system, hosted at Roslin Institute, includes data from several published maps (www.thearkdb.org/). The Commonwealth Scientific and Industrial Research Organization (CSIRO) livestock industries cattle genome marker map is built upon data provided by the University of Sydney's comparative location database (www.livestockgenomics.csiro.au/perl/gbrowse.cgi/cattlemap/). This map combined all publicly-available maps into a single integrated map that currently includes 9 400 markers.

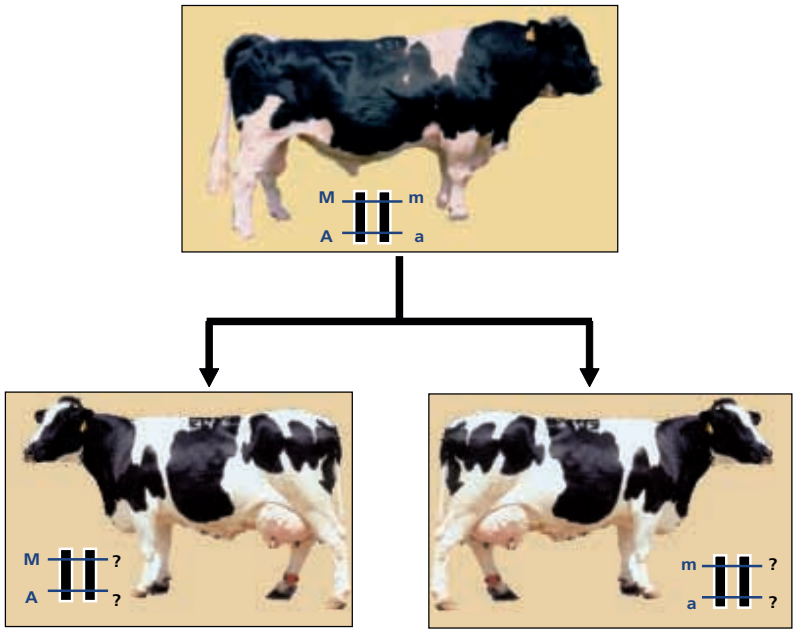
METHODS OF QTL DETECTION SUITABLE FOR COMMERCIAL DAIRY CATTLE POPULATIONS

Detection of QTL requires generation of linkage disequilibrium (LD) between the genetic markers and QTL. In plants, this is generally accomplished by crosses between inbred lines but, for the reasons noted in the introduction, this is not a viable option for dairy cattle in developed countries, in which all analyses must be based on analysis of the existing population. Detection of QTL in developing countries is considered below. For advanced commercial populations, the "daughter" and "granddaughter" designs, which make use of the existence of large half-sib families, are most appropriate for QTL analysis (Weller, Kashi and Soller, 1990). These designs are presented in Figures 2 and 3.

Both designs are similar to the backcross design for crosses between inbred lines in that only the alleles of one parent are followed in the progeny. Thus, similar to the backcross design, dominance cannot be estimated. These designs differ from crosses between inbred lines in that several families are analysed in which the linkage phase between QTL and genetic markers may differ. In addition, any specific QTL will be heterozygous in only a fraction of the families included in the analysis. Thus, QTL effects must be estimated within families, and these designs are therefore less powerful per individual genotyped than designs based on crosses between inbred lines.

The granddaughter design has the advantage of greater statistical power per

FIGURE 2
The daughter design



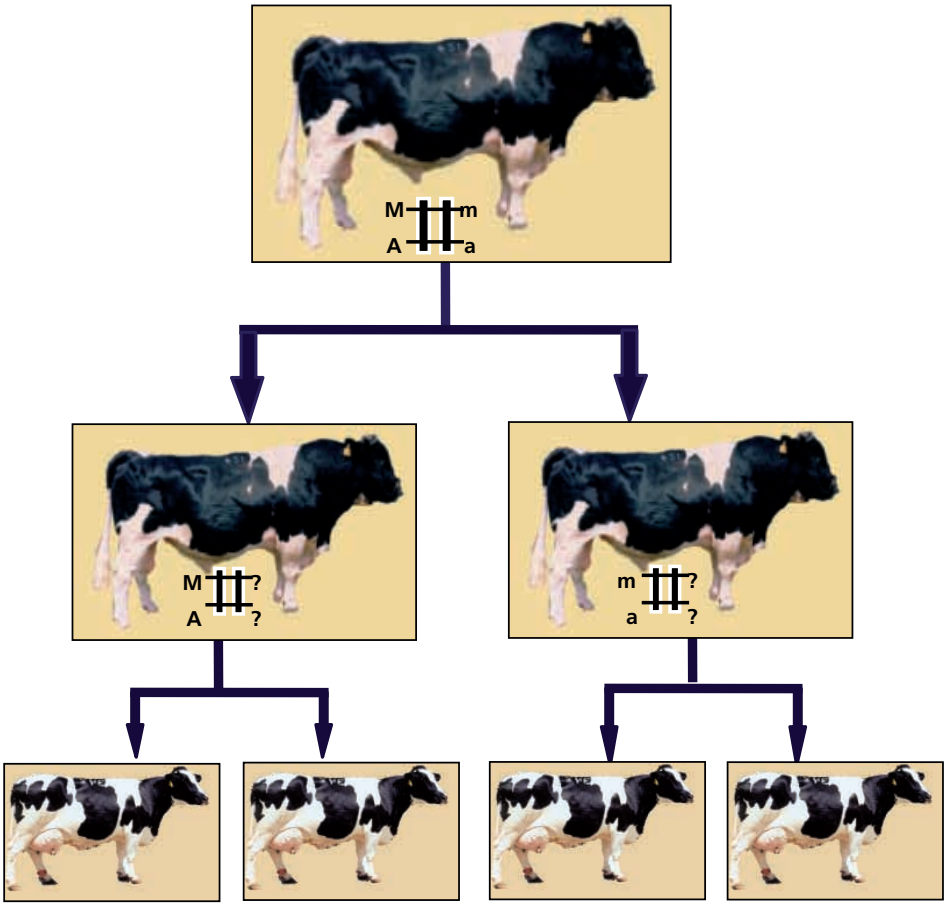
Only a single family is shown, although in practice several families will be analysed jointly. The sire is assumed to be heterozygous for a QTL and a linked genetic marker. The two alleles of the marker locus are denoted "M" and "m", and the two alleles of the QTL are denoted "A" and "a". Alleles of maternal origin are denoted by question marks.

individual genotyped. As each genotype is associated with multiple phenotypic records, the power per individual genotyped in the granddaughter design can be four-fold the power of the daughter design (Weller, Kashi and Soller, 1990). The disadvantage of this design is that the appropriate data structure (hundreds of progeny tested bulls, sons of a limited number of sires) is found only in the largest dairy cattle populations. Both daughter and granddaughter designs are less powerful per individual genotyped than designs based on analysis of inbred lines. Furthermore, the half-sib designs have the disadvantage that progeny with the same genotype as the sire are uninformative, because the progeny could have received either paternal allele.

Additional experimental designs have also been proposed. Coppieters *et al.* (1999) proposed the "great-granddaughter design". One of the disadvantages of the granddaughter design is that the number of progeny-tested sons of most sires is too low to obtain reasonable power to detect QTL of moderate effects. Coppieters *et al.* (1999) proposed that power can be increased by also genotyping progeny-tested grandsons of the grandsire. Inclusion of the grandsons is complicated by the fact that there is another generation of meiosis between the grandsire and his grandson.

A significant drawback of all the designs considered above is that they give no indication of the number of QTL alleles segregating in the population or their rela-

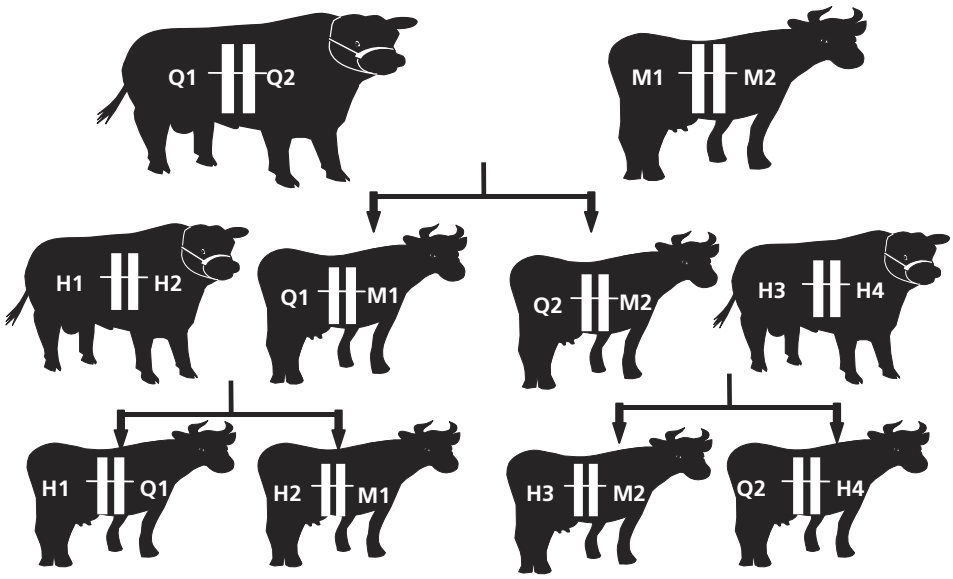
FIGURE 3
The granddaughter design



The grandsire is assumed to be heterozygous for a QTL and a linked genetic marker. As in Figure 2, only a single family is shown. The two alleles of the marker locus are denote "M" and "m", and the two alleles of the QTL are denoted "A" and "a". Alleles of maternal origin are denoted by question marks. Genotypes are not listed for the granddaughters because they were not genotyped.

tive frequencies. To answer this question, Weller *et al*. (2002) proposed the "modified granddaughter design" presented in Figure 4. Assume that a segregating QTL for a trait of interest has been detected and mapped to a short chromosomal segment using either a daughter or a granddaughter design. Consider the maternal granddaughters of a grandsire with a significant contrast between his two paternal alleles. This grandsire will be denoted the "heterozygous grandsire". Each maternal granddaughter will receive one allele from her sire, who is assumed to be unrelated to the heterozygous grandsire, and one allele from her dam, who is a daughter of the heterozygous grandsire. Of these granddaughters, one-quarter should receive the grandpaternal QTL allele with the positive effect, one-quarter should receive the

FIGURE 4
The modified granddaughter design



Only alleles for the QTL are shown. Alleles originating in the heterozygous grandsire are termed "Q1" and Q2". Alleles originating in the grand-dams are termed "M1" and "M2". Alleles originating in the sires are termed "H1", "H2", "H3" and "H4".

negative grandpaternal QTL allele, and half should receive neither grandpaternal allele. In the third case, the granddaughter received one of the QTL alleles of her grand-dam, the mate of the heterozygous grandsire. These grand-dams can be considered a random sample of the general population with respect to the allelic distribution of the QTL. All genetic and environmental effects not linked to the chromosomal segment in question are assumed to be randomly distributed among the granddaughters, or are included in the analysis model. Thus, unlike the daughter or granddaughter designs, it is possible to compare the effects of the two grandpaternal alleles with the mean QTL population effect.

Assuming that the QTL is "functionally biallelic" (i.e. there are only two alleles with differential expression relative to the quantitative trait), and that allele origin can be determined in the granddaughters, the relative frequencies of the two QTL alleles in the population can be determined by comparing the mean values of the three groups of granddaughters for the quantitative trait. Using the modified granddaughter design it is also possible to estimate the number of alleles segregating in the population, and to determine if the same alleles are segregating in different cattle populations. Weller *et al.* (2002) estimated the frequency of the QTL allele that increases fat and protein concentration on BTA6 in the Israeli Holstein population as 0.69 and 0.63, relative to fat and protein percent, by the modified granddaughter design. This corresponded closely to the frequency of 0.69 estimated for the Y581 allele of the ABCG2 gene for cows born during the same time period (Cohen-Zinder *et al.*, 2005).

METHODS TO ESTIMATE QTL EFFECTS AND LOCATION IN DAIRY CATTLE

If a significant effect on a quantitative trait is associated with a genetic marker, the difference between the means of marker genotype classes will be a biased estimate of the QTL effect due to recombination between the QTL and the genetic marker. Weller (1986) first demonstrated that maximum likelihood (ML) methodology could be used to obtain estimates of QTL location and effect unbiased by recombination, while Lander and Botstein (1989) proposed interval mapping, based on ML for a QTL bracketed between two markers. Haley and Knott (1992) and Martinez and Curnow (1992) proposed an interval mapping method based on non-linear regression, which was easier to apply than ML. Their methods are not directly applicable to half-sib designs because, as noted previously, linkage relationships between the QTL and the genetic markers will be different across families, and in some families the common ancestor will be homozygous for the QTL. Furthermore, if multiple QTL alleles are segregating in the population, or if the observed effect is due to several tightly linked QTL, the magnitude of the effect will also differ across families.

A method suitable for interval mapping that accounts for these problems has been developed by Knott, Elsen and Haley (1996) and has been applied to nearly all daughter and granddaughter design analyses. Their method is a modification of the non-linear regression method, and assumes a single QTL location for all families, but estimates a separate QTL effect for each family. This method has the advantage that, unlike ML, it can readily deal with missing and uninformative genotypes for some markers. Mackinnon and Weller (1995) proposed an ML method to estimate both QTL location and effect for half-sib designs under the assumption that only two QTL alleles are segregating in the population. Using this method it is also possible to estimate QTL genotype of the common parent of each family. However, these determinations are accurate only for relatively large QTL. The method of Mackinnon and Weller (1995) is more difficult to apply than the method of Knott, Elsen and Haley (1996), and has not come into general usage.

Lander and Botstein (1989) proposed the LOD-score (logarithm of the odds to the base 10) drop-off method to estimate confidence intervals for QTL location, but several studies have shown that this seriously underestimate the actual value (e.g. Darvasi *et al.*, 1993). The non-parametric bootstrap method (Visscher, Thompson and Haley, 1996) was found to be more accurate, but tends to overestimate confidence intervals. Bennewitz, Reinsch and Kalm (2003) proposed improvements to the bootstrap method that result in shorter but still unbiased confidence intervals.

Most studies to detect QTL in dairy cattle have considered many markers and multiple traits. In some studies nearly the entire genome was analysed, which raises a serious problem with respect to the appropriate threshold to declare significance. If normal point-wise significance levels of 5 or 1 percent are used, many marker-trait combinations will show "significance" by chance. While this is a problem for all QTL genome scans, it is even more severe for dairy cattle in which multiple half-sib families are analysed, in addition to multiple markers and traits. Several solutions to this problem have been proposed, none of which is completely satisfactory. The only solution to deal adequately with both multiple traits and families in addition to multiple markers is the false discovery rate (Weller *et al.*, 1998).

The QTL effects derived from either daughter or granddaughter by ML or non-linear regression will still be biased for several reasons. First, the usual assumptions of interval mapping, a single QTL segregating within the marker interval and no QTL in adjacent intervals, often do not reflect reality. Second, the dependant variable is generally an "adjusted" record, either daughter yield deviations (DYD; VanRaden and Wiggans, 1991) or genetic evaluations. Israel and Weller (1998) demonstrated that QTL effects derived from analysis of either genetic evaluations, yield deviations or DYD will be underestimated. In addition to this downward bias, there are two sources of upward bias for QTL effects. First, the direction of the effects is generally arbitrary, and therefore absolute values are retained and all effects are >0. Third, only the effects deemed "significant" are retained, and this is a selected sample (Georges *et al.,* 1995). Bayesian analysis methods that account for bias of QTL effect due to selection have recently been developed by Weller, Schlezinger and Ron (2005).

CURRENT STATUS OF QTL DETECTION IN DAIRY CATTLE

Genome scans by the granddaughter design have been completed for Holsteins from Canada (Nadesalingam, Plante and Gibson, 2001), the Netherlands (Spelman *et al.,* 1996; Schrooten *et al.,* 2000), France (Bennewitz, *et al.,* 2003a; Boichard *et al.,* 2003), Germany (Bennewitz, *et al.,* 2003a; Kuhn *et al.,* 2003a), New Zealand (Spelman *et al.,* 1999), and the United States (Georges *et al.,* 1995; Ashwell *et al.,* 1996, 1997, 1998a, 1998b, 2004; Ashwell, Van Tassell and Sonstegard, 2001; Zhang *et al.,* 1998; Ashwell and Van Tassell, 1999; Heyen *et al.,* 1999); Finnish Ayrshires (Vilkki *et al.,* 1997; Viitala *et al.,* 2003; Schulman *et al.,* 2004); French Normande and Montbeliarde cattle (Boichard *et al.,* 2003); Norwegian cattle in Norway (Klungland *et al.,* 2001; Olsen *et al.,* 2002); and Swedish Red and White (SRB) (Holmberg and Andersson-Eklund, 2004). Daughter design analyses have been performed for Israeli Holsteins (Mosig *et al.,* 2001; Ron *et al.,* 2004). Most studies have considered the five economic milk production traits: milk, fat and protein production, and fat and protein concentration, although a number of studies have also considered somatic cell score (SCS), female fertility, herd life, calving traits, health traits, temperament and conformation traits. The SCS is a log base 2 function of the concentration of somatic cells, and has been shown to be a useful indicator of udder health. Results are summarized in Table 1.

Results for milk, fat and protein production, fat and protein concentration, and SCS from most of the studies listed above are summarized at www.vetsci.usyd.edu.au/reprogen/QTL_Map/. Results from these traits, and many others including meat production, are summarized at http://bovineqtl.tamu.edu. Significant effects were found on all 29 autosomes, but most effects were found only in single studies and have not been repeated. Khatkar *et al.* (2004) performed a meta-analysis, combining data from most of these studies, and found significant across-study effects on chromosomes 1, 3, 6, 9, 10, 14 and 20.

METHODS OF INCORPORATING INFORMATION FROM GENETIC MARKERS IN GENETIC EVALUATION SYSTEMS

Heritabilities of most economic traits in dairy cattle are low to moderate. Genetic evaluation of dairy cattle is complicated by confounding between genetic and

TABLE 1
Summary of dairy cattle genome scans

Experimental design	Breed	Country	Traits analysed	References
Granddaughter	Ayrshire	Finland	Milk production[1]	Vilkki *et al.,* 1997; de Koning *et al.,* 2001; Viitala *et al.,* 2003
			SCS[2], mastitis, other treatments	Schulman *et al.,* 2004
	Jersey	New Zealand	Conformation	Spelman, Garrick and van Arendonk, 1999
	Holstein	Canadian	Milk production	Plante *et al.,* 2001
		France	Milk production	Boichard *et al.,* 2003
		Germany	Milk production	Thomsen *et al.,* 2001
			Functional	Kuhn *et al.,* 2003
			Conformation, temperament, milking speed	Hiendleder *et al.,* 2003
		Netherlands	conformation, SCS, fertility, calving, milking speed, gestation, birth weight, temperament	Schrooten *et al.,* 2000
		New Zealand	Conformation	Spelman, Garrick and van Arendonk, 1999
		USA	Milk production	Ashwell *et al.,* 1998b; Ashwell and Tassell 1999; Ashwell *et al.,* 1997, 2004; Ashwell, Van Tassell and Sonstegard, 2001;Georges *et al.,* 1995; Heyen *et al.,* 1999; Zhang *et al.,* 1998
			SCS	Ashwell *et al.,* 1996, 1997, 1998b; Ashwell and Van Tassell, 1999; Heyen *et al.,* 1999
			Herdlife	Heyen *et al.,* 1999
			Conformation	Ashwell *et al.,* 1998a, 1998b; Ashwell and Van Tassell, 1999
			Fertility	Ashwell *et al.,* 2004
	Montbeliarde	France	Milk production	Boichard *et al.,* 2003
	Normande	France	Milk production	Boichard *et al.,* 2003
	Norwegian	Norway	Milk production	Olsen *et al.,* 2002
			SCS, mastitis	Klungland *et al.,* 2001
	Swedish	Sweden	SCS, mastitis, other diseases	Holmberg and Andersson-Eklund, 2004
Daughter	Holstein	Israel	Milk production, SCS, fertility, herdlife	Ron *et al.,* 2004
			% protein	Mosig *et al.,* 2001

[1] Milk, fat, and protein production, and fat and protein concentration.
[2] Somatic cell concentration

environmental factors. Cows are scattered over many different herds with different management levels, and distribution of sires across herds is not random or orthogonal. Furthermore, cows generally produce multiple lactations that are correlated. In order to account for the limited heritability, and co-variances among relatives, genetic effects are generally assumed to be random, while most environmental effects are assumed to be fixed. Thus, genetic evaluation is performed by the mixed model using best linear unbiased prediction (BLUP) methodology (Henderson, 1984).

Beginning in the late 1980s, the model of choice for genetic evaluation for milk production traits was the individual animal model, in which a genetic effect is computed

for each animal, including animals that did not have production records (Westall and van Vleck, 1987). Genetic evaluations for these animals are derived via the numerator relationship matrix, which is included in the model. In addition, a "permanent environmental" effect is computed for each animal with records to account for similarities among multiple records of the same cow that are not due to additive genetic effects. As noted previously, analysis of QTL effects has generally been based on analysis of genetic evaluations or DYD, which are the adjusted means of the daughter records of a bull but which, unlike genetic evaluations, are not regressed. However, the statistical properties of DYD are not well understood, and QTL effects derived from analysis of DYD are still biased (Israel and Weller, 1998). Theoretically, it should be possible to derive unbiased QTL estimates if these effects are incorporated into a genetic evaluation scheme based on analysis of the actual records, such as the animal model. In practice, the inclusion of QTL effects into genetic evaluation models is complicated by three main factors:

- actual QTL location is unknown, and there is only partial linkage between genetic markers and QTL;
- linkage phase between genetic markers and QTL differs among individuals, and is generally unknown;
- only a small fraction of the population is genotyped.

An analysis including only genotyped individuals is not a viable option as it will generally not be possible to derive accurate fixed effects, such as herd-year-seasons, from this sample.

Fernando and Grossman (1989) proposed modifying the individual animal model described above to a "gametic" model that assumes the two QTL alleles of each individual are random effects sampled from a distribution with a known variance. They developed a method to estimate breeding values for all individuals in a population, including QTL effects via linkage to genetic markers, provided that all animals are genotyped and the heritability and recombination frequency between the QTL and the genetic marker are known. This model is suitable for any population structure and can also incorporate non-linked polygenic effects and other "nuisance" effects such as herd or block. The basic model assumes only a single record per individual, but can be adapted readily to a situation of multiple records per animal. This method is also denoted "marker-assisted BLUP" or "MA-BLUP".

Each individual with unknown ancestors is assumed to have two unique alleles for the QTL, which are "sampled" from an infinite population of alleles. For animals that are not genotyped, the probability of receiving either allele from either parent will be equal. However, if both the parent and progeny are genotyped for a linked genetic marker, then the probability of receiving a specific parental allele for a QTL linked to the genetic marker will be a function of the progeny marker genotype and recombination frequency. Based on these probabilities, Fernando and Grossman (1989) demonstrated how a variance-co-variance matrix could be constructed for the QTL gametic effects. They further described a simple algorithm to invert this matrix analogous to Henderson's method for inverting the numerator relationship matrix. This method has been extended to handle multiple markers and traits (Goddard, 1992). Cantet and Smith (1991) demonstrated that the number of equations could be significantly reduced by analysis of the reduced animal model.

The disadvantages of this model are that it assumes that both recombination frequency and the variance due to the QTL are known *a priori*. Studies on simulated data have demonstrated that although restricted maximum likelihood methodology can be used to estimate these parameters, they are completely confounded for a single marker locus (van Arendonk *et al.,* 1994). Methods to estimate the variance contributed by QTL with multiple markers were developed by Grignola, Hoeschele and Tier (1996). Furthermore, as each individual with unknown parents is assumed to have two unique alleles, the prediction error variances of the effects for any individual will be quite large and, therefore, not very informative. Finally, the assumption of a normal distribution of possible QTL allele effects may not be realistic.

Israel and Weller (1998) proposed an alternative method that assumes that only two QTL alleles are segregating in the population, and that either a daughter or granddaughter design has been applied to determine QTL genotypes of the family ancestors. The QTL effect is then included in the complete animal model analysis as a fixed effect. For individuals that are not genotyped, probabilities of receiving either allele are included as regression constants. These probabilities can be readily computed for the entire population using the segregation analysis method of Kerr and Kinghorn (1996). Israel and Weller (1998) assumed complete linkage between the QTL and a single marker. Israel and Weller (2002) extended the method to QTL analysis based on flanking marker, using the method of Whittaker, Thompsom and Visscher (1996) to estimate QTL effects and location from the regression estimates of flanking markers. This method has been tested extensively on simulated populations, and was able to derive unbiased estimates of QTL effect and location. It has also been applied to actual data from the Israeli Holstein population for a segregating QTL on chromosome 14 that affected milk production traits (Weller *et al.,* 2003). However, in this case the QTL effect was underestimated. Further research is required to determine the reason for this discrepancy.

METHODS FOR QTL DETECTION AND MAS IN DEVELOPING COUNTRIES

As noted previously, dairy cattle breeding in tropical and subtropical countries is generally based on crossbreeding between high production breeds adapted to temperate climates, and tropical strains which are adapted to the local environment, including resistance to local diseases. In other animal species, synthetic strains have been produced by selecting those individuals that retain the positive characteristics from both strains. For example, the Assaf sheep breed was produced from a cross between the Middle East Awassi breed and the East Friesian breed (www.sheep101.info/breedsA.html). In dairy cattle, the problem of an appropriate strategy for future generations has not been adequately solved, for reasons considered previously. If the economically important genes were identified, then the time and effort required for production of the desired synthetic strains could be reduced.

Visscher, Haley and Thompson (1996) considered the situation in which the recipient strain is an outbred population in an ongoing selection programme, and the introgressed genes are QTL. Markers flanking the QTL will be required in order to select backcross progeny that received the donor QTL allele. As there will be uncertainty with respect to the QTL location,

the flanking markers must be sufficiently close to the QTL so that it will be possible to determine with relative certainty that the QTL is in fact located between the flanking markers. Although marker-assisted introgression does decrease the number of generations required to obtain fixation of the desired allele, it increases two key cost elements. First, with traditional introgression, half of the progeny will carry the donor allele for the introgressed gene, and all of these can be used as parents in the next generation. However, if only a small fraction of the progeny is selected based on genetic markers, then many more individuals must be produced each generation. Second, genotyping costs for a large number of markers at each generation will also be significant.

Crosses between cattle breeds can also be used for QTL detection and they have been used in developing countries. In most plant species, the parental lines are completely inbred, and there will be complete LD in the F_2 or backcross generation. However, cattle are outbreeders and in crosses between breeds there will therefore only be partial LD between segregating QTL and linked genetic markers. Song, Soller and Genizi (1999) proposed the full-sib intercross line (FSIL) design for QTL detection and mapping for crosses between strains of outcrossing species. They assumed that the two parental strains differ in allelic frequencies, but were not at fixation for alternative QTL alleles.

For given statistical power, the FSIL design requires only slightly more individuals than an F_2 design derived from an inbred line cross, but six- to ten-fold fewer than a half-sib or full-sib design. In addition, as the population is maintained by continued intercrossing, DNA samples and phenotypic information can be accumulated across generations. Continued intercrossing in future generations also leads to map expansion, and thus to increased mapping accuracy in the later generations. An FSIL can therefore be used for fine mapping of QTL and this is considered below in detail.

Although these methods have not as yet been applied to detect QTL related to milk production, they have been applied to QTL for disease resistance. Trypanosomosis (sleeping sickness) is a major constraint on livestock productivity in sub-Saharan Africa. Hanotte *et al.* (2003) mapped QTL affecting trypanotolerance in a cross between the "tolerant" N'Dama breed and the susceptible Boran breed. Putative QTL affecting 16 traits associated with disease susceptibility were mapped tentatively to 18 autosomes. Excluding chromosomes with ambiguous effects, the allele associated with resistance was derived from the N'Dama strain for nine QTL and from the Boran strain for five QTL. These results are consistent with many plant crossbreeding experiments in which the strain with overall phenotypic inferiority for the quantitative trait nevertheless harbours QTL alleles that are superior to the alleles present in the phenotypically superior strain (e.g. Weller, Soller and Brody, 1988).

FROM QTL TO QTN – THEORY

As noted by Darvasi and Soller (1997), with a saturated genetic map, the resolving power for QTL will be a function of the experimental design, number of individuals genotyped and QTL effect. Weller and Soller (2004) computed that the 95 percent confidence interval (CI) in percent recombination for half-sib designs, including the daughter and granddaughter designs, was $3073/d^2N$, where d is the QTL substitution effect in units of the standard deviation, and

N is the sample size. In the case of a granddaughter design, the units for the standard deviation will be either units of the bulls' DYD or genetic evaluations. For example, if d is 0.5 and N is 400, the CI will be 31 percent recombination, or approximately 35 cM. Thus, except for the largest QTL, CIs will generally include several tens of cM. Considering that each cattle cM includes ~8 genes and one million bp, detection of the actual polymorphism responsible for the observed QTL effects (the quantitative trait nucleotide, QTN) appears at first glance to be a "mission impossible".

Various strategies have been proposed to reduce the CI based on multiple crosses, but most are not applicable to dairy cattle (e.g. Darvasi, 1998). Meuwissen and Goddard (2000) proposed that CI for QTL location could be reduced to individual cM by application of LD mapping. If a QTL polymorphism is due to a relatively recent mutation or to a relatively recent introduction from another population, then it should be possible to detect population-wide LD between the QTL and closely linked genetic markers. The closer the marker to the QTL, the greater will be the extent of LD. They developed a method to estimate QTL location and CI based on LD between a QTL and a series of closely linked markers. The CI can be further reduced by combining linkage and LD mapping (Meuwissen *et al.,* 2002), and by a multitrait analysis (Meuwissen and Goddard, 2004). However, unless the QTL effect is very large, the CI will still extend over several cM.

In order to determine the actual gene responsible for the QTL, most studies have used the "candidate gene" approach, i.e. to determine a likely candidate among the genes within the CI, based on known gene function, or specific gene expression in the organ of interest. Examples are given in the following section. However, even if a polymorphism is detected in the candidate gene and the polymorphism has a major LD effect on the QTL, how does one prove that this polymorphism is not merely in LD with the actual QTN?

Mackay (2001) proposed two alternatives for proof positive that a candidate polymorphism is in fact the QTN, namely, co-segregation of intragenic recombinant genotypes in a candidate gene with the QTL phenotype, and functional complementation where the trait phenotype is "rescued" in a transgenic organism. Neither of these is applicable to QTL in dairy cattle. In this case, Mackay (2001) postulated that the only option to achieve the standard of rigorous proof for identification of a gene underlying a QTL in commercial animal populations is to collect "multiple pieces of evidence, no single one of which is convincing, but which together consistently point to a candidate gene". Evidence can be provided by concordance of polymorphism with deduced QTL genotype, quantitative differences of gene expression in physiologically relevant organs, SNP capable of encoding a non-conservative amino acid change, protein differences in cows with contrasting genotypes for the QTN, orthologous QTL in other species (genes that are derived from a common ancestral gene) and alteration of gene protein in bovine cell lines by "short interfering RNA" (siRNA) technology. (The siRNA molecules bind with proteins to form a unit called the "RNA-induced silencing complex" that suppresses the expression of the gene to which it corresponds in the viral genome, silencing the gene from which the siRNA is derived.)

For dairy cattle, to date, the most compelling evidence is "concordance", i.e. that

the deduced QTL genotypes of a sample of individuals correspond completely to their genotypes for the putative QTN. All individuals heterozygous for the QTL should be heterozygous for the putative QTN, with the same QTN allele associated with the same QTL allele in all individuals, and all individuals homozygous for the QTL should also be homozygous for the QTN. Theoretically, the sample of individuals analysed should be large enough to reject statistically the hypothesis that concordance was obtained by chance. However, in dairy cattle, the only individuals for which QTL genotype can be derived with any level of reliability are sires that have been analysed by either a daughter or granddaughter design, and the number of these individuals will always be limited. Furthermore, there is at present no accepted theory to compute concordance probabilities by chance, considering that any polymorphism very close to the QTN will display significant LD. Several studies have addressed the problem (Cohen-Zinder *et al.*, 2005; Schnabel *et al.*, 2005). The case for identification of the QTN is clearly more compelling if concordance is obtained in two different populations.

FROM QTL TO QTN – RESULTS

To date, the QTN has been determined in two cases in dairy cattle, on BTA 6 and BTA 14. In both cases the QTL chiefly affected fat and protein concentration and the QTL effect was large enough that the confidence interval for QTL location was <10 cM. A QTL on BTA 14 near the centromere that chiefly affected fat quantity and both fat and protein concentration in both the United States and Israeli Holstein populations was first detected by Ron *et al.* (1998), and further studies were able to map the QTL to a region of approximately 10 cM (Coppieters *et al.*, 1999). In 2002, two studies independently showed that a mis-sense mutation, causing replacement of a lysine residue with alanine in exon VIII of the gene acylCoA:diacyglycerol acyltransferase (DGAT1), is the QTN (Grisart *et al.*, 2002; Winter *et al.*, 2002). Discovery was aided by the fact that DGAT1 was an obvious physiological candidate. In addition to mapping to the putative QTL region, DGAT1 encodes a microsomal enzyme that catalyses the final step of triglyceride synthesis and mice lacking both copies of DGAT1 are completely devoid of milk secretion. Complete concordance between this polymorphism and the QTL was found in three different dairy breeds.

The QTL near the middle of BTA 6 affecting protein concentration was first detected by Georges *et al.* (1995) in the United States Holstein population. This QTL was then detected in several other Holstein populations, including Finnish Ayrshire cattle (Velmala *et al.*, 1999) and Norwegian cattle (Olsen *et al.*, 2002). Ron *et al.* (2001) reduced the CI to 4 cM centred on microsatellite BM143. Olsen *et al.* (2002) used physical mapping and combined linkage and LD mapping to determine that this QTL is located within a 420 000 bp region between the genes ABCG2 and LAP3.

In 2005, two research groups claimed to have found the QTN in two different genes. Schnabel *et al.* (2005) claimed that the QTN is located in a poly-A sequence in the promoter region of the osteopontin gene, also denoted SPP1, while Cohen-Zinder *et al*. (2005) claimed that the QTN is a mis-sense mutation in exon 14 of the ABCG2 gene. Both studies based their claim on gene function and concordance of bulls with known genotypes. Both genes are dif-

ferentially expressed in the mammary gland during lactation, as compared with the liver. Furthermore, anti-sense SPP1 transgenic mice displayed abnormal mammary gland differentiation and milk secretion (Nemir *et al.,* 2000).

Schnabel *et al.* (2005) found concordance based on four heterozygous and four homozygous sires for the United States Holstein population, as determined by a granddaughter design, while Cohen-Zinder *et al.* (2005) found concordance for three heterozygous and 15 homozygous sires from both the United States and Israeli Holstein populations. Cohen-Zinder *et al.* (2005) also analysed the site proposed by Schnabel *et al.* (2005), and found that this site was hyper-variable in that at least four single nucleotide changes were found within the 20 bp region centred on the poly-A sequence. Eight of nine Israeli sires analysed by the daughter design were heterozygous for at least one of these polymorphisms.

Many studies have found a QTL affecting all five milk production traits and SCS near the middle of BTA 20. Blott *et al.* (2003) claimed that a mis-sense mutation in the bovine growth hormone receptor was responsible for the QTL affecting milk yield and composition on BTA 20, but did not find concordance for the bulls heterozygous for the QTL. Thus, this polymorphism may be responsible for only part of the observed effect on BTA 20, or may be a physiologically neutral mutation in LD with the QTN.

For both the QTL on BTA 6 and 14, the polymorphisms analysed apparently do not account for the entire effect observed in these chromosomal regions (Bennewitz *et al.,* 2004a; Kuhn *et al.,* 2004; Cohen-Zinder *et al.,* 2005). The effect associated with the mis-sense mutation in ABCG2 explains the entire effect observed on milk yield and fat and protein concentration, but does not explain the effects associated with fat and protein yield. It is likely that in the near future additional QTN will be resolved. As noted, the meta-analysis (Khatkar *et al.,* 2004) found significant effects on BTA 1, 3, 9 and 10, in addition to the effects described on BTA 6, 14 and 20.

METHODS AND THEORY FOR MAS IN DAIRY CATTLE

Considering the long generation interval, the high value of each individual, the very limited female fertility and the fact that nearly all economic traits are expressed only in females, it would seem that dairy cattle should be a nearly ideal species for application of MAS. However, most theoretical studies have been rather pessimistic with respect to the expected gains that can be obtained by MAS. As noted by Weller (2001), MAS can potentially increase annual genetic gain by increasing the accuracy of evaluation, increasing the selection intensity and decreasing the generation interval.

The following dairy cattle breeding schemes that incorporate MAS have been proposed:

- a standard PT system, with information from genetic markers being used to increase the accuracy of sire evaluations in addition to phenotypic information from daughter records (Meuwissen and van Arendonk, 1992);
- a MOET nucleus breeding scheme in which marker information is used to select sires for service in the MOET population, in addition to phenotypic information on half-sisters (Meuwissen and van Arendonk, 1992);
- PT schemes in which information on genetic markers is used to preselect young sires for entrance into the PT (Kashi,

Hallerman and Soller, 1990; Mackinnon and Georges, 1998);
- selection of bull sires without a PT, based on half-sib records and genetic markers (Spelman, Garrick and van Arendonk, 1999);
- selection of sires in a half-sib scheme, based on half-sib records and genetic markers (Spelman, Garrick and van Arendonk, 1999);
- use of genetic markers to reduce errors in parentage determination (Israel and Weller, 2000).

Meuwissen and van Arendonk (1992) found that inclusion of marker information to increase the accuracy of sire evaluations increased the rate of genetic gain by only 5 percent when the markers explained 25 percent of the genetic variance. This result is not surprising considering that the accuracy of sire evaluations based on a PT of 50 to 100 daughters is already quite high. In "open" and "closed" nucleus breeding schemes, rates of genetic gain were increased by 26 and 22 percent, respectively. The advantage of MAS in this case is greater, because young sires are not progeny tested, and their reliabilities based only on half-sib information are much lower.

Mackinnon and Georges (1998) proposed "top-down" and "bottom-up" strategies to apply the third scheme listed above, preselection of young sires prior to PT. In the "top-down" strategy, QTL genotypes are determined for the elite sires used as bull sires by a granddaughter design. If a dense marker map is available, it will then be possible to determine which QTL allele is passed to each son. Elite bulls from among these sons are then selected as bull sires for the next generation. If the original sire was heterozygous for a QTL, it can be determined which of his sons received the favourable allele. Sons of these sires are then genotyped and selected based on whether they received the favourable grandpaternal QTL alleles. It is assumed that the dams of the candidate sires are also genotyped, and that these cows will be progeny of the sires evaluated by a granddaughter design. Thus, grandpaternal alleles inherited via the candidates' dams can also be traced. A disadvantage of this scheme is that only the grandpaternal alleles are followed. Some of the sons of the original sires that were evaluated by a granddaughter design will also have received the favourable QTL allele from their dams, but not via the genotyped grandsires. However, young sires will be selected based only on the grandpaternal haplotypes.

In the "bottom-up" scheme, QTL genotypes of elite sires are determined by a daughter design. These sires are then used as bull sires. The candidate bulls are then pre-selected for those QTL heterozygous in their sires, based on which paternal haplotype they received. As the QTL phase is evaluated on the sires of the bull calves (the candidates for selection), no selection pressure is "wasted" as in the "top-down" scheme. In addition, this design can be applied to a much smaller population, because only several hundred daughters are required to evaluate each bull sire. On the negative side, more daughters than sons must be genotyped to determine QTL genotype. Mackinnon and Georges (1998) assumed that in either scheme it will not be necessary to increase mean generation interval above that of a traditional PT programme, although this will probably not be the case (Weller, 2001).

Kashi, Hallerman and Soller (1990), Mackinnon and Georges (1998), and Israel and Weller (2004) all addressed the problem that QTL determination will be subject to error. Deciding that a specific sire is homozygous for the QTL when in fact

the sire is heterozygous will be denoted the "type I" error. Deciding that the QTL is heterozygous in a specific sire, while the sire is in reality homozygous will be denoted the "type II" error. In the first case, segregating QTL will be missed while, in the second case, selection for the positive QTL allele will be applied to no advantage. All three studies found that genetic gains will be maximized with a relatively large proportion of type I errors, between 5 and 20 percent. This is due to the fact that as type I error increases, type II error decreases, and more real effects will be detected and applied in selection. A third type of error is theoretically possible, i.e. determining correctly that the ancestor is heterozygous for the QTL, but incorrect determination of QTL phase relative to the genetic markers. However, Israel and Weller (2004) showed by simulation that this error never occurred even when the type I error rate was set at 20 percent.

Spelman, Garrick and van Arendonk (1999) considered three different breeding schemes by deterministic simulation:

- a standard PT with the inclusion of QTL data;
- the same scheme except that young bulls without PT could also be used as bull sires based on QTL information;
- a scheme in which young sires could be used as both bull sires and cow sires in the general population, based on QTL information.

It was assumed that only bulls were genotyped but that, once genotyped, the information on QTL genotype and effect was known without error. It was then possible to conduct a completely deterministic analysis. They varied the fraction of the genetic variance controlled by known QTL from zero to 100 percent. Even without MAS, a slight gain was obtained by allowing young sires to be used as bull sires, and a genetic gain of 9 percent was obtained if young sires with superior evaluations were also used directly as both sires of sires and in general service. As noted previously, the genetic gain was limited where MAS was used only to increase the accuracy of young bull evaluations for a standard PT scheme because the accuracy of the bull evaluations was already high. Thus, even if all the genetic variance was accounted for by QTL, the genetic gain was less than 25 percent. However, if young sires are selected for general service based on known QTL, the rate of genetic progress can be doubled. The maximum rate of genetic gain that can be obtained in the third scheme, the "all bulls" scheme, was 2.2 times the rate of genetic gain in a standard PT. Theoretically, with half of the genetic variance due to known QTL, the rate of genetic gain obtained was greater than that possible with nucleus breeding schemes.

The final scheme, with use of genetic markers to reduce parentage errors, is the most certain to produce gains, as it does not rely on QTL genotype determination, which may be erroneous. Weller *et al.* (2004) genotyped 6 040 Israeli Holstein cows from 181 Kibbutz herds for 104 microsatellites. The frequency of rejected paternity was 11.7 percent, and most errors were due to inseminator mistakes. Most advanced breeding schemes already use genetic markers to confirm parentage of young sires. Israel and Weller (2002) found by simulations that if the parentage of bull dams and the test daughters of young sires are also verified, genetic gain increased by 4.3 percent compared with a breeding programme with 10 percent incorrect paternity. This scheme is economically justified if genotyping costs per individual are no more than US$15.

CURRENT STATUS OF MAS IN DAIRY CATTLE

Two ongoing MAS programmes in dairy cattle have been reported to date, in French and German Holsteins (Boichard *et al.,* 2002, 2006; Bennewitz *et al.,* 2004b). Currently in the German programme, markers on three chromosomes are used. The MA-BLUP evaluations (Fernando and Grossman, 1989) are computed at the VIT-computing centre in Verden, and are distributed to Holstein breeders who can use these evaluations for selection of bull dams and preselection of sires for progeny testing. The MA-BLUP algorithm only includes equations for bulls and bull dams, and the dependent variable is the bull's DYD (Bennewitz *et al.*, 2003b). Linkage equilibrium throughout the population is assumed. To close the gap between the grandsire families analysed in the German granddaughter design and the current generation of bulls, 3 600 bulls were genotyped in 2002. As then, about 800 bulls have been evaluated each year (N. Reinsch, personal communication). Only bulls and bull dams are genotyped as tissue samples are already collected for paternity testing. Thus additional costs due to MAS are low and even a very modest genetic gain can be economically justified. This scheme is similar to the "top-down" scheme of Mackinnon and Georges (1998) in that evaluation of the sons is used to determine which grandsires are heterozygous for the QTL and their linkage phase. This information is then used to select grandsons based on which haplotype was passed from their sires. It differs from the scheme of Mackinnon and Georges (1998) in that the grandsons are preselected for PT based on MA-BLUP evaluations, which include general pedigree information in addition to genotypes.

The French MAS programme includes elements of both the "top-down" and "bottom-up" MAS designs. Similar to the German programme, genetic evaluations including marker information were computed by a variant of MA-BLUP, and only genotyped animals and non-genotyped connecting ancestors were included in the algorithm. Genotyped females were characterized by their average performance based on pre-corrected records (with the appropriate weight), whereas males were characterized by twice the yield deviation of their non-genotyped daughters. Twelve chromosomal segments, ranging in length from 5 to 30 cM, are analysed. Regions with putative QTL affecting milk production or composition are located on BTA 3, 6, 7, 14, 19, 20 and 26; segments affecting mastitis resistance are located on BTA 10, 15 and 21; and chromosomal segments affecting fertility are located on BTA 1, 7 and 21. Each region was found to affect one to four traits and on average three regions with segregating QTL were found for each trait. Each region is monitored by two to four evenly spaced microsatellites, and each animal included in the MAS programme is genotyped for at least 43 markers. Sires and dams of candidates for selection, all male AI ancestors, up to 60 AI uncles of candidates, and sampling daughters of bull sires and their dams are genotyped. The number of genotyped animals was 8 000 in 2001 and 50 000 in 2006. An additional 10 000 animals are genotyped per year, with equal proportions of candidates for selection and historical animals.

FUTURE PROSPECTIVE FOR MAS IN DAIRY CATTLE

Although the first large experiment in QTL detection in dairy cattle was published in 1961 by Neimann-Sørensen and Robertson, in 1985 it still looked as if

MAS was a long way off for commercial animal populations as there were very few known genetic markers and methodology was rudimentary. In the last 20 years there have been huge advances in both DNA technology and statistical methodology, and it can now be stated with near certainty that the technology is available to detect and map accurately segregating QTL in dairy cattle. Furthermore, although many effects reported in the literature are "false positives", there is a wealth of evidence that several QTL are in fact real as a number of effects have been repeated across numerous experiments, and the actual QTN have been identified for at least two QTL.

The main limitation at this point to detecting and mapping more QTL is the sample sizes available, especially the number of progeny tested bulls per family. To map QTL of smaller magnitude accurately, it will be necessary to combine data across experiments (e.g. Khatkar *et al.,* 2004) or significantly increase sample sizes. This can only be done by genotyping cows, even though power per individual genotyped will be lower.

The fact that only two countries have actually started MAS programmes highlights the current limitations to practical application of MAS. To date, very few segregating QTL with economic impact have been identified in commercial dairy cattle populations. Of the two QTNs that have been detected, each has disadvantages with respect to application in MAS. The allele of DGAT1 that increases fat production and decreases water content in the milk, both desirable, also decreases protein yield, which is undesirable (Weller *et al.,* 2003). The allele of ABCG2 that decreases milk production and increases protein percent is clearly the favourable allele in nearly all current selection indices, but this allele is already at a very high frequency in all major dairy cattle populations (Ron *et al.,* 2006).

In addition to the limitation of definitively identified QTL with economic value, suitable software for genetic evaluation including QTL effects is also a limiting factor. At present, those countries that are applying MAS are using two-step procedures, i.e. a preliminary analysis to compute genetic evaluations based only on pedigree and phenotypic data, and then a second analysis in which the genetic evaluations are "adjusted" for QTL effects. Ideally a single algorithm should be used to derive genetic evaluations for the entire population including the effects of known QTL.

ACKNOWLEDGEMENTS

I thank M. Ron, E. Seroussi and M. Ashwell for their input.

REFERENCES

Ashwell, M.S. & Van Tassell, C.P. 1999. Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. *J. Dairy Sci.* 82: 2497–2502.

Ashwell, M.S., Van Tassell, C.P. & Sonstegard, T.S. 2001. A genome scan to identify quantitative trait loci affecting economically important traits in a US Holstein population. *J. Dairy Sci.* 84: 2535–2542.

Ashwell, M.S., Rexroad, C.E., Miller, R.H. & VanRaden, P.M. 1996. Mapping economic trait loci for somatic cell score in Holstein cattle using microsatellite markers and selective genotyping. *Anim. Genet*. 27: 235–242.

Ashwell, M.S., Rexroad, C.E., Miller, R.H., VanRaden, P.M. & Da, Y. 1997. Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers. *Anim. Genet*. 28: 216–222.

Ashwell, M.S., Da, Y., VanRaden, P.M., Rexroad, C.E. & Miller, R.H. 1998a. Detection of putative loci affecting conformational type traits in an elite population of United States Holsteins using microsatellite markers. *J. Dairy Sci*. 81: 1120–1125.

Ashwell, M.S., Da, Y., Van Tassell, C.P., VanRaden, P.M., Miller, R.H. & Rexroad, C.E. 1998b. Detection of putative loci affecting milk production and composition, health, and type traits in a United States Holstein population. *J. Dairy Sci*. 81: 3309–3314.

Ashwell, M.S., Heyen, D.W., Sonstegard, T.S., Van Tassell, C.P., DA, Y., VanRaden, P.M., Ron, M., Weller, J.I. & Lewin, H.A. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Dairy Sci*. 87: 468–475.

Bennewitz, J., Reinsch, N. & Kalm, E. 2003. Comparison of several bootstrap methods for bias reduction of QTL effect estimates. *J. Anim. Breed. Genet*. 120: 403–416.

Bennewitz, J., Reinsch, N., Grohs, C., Leveziel, H., Malafosse, A., Thomsen, H., Xu, N.Y., Looft, C., Kuhn, C., Brockmann, G.A., Schwerin, M., Weimann, C., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Forster., M., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Blumel, J., Boichard, D. & Kalm, E. 2003a. Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genet. Sel. Evol.* 35: 319–38.

Bennewitz, J., Reinsch, N., Thomsen, H., Szyda, J., Reinhart, F., Kuhn, C., Schwerin, M., Erhardt, G., Weimann, C. & Kalm, E. 2003b. Marker assisted selection in German Holstein dairy cattle breeding: outline of the program and marker assisted breeding value estimation. *Ann. Meet. Europ. Ass. Anim. Prod*. 54th Session G1.9. Rome, Italy.

Bennewitz, J., Reinsch, N., Paul, S., Looft, C., Kaupe, B., Weimann, C., Erhardt, G., Thaller, G., Kuhn, C., Schwerin, M., Thomsen, H., Reinhardt, F., Reents, R. & Kalm, E. 2004a. The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. *J. Dairy Sci*. 87: 431–442.

Bennewitz, J., Reinsch, N., Reinhardt, F., Liu, Z. & Kalm, E. 2004b. Top down preselection using marker assisted estimates of breeding values in dairy cattle. *J. Anim. Breed. Genet*. 121: 307–318.

Blott, S., Kim, J., Moisio, S., Schmidt-Kuntzel, A., Cornet, A., Berzi, P., Cambisano, N., Ford, C., Grisart, B., Johnson, D., Karim, L., Simon, P., Snell, R., Spelman, R., Wong, J., Vilkki, J., Georges, M., Farnir, F. & Coppieters, W. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163: 253–266.

Boichard, D., Fritz, S., Rossignol, M.N., Boscher, M.Y., Malafosse, A. & Colleau, J.J. 2002. Implementation of marker-assisted selection in French dairy cattle. *Proc. 7th World Cong. Genet. Appl. Livest. Prod.* 33: 19–22. Montpellier, France.

Boichard, D., Fritz, S., Rossignol, M.N., Guillaume, F., Colleau, J.J. & Druet, T. 2006. Implementation of marker-assisted selection: practical lessons from dairy cattle. *Proc. 8th World Cong. Genet. Appl. Livest. Prod.* 22: 11. 13–18 August 2006, Belo Horizonte, MG, Brazil.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M.Y. & Leveziel, H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet. Sel. Evol*. 35: 77–101.

Brascamp, E.W., van Arendonk J.A.M. & Groen A.F. 1993. Economic appraisal of the utilization of genetic markers in dairy cattle breeding. *J. Dairy Sci*. 76: 1204–1213.

Cantet, R.J.C. & Smith, C. 1991. Reduced animal model for marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol*. 23: 221–233.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-van der Wind, A., Lee, J.H., Drackley, J.K., Mark R. Band, M.R., Hernandez, A.G., Shani, M., Lewin, H.A., Weller, J.I. & Ron, M. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* 15: 936–944.

Coppieters, W., Kvasz, A., Arranz, J.J., Grisart, B., Riquet, J., Farnir, F. & Georges, M. 1999. The great-grand-daughter design: a simple strategy to increase the power of a grand-daughter design for QTL mapping. *Genetical Research* 74: 189–199.

Cunningham, E.P. 1989. Formulation of breeding plans for dairy and dual-purpose cattle. *Brazil. J. Genet*. 12: 81.

Darvasi, A. 1998. Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet*. 18: 19–24.

Darvasi, A. & Soller, M. 1997. A simple method to calculate resolving power and confidence interval of QTL map location. *Behavior. Genet*. 27: 125–132.

Darvasi, A., Vinreb, A., Minke, V., Weller, J.I. & Soller, M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943–951.

de Koning, D.J., Schulman, N.F., Elo, K., Moisio, S., Kinos, R., Vilkki, J. & Maki-Tanila, A. 2001. Mapping of multiple quantitative trait loci by simple regression in half-sib designs. *J. Anim. Sci*. 79: 616–622.

de Vaccaro, L. 1990. Survival of European dairy breeds and their crosses with Zebus in the tropics. *Anim. Breed. Abstr*. 58: 475–494.

Fernando R.L. & Grossman, M. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol*. 21: 467–477.

FAO. 1987. Crossbreeding *Bos indicus* and *Bos taurus* for milk production in the tropics, by E.P. Cunningham & O. Syrstad. *Anim. Prod. Health Paper No. 68*. Rome.

Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq., F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* 10: 220–227.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargent, L.S., Sorensen, A., Steele, M.R., Zhao, X., Womack, J.E. & Hoeschele, I. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139: 907–920.

Goddard, M.E. 1992. A mixed model for analyses of data on multiple genetic markers. *Theor. Appl. Genet.* 83: 878–886.

Grignola, F.E., Hoeschele, I. & Tier, B. 1996. Mapping quantitative trait loci in outcross populations via residual maximum likelihood. I. Methodology. *Genet. Sel. Evol*. 28: 479–490.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C. Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, R. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* 12: 222–231.

Haley, C.S. & Knott, S.A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315–324.

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., Sugimoto, Y., Kemp, S., Gibson, J., Korol, A., Soller, M. & Teale, A. 2003. Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Nat. Acad. Sci. USA* 100: 7443–7448.

Henderson, C.R. 1984. *Applications of linear models in animal breeding*. University of Guelph, Ontario, ON, Canada.

Heyen, D.W., Weller J.I., Ron, M., Band, M., Beever, J.E., Feldmesser, E., Da, Y., Wiggans, G.R., VanRaden, P.M. & Lewin, H.A. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiol. Genom*. 1: 165–175.

Hiendleder, S., Thomsen, H., Reinsch, N., Bennewitz, J., Leyhe-Horn, B., Looft, C., Xu, N., Medjugorac, I., Russ, I., Kuhn, C., Brockmann, G.A., Blumel, J., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Schwerin, M., Forster, M., Kalm, E. & Erhardt, G. 2003. Mapping of QTL for body conformation and behavior in cattle. *J. Heredity* 94: 496–506.

Hill, W.G. 1971. Investment appraisal for national breeding programmes. *Anim. Prodn*. 13: 37–50.

Holmberg, M. & Andersson-Eklund, L. 2004. Quantitative trait loci affecting health traits in Swedish dairy cattle. *J. Dairy Sci.* 87: 2653–2659.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945.

Israel, C. & Weller, J.I. 1998. Estimation of candidate gene effects in dairy cattle populations. *J. Dairy Sci*. 81: 1653–1662.

Israel, C. & Weller, J.I. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *J. Dairy Sci*. 83: 181–187.

Israel, C. & Weller, J.I. 2002. Estimation of quantitative trait loci effects in dairy cattle populations. *J. Dairy Sci*. 85: 1285–1297.

Israel, C. & Weller, J.I. 2004. Effect of type I error threshold on marker-assisted selection in dairy cattle. *Livest. Prod. Sci.* 85: 189–199.

Kashi, Y., Hallerman, E. & Soller, M. 1990. Marker-assisted selection of candidate bulls for progeny testing programmes. *Anim. Prod*. 51: 63–74.

Kerr, R. J. & Kinghorn, B.P. 1996. An efficient algorithm for segregating analysis in large populations. *J. Anim. Breed. Genet*. 113: 457–469.

Khatkar, M.S., Thomson, P.C., Tammen, I. & Raadsma, H.W. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol*. 36: 163–190.

Klungland, H., Sabry, A., Heringstad, B., Olsen, H.G., Gomez-Raya, L., Vage, D.I., Olsaker, I., Odegard, J., Klemetsdal, G., Schulman, N., Vilkki, J., Ruane, J., Aasland, M., Ronningen, K. & Lien, S. 2001. Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mamm. Genome* 12: 837–842.

Knott, S.A., Elsen J.M. & Haley, C.S. 1996. Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor. Appl. Genet*. 93: 71–80.

Kosgey, I. S., Kahi, K.A. & van Arendonk J.A.M. 2005. Evaluation of closed adult nucleus multiple ovulation and embryo transfer and conventional progeny testing breeding schemes for milk production in tropical crossbred cattle. *J. Dairy Sci*. 88: 1582–1594.

Kuhn, C., Bennewitz, J., Reinsch, N., Thomsen, H., Looft, C., Brockmann, G.A., Schwerin, M., Weimann, C., Hiendieder, S., Erhardt, G., Medjugorac, I., Forster, M., Brenig, B., Reinhardt, F., Reents, R., Russ, I., Averdunk, G., Blumel, J. & Kalm, E. 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J. Dairy Sci*. 86: 360–368.

Kuhn, C., Thaller, G., Winter, A., Bininda-Emonds, O.R.P., Kaupe, B., Erhardt, G., Bennewitz, J., Schwerin, M. & Fries, R. 2004. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative tip trait locus with major effect on milk fat content in cattle. *Genetics* 167: 1873–1881.

Lander, E.S. & Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Mackay, T.F. 2001. The genetic architecture of quantitative traits. *Ann. Rev. Genet.* 35: 303–339.

Mackinnon, M.J. & Georges, M.A.J. 1998. Marker-assisted preselection of young dairy sires prior to progeny-testing. *Livest. Prod. Sci.* 54: 229–250.

Mackinnon, M.J. & Weller, J.I. 1995. Methodology and accuracy of estimation of quantitative trait loci parameters in a half-sib design using maximum likelihood. *Genetics* 141: 755–770.

Madalena, F.E. 1993. A simple scheme to utilize heterosis in tropical dairy cattle. *World Anim. Rev.* 74/75: 17–25.

Martinez, O. & Curnow, R.N. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet*. 85: 480–488.

Meuwissen, T.H.E. & Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155: 421–430.

Meuwissen, T.H.E. & Goddard, M.E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol* 36: 261–279.

Meuwissen, T.H.E. & van Arendonk, J.A.M. 1992. Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. *J. Dairy Sci.* 75: 1651–1659.

Meuwissen, T.H.E., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M.E. 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 16: 373–379.

Mosig, M.O., Lipkin, E., Khutoreskaya, G., Tchourzyna, E., Soller, M. & Friedmann, A. 2001. A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* 157: 1683–1698.

Nadesalingam, J., Plante, Y. & Gibson, J.P. 2001. Detection of QTL for milk production on chromosomes 1 and 6 of Holstein cattle. *Mammal. Genome* 12: 27–31.

Neimann-Sørensen, A. & Robertson, A. 1961. The association between blood groups and several production characters in three Danish cattle breeds. *Acta. Agr. Scand*. 11: 163–196.

Nemir, M., Bhattacharyya, D., Li, X., Singh, K., Mukherjee, A.B. & Mukherjee, B.B. 2000. Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency. *J. Biol. Chem*. 275: 969–976.

Nicholas, F.W. & Smith C. 1983. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim Prod*. 36: 341–353.

Olsen, H.G., Gomez-Raya, L., Vage, D.I., Olsaker, I., Klungland, H., Svendsen, M., Adnoy, T., Sabry, A., Klemetsdal, G., Schulman, N., Kramer, W., Thaller, G., Ronningen, K. & Lien, S.

2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *J. Dairy Sci.* 85: 3124–3130.

Plante, Y., Gibson, J.P., Nadesalingam, J., Mehrabani-Yeganeh, H., Lefebvre, S., Vandervoort, G. & Jansen, G.B. 2001. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *J. Dairy Sci*. 84: 1516–1524.

Powell, R.L. & Norman, H.D. 1988. Accuracy of cow indexes according to repeatability, evaluation, herd yield, and registry status. *J. Dairy Sci*. 71: 2232–2240.

Ron, M., Heyen, D.W., Weller, J.I., Band, M., Feldmesser, E., Pasternak, H., Da, Y., Wiggans, G.R., Vanraden, P.M., Ezra, E. & Lewin, H.A. 1998. Detection and analysis of a locus affecting milk concentration in the US and Israeli dairy cattle populations. *Proc. 6th World Cong. Genet. Appl. Livest. Prod*. 26: 422–425. Armidale, Australia.

Ron, M., Kliger D., Feldmesser E., Seroussi E., Ezra E. & Weller J. I. 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design. *Genetics* 159:727–735.

Ron, M., Cohen-Zinder, M., Peter, C., Weller, J.I. & Erhardt, G. 2006. ABCG2 polymorphism in *Bos indicus* and *Bos taurus* cattle breeds. *J. Dairy Sci.* **89**: 4921–4923.

Ron, M., Feldmesser, E., Golik, M., Tager-Cohen, I., Kliger, D., Reiss, V., Domochovsky, R., Alus, O., Seroussi, E., Ezra, E. & Weller, J.I. 2004. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. *J. Dairy Sci*. 87: 476–490.

Schnabel, R.D., Kim, J.J., Ashwell, M.S., Sonstegard, T.S., Van Tassell, C.P., Connor, E.E. & Taylor, J.F. 2005. Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine osteopontin gene. *Proc. Nat. Acad. Sci. USA* 102: 6896–6901.

Schrooten, C., Bovenhuis, H., Coppieters, W. & van Arendonk, J.A.M. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J. Dairy Sci*. 83: 795–806.

Schulman, N.F., Viitala, S.M., de Koning, D.J., Virta, J., Maki-Tanila, A. & Vilkki, J.H. 2004. Quantitative trait loci for health traits in Finnish Ayrshire cattle. *J. Dairy Sci*. 87: 443–449.

Smith, C. 1988. Genetic improvement of livestock, using nucleus breeding units. *World Anim. Rev*. 65: 2–10.

Song, J-Z., Soller, M. & Genizi, A. 1999. The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. *Genet. Research* 73: 61–73.

Spelman, R.J., Garrick, D.J. & van Arendonk, J.A.M. 1999. Utilization of genetic variation by marker assisted selection in commercial diary cattle populations. *Livest. Prod. Sci.* 59: 51–60.

Spelman, R.J., Coppieters, W., Karim, L., van Arendonk, J.A.M. & Bovenhuis, H. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144: 1799–1808.

Spelman, R.J., Huisman, A.E., Singireddy, S.R., Coppieters, W., Arranz, J., Georges, M. & Garrick, D.J. 1999. Quantitative trait loci analysis on 17 nonproduction traits in the New Zealand dairy population. *J. Dairy Sci*. 82: 2514–2516.

Thomsen, H., Reinsch, N., Xu, N., Looft, C., Grupe, S., Kuhn, C., Brockmann, G.A., Schwerin, M., Leyhe-Horn, B., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Forster, M., Brenig, B., Reinhardt, Kalm, E. 2001. Comparison of estimated breeding values, daughter yield deviations and deregressed proofs within a whole genome scan for QTL. *J. Anim. Breed. Genet*. 118: 357–370.

van Arendonk, J.A.M., Bovehuis, H., van der Beek, S. & Groen, A.F. 1994. Detection and exploitation of markers linked to quantitative traits in farm animals. *Proc. 5th World Cong. Genet. Appl. Livest. Prod*. 21:193–200. Guelph, Canada.

VanRaden, P.M. & Wiggans, G.R. 1991. Derivation, calculation and use of national animal model information. *J. Dairy Sci*. 74: 2737–2746.

Velmala, R.J., Vilkki, H.J., Elo, K.T., de Koning, D.J. & Maki-Tanila, A.V. 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. *Anim. Genet*. 30: 136–143.

Viitala, S.M., Schulman, N.F., de Koning, D.J., Elo, K., Kinos, R., Virta, A., Virta, J., Maki-Tanila, A. & Vilkki, H.J. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *J. Dairy Sci*. 86: 1828–1836.

Vilkki, H.J., de Koning, D.J., Elo, K.T., Velmala, R. & Maki-Tanila, A. 1997. Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression, *J. Dairy Sci.* 80: 198–204.

Visscher, P.M., Haley, C.S. & Thompson, R. 1996. Marker-assisted introgression in backcross breeding programs. *Genetics* 144: 1923–1932.

Visscher, P.M., Thompson, R. & Haley, C.S. 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143: 1013–1020.

Weller, J.I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42: 627–640.

Weller, J.I. 1994. *Economic aspects of animal breeding.* London, Chapman & Hall. 244 pp.

Weller, J.I. 2001. *Quantitative trait loci analysis in animals*. London, CABI Publishing. 287 pp.

Weller, J.I. & Soller, M. 2004. An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. *Theor. Appl. Genet.* 109: 1224–1229.

Weller, J. I., Kashi, Y. & Soller, M. 1990. Power of "daughter" and "granddaughter" designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. *J. Dairy Sci*. 73: 2525–2537.

Weller, J.I., Shlezinger, M. & Ron, M. 2005. Correcting for bias in estimation of quantitative trait loci effects. *Genet. Sel. Evol*. 37: 501–522.

Weller, J.I., Soller, M. & Brody, T. 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* x *L. pimpinellifolium*) by means of genetic markers. *Genetics* 118: 329–339.

Weller, J.I., Song, J.-Z., Heyen, D.W., Lewin, H.A. & Ron M. 1998. A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. *Genetics* 150: 1699–1706.

Weller, J.I., Weller, H., Kliger, D. & Ron, M. 2002. Estimation of quantitative trait locus allele frequency via a modified granddaughter design. *Genetics* 162: 841–849.

Weller, J.I., Golik, M., Seroussi, E., Ezra, E. & Ron, M. 2003. Population-wide analysis of a QTL affecting milk-fat production in the Israeli Holstein population. *J. Dairy Sci*. 86: 2219–2227.

Weller, J.I., Feldmesser, E., Golik, M., Tager-Cohen, I., Domochovsky, R., Alus, O., Ezra, E. & Ron, M. 2004. Factors affecting incorrect paternity assignment in the Israeli Holstein population. *J. Dairy Sci.* 87: 2627–2640.

Westell, R.A. & van Vleck, L.D. 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *J. Dairy Sci*. 70: 1106–1017.

Whittaker, J.C., Thompsom, R. & Visscher, P.M. 1996. On the mapping of QTL by regression of phenotype on marker-type. *Heredity* 77: 23–32.

Winter, A., Kramer, W., Werner, F.A.O., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J.E., Thaller, G. & Fries, R. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc. Nat. Acad. Sci. USA* 99: 9300–9305.

Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L.A., Grignola, F.E., Uimari, P., Thaller, G. & Bishop, M.D. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149: 1959–1973.

CHAPTER 13

Marker-assisted selection in sheep and goats

Julius H.J. van der Werf

SUMMARY

Sheep and goats are often kept in low input production systems, often at subsistence levels. In such systems, the uptake of effective commercial breeding programmes is limited, let alone the uptake of more advanced technologies such as those needed for marker-assisted selection (MAS). However, effective breeding programmes exist in a number of countries, the largest ones in Australia and New Zealand aiming for genetic improvement of meat and wool characteristics as well as disease resistance and fecundity. Advances have been made in sheep gene mapping with the marker map consisting of more than 1 200 microsatellites, and a virtual genome sequence together with a very dense single nucleotide polymorphism (SNP) map are expected within a year. Significant research efforts into quantitative trait loci (QTL) are under way and a number of commercial sheep gene tests have already become available, mainly for single gene effects but some for muscularity and disease resistance. Gene mapping in goats is much less advanced with mainly some activity in dairy goats. Integration of genotypic information into commercial genetic evaluation and optimal selection strategies is a challenge that deserves more development.

INTRODUCTION

The benefits of marker-assisted selection (MAS) to sheep and goat breeding programmes depend on a number of conditions that are relevant for most breeding programmes across species. These conditions include the existence of a genotype test predicting phenotypic differences, the economic value of these differences and the value of the genotypic information within the breeding programme. The value of genetic information will depend heavily on the socio-economic context of the breeding programme and the production system. In a technical sense, the value of this information is basically driven by the increase in selection accuracy resulting from knowledge of genotypes, which in turn will differ between animals from different age classes. In particular, the relative increase in selection accuracy of the youngest selection candidates will be critical to the value of MAS. However, technical arguments about increased selection accuracy are of little value if these selection criteria are poorly developed or accepted within the production system.

The application of new technologies such as MAS in animal breeding programmes therefore depends not only on a number of technical aspects associated with increased rates of genetic improvement, but also on the commercial structures of the industry. For example, the uptake of MAS in breeding programmes depends on the willingness of breeders to invest in genotypic information, and their ability to turn this into knowledge that helps them improve their commercial breeding activities. A basic understanding of breeding programme characteristics, the possible role of genetic information within these programmes, and the commercial relationships among the different players are needed to assess the value and predict the application of MAS in breeding programmes. These commercial relationships are distinctly different in sheep and goat breeding programmes from those in the more intensive animal industries, and the application of MAS will therefore be different. For example, 96 percent of the world goat population is kept by smallholders in developing countries, and genetic improvement programmes are rare (Olivier *et al.*, 2005).

The purpose of this chapter is to describe the use of MAS in breeding programmes for sheep and goats and the likely rate of uptake of this technology in these species. It begins by characterizing such programmes and describing and comparing existing programmes. MAS is most useful for traits that cannot be improved easily by phenotypic selection, either because they are difficult to measure on young animals (before sexual reproduction), or because of low heritability. Therefore, breeding objectives are discussed in general terms and the traits that are particularly suitable for MAS are identified. Based on some general well-known advantages of MAS, its possible role within breeding programmes can be predicted and examples of these are provided. Examples of marked genes are then described and an overview given of the status of "gene discovery" and gene mapping projects in sheep and goats. The chapter concludes by describing cases of using this information in actual breeding programmes. Some gene tests are based on actual functional mutations, many of which do not affect quantitative traits that are generally targeted in breeding programmes. Although, the term "MAS" should be replaced in some cases by "genotype assisted selection" (GAS), the term MAS is used loosely to refer to all selection based on genotypic information. It will become clear that

currently most real applications of MAS for sheep and goat breeding are based on research projects and therefore subsidized. However, the first commercial applications are now also emerging. The main requirements for a successful commercial and long-term application of MAS in sheep and goat breeding are discussed and illustrated based on examples.

MAS applications are often illustrated or simulated for pure breeding programmes. However, MAS could be particularly useful in crossbreeding programmes where desirable genotypes in unfavourable backgrounds are introgressed into productive local breeds with overall better breeding values. The opposite is also possible, where disease resistance genes of local breeds are specifically targeted in upgrading programmes with imported stock with higher productivity being crossed to local breeds. Crossbreeding and introgression programmes are discussed and, as sheep and goat production is relatively predominant in developing countries, particular attention is given to breeding programmes for low to medium input production systems.

CHARACTERISTICS OF SHEEP AND GOAT BREEDING PROGRAMMES

Breeding structures

Breeding programmes for sheep and goats generally operate within an industry that is based on low levels of resource inputs, i.e. low levels of feeding and low labour costs on a per animal basis. Goat production takes place largely in developing countries where selective breeding based on performance recording is often absent. A more substantial proportion of sheep production is found in developed countries such as Australia, France, New Zealand, South Africa and the United Kingdom. These systems are also predominantly pastoral-based and extensive in nature. An FAO working group report (Hoste, 2002; Olivier *et al*., 2005) made the following distinction between production systems and the opportunities within them for breeding programmes: 1) subsistence-based production, among the world's poorest, with limited market development and limited inputs and scope for genetic improvement; 2) market-based production, with better developed markets targeting urban populations, higher input levels and more specialized production systems, with scope for genetic improvement depending on cost of inputs and also on skills and information literacy of breeders and producers; and 3) high-input production, with further specialization, emphasis on increased land and labour efficiency, and much more concern for food quality, food safety, animal welfare and the environment. Most of the world's goat production as well as many of the sheep systems would fall into the first category, whereas sheep production in developed countries would mainly fall into the second category, with some of these working towards the third category.

Sheep and goat breeding programmes are characterized by a flat breeding structure, meaning that compared with intensive livestock industries many operations participate in genetic improvement, thereby forming a wide base for the nucleus breeding sector. Reproductive levels of breeding animals, especially males, are relatively low compared with other species. In such a system, the multiplication factor, i.e. the number of commercial expressions resulting from investments in improved genotypes in the breeding nucleus, is relatively low. This makes it more difficult to introduce new technologies and justify large investments in improving individual animals. However, like other breeding programmes, there remains

a significant return on overall investment in genetic improvement. Also, in some more advanced sheep breeding programmes, the use of artificial insemination (AI) and across-flock evaluation has boosted the use of high profile rams and raised the value of individual breeding animals.

The main investment in breeding programmes is for performance recording. The extent of trait measurement is often quite closely aligned with the intensity of the production system. Input levels for sheep production vary, depending on breed type and market. In Australia, for example, there is a significant difference between wool producing Merino sheep that are kept extensively in harsh environments, and more intensive lamb production systems that are found in higher rainfall areas or on irrigated land. The proportion of breeding flocks for which objective trait and pedigree measurements are undertaken is relatively much higher in the Australian terminal sire breeds.

Selection takes place within the breeding studs. AI is common in the stud breeding sector, enabling the genetic linkage of flocks. There are breeder groups with organized progeny testing of young sires across flock programmes. In Australia, a national genetic evaluation system known as "Lambplan" has driven genetic evaluation for terminal sires and maternal breeds across flocks for more than a decade. Breeders as well as ram buyers are increasingly basing their ram assessment on estimated breeding value (EBV) or dollar index value. Such a system gives breeders incentives to invest in trait measurement and to create genetic links between their flocks, otherwise it would be difficult for a ram to rise to the top of the across-flock EBV list. Hence, there is increasingly an exchange of genetic material between flocks, mainly through the use of AI. Obviously, such a breeding structure would be more conducive to breeders investing in gene marker technology.

By contrast, the Australian Merino industry has had a much lower proportion of breeders taking up trait and pedigree recording. The industry is more traditional and selection is most often based on visual assessment. While this might be due partly to the sector being more extensive, AI has been commonly used in the Merino stud sector and top Merino rams have always been sold for high prices. Therefore, the extensive nature of the industry does not fully explain the lack of investment in performance recording. The traditional nature of the industry that has hampered the uptake of quantitative genetic principles is also the result of socio-economic factors, with wool producers being traditionally a prominent and relatively wealthy social class. The lamb industry has long been the wool person's "poor brother", but this lack of status has accelerated innovation with the introduction of new approaches such as formal recording and across-flock evaluation. Hence, economic as well as social and cultural reasons may explain why sheep breeding programmes have different levels of sophistication in terms of recording, genetic evaluation and across-flock selection.

BREEDING PROGRAMMES AND TRAITS TARGETED

Meat sheep

Large-scale genetic evaluation programmes for sheep are found in Australia, France, New Zealand, South Africa and the United Kingdom. In all of these, performance recording for meat traits is well advanced, with not only weight traits measured, but also traits related to carcass quality such as body fat and muscle (based on ultrasound

scanning and in some cases computer tomography [CT] scanning), disease (mainly resistance to internal parasites) and reproduction. The national evaluation system in Australia ("Lambplan") now has about 120 000 new animals from about 450 flocks recorded each year for terminal sire breeds and maternal breeds (A. Ball, personal communication). Performance recording takes place only at the stud level, which in a sense is a dispersed nucleus, and a large proportion of the genetic basis of the commercial population stems from these recorded flocks. The proportion of pedigree recorded individuals is high at the stud level, allowing best linear unbiased prediction (BLUP) of EBV. In New Zealand, a similar programme exists ("Sheep Improvement Limited" [SIL]), in which pedigree and performance records are registered with genetic service providers and the information "retailed" back to the breeders. SIL enters more than 250 000 new animals per year from some 750 recorded flocks, all pedigree recorded, and has a database of more than 5 million animal records. Across-flock EBVs are estimated for a proportion of these. In the United Kingdom, about 50 000 breeding ewes and their lamb records are recorded every year from 37 different breeds, and indices have been developed for terminals and maternal ("hill") breeds (Conington *et al.,* 2004). Across-flock genetic evaluation programmes for meat sheep breeds exist also on smaller scales in France, Norway and South Africa.

Most breeding programmes for meat sheep focus on weight traits, and ultrasound scanning is commonly used for fat and muscle traits. Reproduction traits are recorded as numbers of lambs born and weaned. Selection for resistance to internal parasites can be based on faecal worm egg counts (WECs) associated with natural challenge in the field, e.g. in Australia and New Zealand, and this has been shown to be reasonably heritable in Merino sheep (e.g. Khusro *et al.,* 2004). EBVs for WEC are produced for an increasing number of flocks in Australia and New Zealand.

The traits that would most obviously benefit from MAS in meat sheep would be traits related to carcass and carcass quality, reproduction and disease resistance. Ultrasound measurements are currently used to predict carcass fat and muscling. However, genetic correlations with traits measured on carcass are only moderate (Safari, Fogarty and Gilmour, 2005) and specific meat quality attributes such as tenderness and colour might not be well captured by current measurement. Carcass traits are prime targets for MAS as they cannot be measured on breeding animals and progeny or sib testing would be needed as an alternative. Reproduction traits as well as maternal behaviour and ewe survival are also good MAS targets as they are sex limited and are only expressed after the first round of reproduction. Disease resistance traits are generally hard to measure under uniform conditions and would also greatly benefit from MAS.

Wool sheep

Breeding for and recording of wool traits is limited to a few countries. The largest across-flock scheme is found in Australia (mainly for the Merino breed), and smaller genetic evaluation schemes are run in New Zealand, South Africa and South America (Merino and Corriedale). In Australia, the proportion of breeders participating in formal recording and genetic evaluation is smaller for wool than for meat sheep. However, the Merino industry is very large, constituting the vast majority of the Australian flock that consists of about 100 million sheep. By

the end of 2005, a new single system for a national across-flock genetic evaluation of Merinos had been introduced in Australia, combining data from previously separate schemes. The number of animals performance recorded per year is growing rapidly, with about 100 000 new animals now being entered annually.

Wool production efficiency is mainly determined by fleece weight and wool quality. Wool quality traits are mainly fibre diameter and staple strength, and these are economically much more important for fine wools. Staple strength is more expensive to measure, but has a high correlation with the coefficient of variation of fibre diameter, which is therefore a good predictor. Wool traits have generally high levels of heritability, especially fleece weight and fibre diameter.

Reproductive rate in wool sheep has been hard to select for as pedigree recording has been limited and the heritability is low. Moreover, genetic improvement of reproductive rate has been less important for wool production because of the positive net economic benefit of wool producing breeding females. However, with an increasing meat/wool price ratio, the situation is changing and reproductive rate is currently becoming more important. Also, meat attributes of Merino sheep are now receiving increased attention, including measurements of body weight at different ages, fat depth and eye muscle depth (ultrasound scanned).

In pure wool production systems, MAS would be expected to have limited benefit for wool production traits because of their high heritability and the ability to measure the traits before the age of first selection. MAS for reproductive traits and mothering ability would be more beneficial because of low heritability and sex-limited recording. Parasite resistance is becoming a trait of greater economic importance due to the development of resistance to all the major classes of anthelmintics used and the lack of new anthelmintic classes being developed. Host resistance to internal parasites is particularly poor in the Merino breed. The trait can be selected for using field records of WEC. EBVs are being produced for this trait and genetic progress is being achieved. However, the procedure is laborious and there is also some concern about uniformity of measurement and trait definition, as well as the existence of different species of parasites in different regions. Various studies have looked at genotype x environment interactions for parasite resistance and, although some interaction exists, relatively high correlations (~0.8) were found between breeding values in different environments, when environments were defined either through worm type (McEwan *et al.,* 1997) or by high and low flock averages for WEC (Pollot and Greeff, 2004). In any case, many of these trait attributes make parasite resistance a good target for MAS. Identifying QTL for parasite resistance might also shed more light on the biology of immunity, and possibly help to find other modes of improvement.

Feed efficiency, and particularly maternal efficiency, are important determinants of pastoral production systems (Ferrell and Jenkins, 1984) and genetic improvement would benefit from MAS because of the cost of their measurement. However, feed availability and feed costs are quite variable within and between years, and the ability of sheep to cope with harsh environments and periods of drought is perceived by industry as being of greater importance. Hardiness and ewe survival are not well defined characteristics and are not normally measured in breeding programmes.

Ewe fitness and adaptation are often used as the main argument for the existence of genotype x environment interaction in wool production, inhibiting the exchange of genetic material among regions. Carrick (2005) found moderate to high genetic correlations between wool production traits in flock groups differentiated by their phenotypic means for a range of production traits. Discovering QTL for fitness and survival traits in different environments would be useful, but these are unlikely to be found unless the traits themselves are clearly defined and measured.

Dairy sheep

Dairy sheep are predominantly found in the Mediterranean region with both milk and meat production being economically relevant traits to farmers. A great variety of breeds are being targeted in selection programmes for the improvement of milk yield and milk composition but the importance of functional traits such as udder characteristics and mastitis susceptibility is increasing (Barillet, 1997; Barillet, Arranz and Carta, 2005). Genetic improvement for dairy traits, being sex-limited and measured after the first offspring are born, would particularly benefit from MAS.

Goats

Most goat farming systems focus on meat production (about 80 percent), with more emphasis in developed countries on dairy goat production (Olivier *et al.*, 2005) and fibre production (cashmere, mohair). In dairy goat breeding, the most developed breeding programmes are found in France and are based on a strong goat cheese market. Based on AI and milk recording, Caprigene France runs selection schemes for the Saanen and Alpine breeds, with 300 000 goats in 2 500 herds being recorded for milk traits. Dairy goat production is also recorded on smaller scales in Italy, Norway and Spain, with no more than a few thousand animals recorded in other countries (Montaldo and Manfredi, 2002). The main traits in dairy goat production are milk yield and protein and fat content of milk. Being sex-limited and measured only after first production of progeny, these traits would benefit from MAS.

Goat meat production is widely spread throughout the developing world but there are few breeding programmes of any significance. Genetic evaluation for Boer goats and other meat breeds is taking place in Australia and South Africa with weaning weight usually being the main trait measured. Ultrasound measurement of fat and muscle traits is less common in goats, while reproductive traits have had less attention, possibly because of their low heritability and multiparous nature. There are few studies concerning resistance to internal parasites in goats (Olayemi *et al.*, 2002), but these seem to indicate that faecal WECs could be a similar selection trait as in sheep. However, the trait is hard to measure and there is no systematic recording and evaluation in breeding programmes.

DEVELOPMENT OF SHEEP AND GOAT GENOME MAPS

Several key publications have reported progress on the linkage map of the sheep genome based on an international mapping flock developed in New Zealand (Crawford *et al.*, 1995; Maddox *et al.*, 2001). The latest sheep linkage map (version 4.3) comprises 1 256 gene markers mapped to unique locations (Maddox, 2004) and most genomic regions are well covered with a maximum gap of 20 cM. However, there are quite a number of markers of low quality, so a typical genome scan would leave a number

of gaps. Most of the markers are microsatellites. The total number of sheep loci listed in the ARKdb database (http://iowa.thearkdb.org) contains more than 2 000 markers, but many of these are not on the linkage map. The development of the sheep genome map runs somewhat behind developments for other livestock species because of substantially lower investments. Nevertheless, at the DNA level where the sequence can be aligned, there is a ~90 percent homology with the cattle sequence and through gene coding regions ~96 percent, and the sequencing of the cattle genome will therefore greatly enhance the development of the genome map in sheep. There is generally good agreement between sheep and cattle maps, with 598 mainly anonymous common microsatellite loci, i.e. gene markers can be linked to a comparative map. Based on sequence information in other mammals (mainly cattle) and sheep GeneBank sequences, comparative mapping can be used to construct a predicted sheep map. This can be accessed from the Australian Gene Mapping Web site (Maddox, 2005a). The number of single nucleotide polymorphism (SNP) markers in sheep is still very low, but with the cattle sequence known and with an international collaborative sheep bacterial artificial chromosome (BAC)-end sequencing project under way, it is expected that there will be a large number (~16 000) of SNPs available for sheep towards the end of 2006. This will form a set of markers that would allow high-density genome-wide scans.

The goat map is more sparse than the sheep map and contains about half the number of markers known in sheep: 731 loci with 271 genes and 423 microsatellites (http://locus.jouy.inra.fr/). The last pub-lished linkage map for goats contains only 307 markers (Schibler *et al.,* 1998), with coverage of the whole goat genome being far from complete. Although the sparsity of the sheep map makes it difficult to develop a good homology between the maps, about two-thirds of the mapped goat markers can also be linked to the sheep map (Maddox, 2005b).

QTL AND GENE MAPPING

An excellent overview of mapping experiments in sheep can be found on the Australian Gene Mapping Web site (Maddox, 2005a), including references to identified QTL and genes. Successfully identified genes and QTL are related mainly to fecundity, disease resistance and meat quality.

Fecundity

Two genetic mutations have been reported for fecundity: the Booroola mutation: FecB on chromosome 6 (Wilson *et al.,* 2001; Mulsant *et al.,* 2001; Souza *et al.,* 2001) and the Inverdale gene: FecX on the X chromosome (Galloway *et al.*, 2000). The Booroola gene has a substantial additive effect on ovulation rate with each copy increasing this by about 1.5 eggs (i.e. scanned foetuses). The additional allelic effect of the Booroola mutation on litter size is about 0.8 to 0.9 lambs (Davis *et al.,* 1982; Piper and Bindon, 1982; Gootwine *et al.,* 2003) whereas a second copy of the mutation has a slightly smaller effect (0.4–0.6 lambs). The effect on number of lambs weaned is somewhat lower. The effect of the Booroola gene is often perceived as too large and the survival of twin and triplet lambs decreases substantially in extensive and harsh conditions, typical for many sheep flocks. For example, in the Australian Merino industry, the Booroola mutation is not seen as a desirable characteristic. However, the Booroola gene has been introduced in many sheep populations around the world.

The Booroola mutation possibly originates from the Indian Garole (Davis *et al.*, 2002) and, interestingly, the gene effect appeared to be smaller (0.6 lambs born alive) in an Indian introgression programme with Deccani sheep (Nimbkar, Pardeshi and Ghalsasi, 2005). This increase in litter size appears to be easily managed in shepherd flocks. A smaller effect would be more desirable for extensive production systems. It is not clear whether the reduced gene effect arises from a modification due to environmental effects or the genetic background. As the reproductive rate is a trait of high economic value, and due to the availability of a test for the actual gene mutation, Booroola remains a very interesting gene for MAS and marker-assisted introgression (MAI) programmes.

The Inverdale gene has been mapped to the X chromosome and has an effect of about 0.6 lambs per ewe lambing. However, the homozygous ewe is infertile. As carrier rams as well as non-carrier ewes need to be maintained in a crossbreeding system, using this gene in the industry is more complex. However, the 100 percent accurate test has made the use of this gene more manageable.

A number of other major genes for fecundity have been described by Davis (2005), but the molecular basis of these effects has not been formally described.

Disease

Internal parasites are the main cause of economic losses due to health problems in sheep and goat production systems. Although there is significant research under way to detect and map QTL for host resistance to internal parasites, there have not yet been any major breakthroughs in terms of detected polymorphisms in functional genes. Few QTL have been reported for resistance to internal parasites (see review by Dominik, 2005) but not all results are reported in the literature. A major gene effect for resistance to *Haemonchus contortus* was found based on segregation analysis (Meszaros *et al.*, 1999) but this has not been confirmed based on gene markers. The problem of finding distinct QTL for resistance to internal parasites may be due to the complexity of the underlying biological mechanism as well as the difficulty of finding well-defined phenotypes that measure resistance.

Transmissible spongiform encephalopathy (TSE) is a prion disease like scrapie and is characterized by the accumulation of a modified form of a protein known as *PrP*. The *PrP* gene has been associated with variation in scrapie susceptibility in sheep (Moreno *et al.*, 2002), mice (Moreno *et al.*, 2003), and goats (Acin *et al.*, 2003). The gene only explains a proportion of the overall variation for increased resistance to scrapie. Commercial gene tests are available for the *PrP* gene mutation.

A causative mutation has been found for the Spider Lamb Syndrome. This is a relatively rare recessive skeletal disorder with the responsible mutation being assigned to chromosome 6 (Cockett *et al.*, 1999). A commercial test is available for this syndrome.

A gene test based on the *DQA2* gene that resides on the MHC complex (Hickford *et al.*, 2004) and predicts susceptibility to foot rot has been developed at Lincoln University in New Zealand. Thirty-one different alleles have been identified for *DQA2* and a gene marker test rating has been developed based on a susceptibility score of the two alleles of a genotype. There is a clear association between the test rating and the relative risk of contracting foot rot. A gene marker test has been available since

2001 and has been used extensively (over 40 000 tests).

The β-3 adrenergic receptor gene has been sequenced (Forrest and Hickford, 2000) and eight different alleles have been found. This allelic variation is significantly associated with increased risk of cold-related mortality of lambs.

Meat traits

The first causal mutation found for meat traits in sheep is the callipyge gene causing muscular hypertrophy. The gene has been mapped to chromosome 18 and the causative mutation has been identified. However, the trait is expressed in a rather complex manner, termed polar over-dominance; only lambs that inherit the callipyge mutation from their father but not their mother develop the trait. Several interacting genes are involved and the complete molecular basis of callipyge phenotypes has not yet been fully resolved (Freking *et al.,* 2002; Cockett *et al.,* 1996, 2005).

The Carwell gene somewhat resembles the callipyge gene, as it has been mapped to the same genomic region (distal end of chromosome 18) and it also affects muscling (McLaren *et al.,* 2001). However, the overall phenotypic effect is not exactly the same in that the Carwell gene affects only the *longissimus dorsi* and unlike the callipyge gene it has not been associated with a decreased tenderness if the meat is aged appropriately and neither does it seem to be affected by the parent of origin (Jopson *et al.,* 2001). The functional mutation of the Carwell gene, also known as the rib-eye muscling (REM) gene, has not yet been found but close markers in linkage disequilibrium with the putative gene are being developed in Australia, New Zealand and the United Kingdom. A commercial gene test termed "LoinMax" was introduced towards the end of 2005 by Ovita in New Zealand.

A number of gene detection projects have resulted in significant QTL for muscle, fat and other carcass traits, but not all of these have been published, confirmed or fine mapped. A number of studies have reported on QTL for meat traits in sheep (Broad *et al.,* 2000; Walling *et al.,* 2004; Johnson *et al.,* 2005; McRae *et al.,* 2005) and there are probably some unpublished QTL being further developed. Some of these sheep QTL are based on related cattle genes, e.g. the myostatin gene for double muscling (Grobet *et al.,* 1997) and the thyroglobulin gene affecting intramuscular fat (Barendse *et al.,* 2004).

Wool traits

In a recent paper, Purvis and Franklin (2005) reviewed QTL for wool production traits and wool quality. Although wool traits can be measured easily and have high heritability, these authors suggested that research into certain wool production genes was still justified, for example, to break antagonistic correlations (between fleece weight and fibre diameter) or to target specific wool quality traits important for the processing of the product.

A few Mendelian (single locus) characteristics have been described for wool. There is a known mutation of the halo hair gene (HH1) causing extreme hairiness. This has been found in the New Zealand Romney breed and several lines have been developed for the production of "carpet wool" using this specific mutation,. A recessive gene for hairlessness (*hr*) has been described by Finocchiaro *et al.* (2003). Several QTL for wool traits have been published (see Purvis and Franklin, 2005 for an overview), but few of these have been confirmed. On the other hand, it is

probable that a number of QTL identified have not been published. It is likely that some of these wool QTL will be confirmed and available for gene testing over the next few years. Polymorphisms associated with candidate genes for the wool proteins keratin and sulphur have been described (Rogers, Hickford and Bickerstaffe, 1994; McLaren *et al.*, 1997) and seem to be associated significantly with fibre diameter and staple strength.

The genetic regulation of some forms of pigmented wool fibres has often been associated with the Agouti gene (chromosome 13) but this has proven to be a complex pattern of inheritance with several mutations seemingly involved (Smith *et al.*, 2002). More specifically, there appear to be two Agouti loci and at least two different polymorphisms (deletions). Currently, a genetic test for self coloured black wool is not yet available. Other pigmented phenotypes such as badger face and piebald also have a distinct Mendelian inheritance pattern (Sponenberg, 1997) but the molecular basis of these phenotypic variations has not been found.

Dairy traits

Research in dairy sheep has mainly focused on milk protein polymorphisms, in particular αs1-casein and β-lactoglobulin, but results have been inconclusive, unlike those in goats. Together with unfavourable allele frequencies, these results make it unlikely that these polymorphisms will be very useful in a MAS programme. Further QTL mapping work is under way, focusing on production and functional traits (Barillet, Arranz and Carta, 2005).

Other

The Horns gene has been found in sheep as described by Montgomery *et al.* (1996), allowing improved selection efficiency for polled sheep.

Goats

Two goat genes have been well studied. Substantial mapping work has been dedicated towards finding a gene associated with Polled Intersex Syndrome (PIS), and the actual mutation for PIS has been described (Pailhoux *et al.*, 2005). Furthermore, the effects of the αs1-casein gene on milk solids content (protein, fat, casein, casein/protein ratio) have been described in French dairy goat breeds (Barbieri *et al.*, 1995) and the molecular basis has been unravelled (Yahyaoui *et al.*, 2003).

EXAMPLES OF SHEEP AND GOAT MAS BREEDING PROGRAMMES

There is little formal literature about actual applications of MAS in breeding programmes for any livestock species, let alone for sheep and goats. In fact, gene testing and MAS in sheep and goats have only very recently been introduced, and therefore the information compiled in this section is based mainly on information obtained from communication with colleagues in a number of countries (see Acknowledgements).

There are currently two types of MAS programmes. One is the use of gene markers in selection programmes within research projects. Usually the genotyping is subsidized and the purpose of the project is to create additional data for confirmatory studies of the QTL effect, or simply to obtain "proof of concept" where predictions based on simulation and modelling are being verified based on real data. In the other type of application, commercial gene testing is used. This is the scenario required for long-term and sustained use of the technology, but there are few

breeding programmes where commercial applications are viable. The basic condition is that ram breeders and ram buyers are prepared to pay for genetic information arising from genetic testing. This is more likely to happen in places where across-flock genetic evaluations already exist, combined with objective trait measurement and trait valuation in the form of indices. However, not all genetic information can be translated into dollar index terms and genetic testing is often valued beyond the existing index framework.

Experimental sheep MAS

The purpose of "experimental MAS" programmes is to demonstrate that genetic changes can be achieved based on genotype selection and thereby to encourage uptake of MAS by commercial breeders. Usually, the programmes are also designed either to estimate QTL effects more clearly, or to confirm earlier experimental results in industry flocks. Examples of such MAS programmes are:

- selection of sheep against susceptibility to scrapie, being conducted in France and the United Kingdom;
- the MAS Applied to Commercial Sheep (MASACS) Programme in the United Kingdom, coordinated by Oswald Matika from the Roslin Institute. The research team in this programme collaborates with commercial breeders. Three gene marker tests for muscling are being trialled in the first year and it is envisaged that a test for parasite resistance will be introduced in 2006. The three QTL are termed "Texel muscling" (chromosome 18), "Suffolk muscling" (chromosome 1) and "Charollais muscling" (chromosome 1) as described by McRae *et al.* (2005), and the tests will be applied within the respective breeds.

Commercial sheep MAS

Commercial gene testing in sheep is limited mostly to service providers in New Zealand, mainly Ovita and the University of Lincoln, whereas it is absent in goats. Details about gene tests can be found on the Australian Gene Mapping Web site (Maddox, 2005). Gene tests currently available are:

- Foot rot, a gene test commercialized by the University of Lincoln;
- Inverdale gene, through Ovita;
- Booroola gene, through Genomnz;
- Scrapie, (*PrP* gene), available through many companies (see Maddox, 2005);
- Carwell gene, available through Ovita as Loinmax;
- Texel Muscling gene (Chrom 2), available through Ovita as MyoMax.

None of these tests is currently integrated with formal genetic evaluation systems. Rather, gene test results and index values based on polygenic quantitative traits will have to be used separately, and holistic approaches are needed to devise selection rules. The gene tests for reproductive traits are not straightforward to use, while the Inverdale gene is only useful in a heterozygous state and requires specific crossing programmes. The Booroola inheritance model is more straightforward but the effect is too large for most management systems found in Australia.

It should also be noted that most of the commercial gene tests are for traits that are not captured by formal EBVs, and cannot be incorporated easily into existing EBVs, e.g. gene tests for disease traits such as scrapie and foot rot. In principle, the tests for muscle traits could be part of the EBV calculation, but from a ram marketing perspective it might be more useful to exploit the genotype information obtained more explicitly. Furthermore, the proportion of

breeding animals genotyped will be small in relation to the total number of animals evaluated based on phenotype, making an integration of genotypic information with the full evaluation procedure less sensible at this stage. Finally, service providers offering the genotyping results are often different from the service providers of EBV which inhibits a full integration of genetic information to the breeder.

In New Zealand, a significant and rapidly expanding part of the performance recording sheep industry already uses DNA parentage and the above-mentioned tests are now often provided as part of that system. At this point in time, DNA fractional parentage (Dodds, Tate and Sise, 2005) is included within the SIL system, but MAS EBVs for some of the above-mentioned tests (Inverdale, LoinMax) are only carried out on a stand alone basis, in the case of LoinMax since 1997.

Goat MAS

A GAS programme is operational for the alpha-S1 casein gene for dairy goats in France (Manfredi, 2003). The gene is associated with protein content and protein yield. In this programme, young bucks are preselected within families based on genotype. The programme is run by a cooperative AI centre (Capri-IA) and, although started up with government funding, it is now almost running on a fully commercial basis.

MAS in developing countries

Most breeding programmes in developing countries, if existing at all, are small-scale with modest objectives. Usually, the challenge is to foster the flow of information (measurement and evaluation) as well as the flow of genes (dissemination of improved stock). These processes are often inhibited by infrastructural, logistical and socio-economic factors. Clearly, gene marker technology will not be the first priority in many of these programmes. However, where gene tests exist for clearly defined characters with substantial economic benefit, gene markers and MAS could be very beneficial. Introgression of disease resistance genes into productive breeds could be of great value, but few of these examples exist in sheep and goats.

A good example of a clear gene effect successfully implemented in a MAI programme is found in India (Nimbkar *et al.*, 2005). The Booroola gene is being introgressed here from the small Garole breed into the local Deccani breed that is suitable for meat production but has a limited reproductive performance. The Booroola gene has tremendous economic effects in this production system, raising the weaning rate by nearly 50 percent. The breeding programme is undertaken by a research institute, but there are clear strategies and activities to ensure that the improved stock finds its way to shepherd flocks. Evaluation of the results in these shepherd flocks is an explicit part of the project, and initial results look very promising. Therefore, MAS and MAI should not be ruled out for breeding programmes in developing countries, but should be assessed based on the merit of each case. However, implementation of gene marker technology will only work within the framework of a sound existing breeding programme, ensuring the prerequisite that genetic information is valued and that the gene marker accounts for substantial economic merit.

CONCLUSION

Sheep and goat breeding programmes exist in low- to medium-input agricultural systems where there are many independent breeding units and where trait recording

and genetic evaluation are provided by external service agents. This situation is different from that in poultry and pigs and to some extent dairy, and more similar to that in beef cattle, in the sense that the business units that invest in genetic information are not the same as those providing genetic evaluation, and EBVs are available in the public domain. Also, genotypic information is an explicit part of the marketing of genetic material. The result is that genotypic information is more likely to be used outside the usual EBV system, with the chance of being overvalued once the investment is made. There is a place for MAS and MAI based on genetic tests for clearly demonstrated phenotypic effects with economic benefit, for example for disease, fecundity and meat quality.

The number of detected and confirmed QTL is low for sheep and goats and gene mapping is less advanced than in other livestock species. There is significant investment and progress being made in marker development and gene discovery, but it will take some years before large amounts of genetic information become available at little cost, e.g. in the form of SNP chips. Until then, genotypic information will provide additional selection criteria, making optimal selection a greater challenge.

Ultimately, the additional value of gene markers will be greatest in breeding programmes that already use intensive pedigree and performance recording, and it will help to shift selection pressure towards traits that are hard to improve based on phenotypic (BLUP) selection (i.e. traits such as fertility, disease resistance and carcass quality). It is not essential that genetic tests are based on functional mutations, as gene markers can have predictive value due to being in linkage disequilibrium with functional genes. In breeding programmes without extensive recording, it is more important to rely on direct markers, but this will only be valuable in practice if genes have very large economic effects. The same holds for genetic tests for distinct Mendelian traits, but the overall value of these traits in breeding programmes is limited. In less-developed breeding programmes, investments in pedigree and performance recording will most likely be more profitable than investments in gene technology.

Application of MAS or MAI in many sheep and goat breeding programmes in developing countries is not a priority, but opportunities exist, conditional on having a clearly visible phenotypic effect and a programme based on well-defined objectives and performance based selection.

ACKNOWLEDGEMENTS

The author would like to thank the following colleagues for providing information about breeding and MAS programmes: Richard Apps, Alex Ball, Stephen Bishop, Didier Boichard, Joanne Conington, Schalk Cloete, Oswald Matika, Eduardo Manfredi and John McEwan. Information from Jill Maddox, her Web site and her comments on the marker maps were very helpful, and Robert Banks made useful comments and improvements to the manuscript.

REFERENCES

Acin, C., Martin-Burriel, I., Monleon, E., Rodellar, C., Badiola, J. & Zaragoza, P. 2003. Characterization of the caprine *PrP*-gene study of new polymorphisms and relationship with the resistance/susceptibility to the scrapie disease. *Proc. Internat. Workshop on Major Genes in sheep and Goats.* 8–11 December 2003. CD-ROM Communication No. 2-31. Toulouse, France.

Barbieri, M.E., Manfredi, E., Elsen, J.M., Ricordeau, G., Bouillon, J., Grosclaude, F., Mahe, M.F. & Bibe, B. 1995. Effects of the alpha-s1-casein locus on dairy performances and genetic parameters of Alpine goats. *Genet. Sel. Evol.* 27: 437–450.

Barendse, W., Bunch, R., Thomas, M., Armitage, S., Baud, S. & Donaldson, N. 2004. The TG5 thyroglobulin gene test for a marbling quantitative trait loci evaluated in feedlot cattle. *Austr. J. Exper. Agric*. 44: 669 – 674.

Barillet, F. 1997. Genetics of milk production. *In* L. Piper & A. Ruvinsky, eds. *The genetics of sheep*, pp 539–564. Wallingford, UK, CABI.

Barillet, F., Arranz, J.J. & Carta, A. 2005. Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. *Genet. Sel. Evol*. 37 (Suppl. 1): S109–S123.

Broad, T.E., Glass, B.C., Greer, G.J., Robertson, T.M., Bain, W.E., Lord, E.A. & McEwan, J.C. 2000. Search for a locus near to myostatin that increases muscling in Texel sheep in New Zealand. *Proc. New Zealand Soc. Anim. Prod.* 60: 110–112.

Carrick, M. 2005. Genotype by environment interaction in wool sheep. Report SHGEN.036. North Sydney, NSW, Australia, Meat & Livestock Australia.

Cockett, N.E., Jackson, S.P., Shay, T.L., Farnir, F., Berghmans, S., Snowder, G.D., Nielsen, D.M. & Georges, M. 1996. Polar overdominance at the ovine callipyge locus. *Science* 273: 236–238.

Cockett, N.E, Shay, T.L., Beever, J.E., Nielsen, D., Albretsen, J., Georges, M., Peterson, K., Stephens, A., Vernon, W., Timofeevskaia, O., South, S., Mork, J., Maciulis, A. & Bunch, T.D. 1999. Localization of the locus causing Spider Lamb Syndrome to the distal end of ovine chromosome 6. *Mamm. Genome* 10: 35–38.

Cockett, N.E., Smith, M.A., Bidwell, C.A., Segers, K., Hadfield, T.L., Snowder, G., Georges, M. & Charlier, C. 2005. The callipyge mutation and other genes that affect muscle hypertrophy in sheep. *Genet. Sel. Evol*. 37 (Suppl. 1): S65–S81.

Conington, J., Bishop, S.C., Waterhouse, A. & Simm, G. 2004. A bioeconomic approach to derive economic values for pasture-based sheep genetic improvement programmes. *J. Anim. Sci*. 82: 1290–1304.

Crawford, A.M., Dodds, K.G., Pierson, C.A., Ede, A.J., Montgomery, G.W., Garmonsway, H.G., Beattie, A.E., Davies, K., Maddox, J.F., Kappes, S.W., Stone, R.T., Nguyen, T.C., Penty, J.M., Lord, E.A., Broom, J.E., Buitkamp, J., Schwaiger, W., Epplen, J.T. Matthew, P., Matthews, M.E., Hulme, D.J., Beh, K.J., McGraw, R.A. & Beattie, C.W. 1995. An autosomal genetic linkage map of the sheep genome. *Genetics* 140: 703–724.

Davis, G.H. 2005. Major genes affecting ovulation rate in sheep. *Genet. Sel. Evol.* 37 (Suppl. 1): S11–S23.

Davis, G.H., Montgomery, G.W., Allison, A.J., Kelly R.W. & Bray, A.R. 1982. Segregation of a major gene influencing fecundity in progeny of Booroola sheep. *New Zealand J. Agric. Res*. 25: 525–529.

Davis, G.H., Galloway, S.M., Ross, I.K., Gregan, S.M., Ward, J., Nimbkar, B.V., Ghalsasi, P.M., Nimbkar, C., Gray, G.D., Subandriyo, I.I., Tiesnamurti, B., Martyniuk, E., Eythorsdottir, E., Mulsant, P., Lecerf, F., Hanrahan, J.P., Bradford, G.E. & Wilson, T. 2002. DNA tests in prolific sheep from eight countries provide new evidence on origin of the Booroola (FecB) mutation. *Biol. Reprod*. 66: 1869–1874.

Dodds, K.G., Tate, M.L. & Sise, J.A. 2005. Genetic evaluation using parentage information from genetic markers. *J. Anim Sci*. 83: 2271–2279.

Dominik, S. 2005. Quantitative trait loci for internal nematode resistance in sheep: a review. *Genet. Sel. Evol*. 37 (Suppl. 1): S83–S96.

Ferrell, C.L. & Jenkins, T.G. 1984. Energy utilisation by mature, non-pregnant, non-lactating cows of different types. *J. Anim. Sci*.58: 234–243.

Finocchiaro, R., Portolano, B., Damiani, G., Caroli, A., Budelli, E., Bolla, P. & Pagnacco, G. 2003. The hairless (hr) gene is involved in the congenital hypotrichosis of Valle del Belice sheep. *Genet. Sel. Evol.* 35 (Suppl 1): S147–S156.

Forrest, R.H. & Hickford, J.G.H. 2000. Rapid communication: nucleotide sequences of the bovine, caprine and ovine β3-adrenergic receptor genes. *J. Anim. Sci*. 78: 1397–1398.

Freking, B.A., Murphy, S.K., Wylie, A.A., Rhodes, S.J., Keele, J.W., Leymaster, K.A., Jirtle, R.L. & Smith, T.P. 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* 12: 1496–1506.

Galloway, S.M., McNatty, K.P., Cambridge, L.M., Laitinen, M.P.E., Juengel, J.L., Jokiranta, T.S., McLaren, R.J., Luiro, K., Dodds, K.G., Montgomery, G.W., Beattie, A.E., Davis, G.H. & Ritvos, O. 2000. Mutations in an oocyte derived growth factor gene (BMP15) causes increased ovulation rate and infertility in a dosage sensitive manner. *Nature Genet*. 25: 279–283.

Gootwine, E., Rozov, A., Bor, A. & Richer, S. 2003. Effect of the FecB (Booroola) gene on reproductive and productive traits in the Assaf breed. *Proc. Internat. Workshop on Major Genes and QTL in Sheep and Goats*. 8–11 December 2003. CD-ROM Communication No. 2-12. Toulouse, France.

Grobet, L., Royo Martin, L.J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Menissier, F., Massabanda, J., Fries, R., Hanset, R. & Georges, M. 1997. A deletion in the myostatin gene causes double-muscling in cattle. *Nature Genet.* 17: 71–74.

Hickford, J.G.H., Zhou, H., Slow, S. & Fang, Q. 2004. Diversity of the ovine DQA2 gene. 2004 *J. Anim. Sci.* 82: 1553–1563.

Hoste, C.H. 2002. Research and development challenges for animal breeding in developing countries. *Proc. 7th Wrld. Congr. Genetics Appl. Livest. Prodn*. 19–23 August 2002. CD-ROM Communication No. PS-02. Montpellier, France.

Johnson, P.L., McEwan, J.C., Dodds, K.G., Purchas, R.W. & Blair, H.T. 2005. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. *J. Anim. Sci.* 83: 1988–2000.

Jopson, N.B., Nicoll, G.B., Stevenson-Barry, J.M., Duncan, S., Greer, G.J., Bain, W.E., Gerard, E.M., Glass, B.C., Broad, T.E. & McEwan, J.C. 2001. Mode of inheritance and effects on meat quality of the rib-eye muscling (REM) QTL in sheep. *Proc. Assoc. Adv. Anim. Breed. Genet*. 14: 111–114.

Khusro, M., van der Werf, J.H.J., Brown, D. & Ball, A. 2004. Across flock (co)variance components for faecal worm egg count, live weight, and fleece traits for Australian merinos. *Livest. Prod. Sci.* 91: 35–43.

Maddox, J.F. 2004. Enhancing the sheep genome map. *Proc. 29th Int. Conf. Anim. Genet.* p. 86. 11–16 September 2004. Tokyo, Japan, International Society for Animal Genetics and Japanese Society of Animal Breeding and Genetics (available at www.isag.org.uk/ISAG/all/ISAG_Tokyo_2004_Proceedings.pdf).

Maddox, J.F. 2005a. Australian Gene Mapping Web site (http://rubens.its.unimelb.edu.au/~jillm/jill.htm).

Maddox, J.F. 2005b. A presentation of the differences between the sheep and goat maps. *Genet. Sel. Evol*. 37 (Suppl. 1): S1–S10.

Maddox, J.F., Davies, K.P., Crawford, A.M., Hulme, D.J., Vaiman, D., Cribiu, E.P., Freking, B.A., Beh, K.J., Cockett, N.E., Kang, N., Riffkin, C.D., Drinkwater, R., Moore, S.S., Dodds, K.G., Lumsden, J.M., van Stijn, T.C., Phua, S.H., Adelson, D.L., Burkin, H.R., Broom, J.E., Buitkamp, J., Cambridge, L., Cushwa, W.T., Gerard, E., Galloway, S.M., Harrison, B., Hawken, R.J., Hiendleder, S., Henry, H.M., Medrano, J.F., Paterson, K.A., Phua, S.H., Schibler, L., Stone, R.T. & van Hest, B. 2001. An enhanced linkage map of the sheep genome comprising more than 1 000 loci. *Genome Res.*11: 1 275–1 289.

Manfredi, E. 2003. The tale of goat αs1-casein. *Proc. 3rd Internat. Workshop on Major Genes and QTL in Sheep and Goats.*CD-ROM Communication No. 2-31. Toulouse, France.

McEwan, J.C., Dodds, K.G., Greer, G.J., Bain, W.E., Wright, C.S., Green, R.S. & Watson, T.G. 1997. Genotype rankings for host resistance to internal parasites for sheep grazed in contrasting environments. *Proc. Assoc. Adv. Anim. Breed. Genet*. 12: 40–43.

McLaren, R.J., Broad, T.E., McEwan, J.C., Jopson, N.B., Robertson, T.R., Glass, B.C., Gerard, E.M., Greer, G.J., Bain W.E. & Nicoll, G.B. 2001. Identification of positional candidates for the Carwell locus for rib-eye muscling in sheep. *Proc. Plant and Animal Genome IX,* W46. 13–17 January 2001. San Diego, CA, USA (available at www.intl-pag.org/9/abstracts/W17_03.html).

McLaren, R., Rogers, G., Davies, K., Maddox, J. & Montgomery, G. 1997. Linkage mapping of wool keratin and keratin associated protein genes in sheep. *Mamm. Genome* 8: 938–940.

McRae, A.F., Bishop, S.C., Walling, G.A., Wilson, A.D. & Visscher, P.M. 2005. Mapping of multiple quantitative trait loci for growth and carcass traits in a complex commercial sheep pedigree. *Anim. Sci.* 80: 135–141.

Meszaros, S.A., Henshall, J.M., Burgess, S.K., Gray, G.D. & Tier, B. 1999. Detection of a quantitative trait locus associated with a reduction of faecal egg count in Merino sheep. *Proc. Assoc. Adv. Anim. Breed. Genet*. 13: 211–218.

Montaldo, H.H. & Manfredi, E. 2002. Organization of selection programs for dairy goats. *Proc. 7th Wrld. Cong. Genet. Appl. Livest. Prodn.* Communication No. 01-35. Montpellier, France.

Montgomery, G.W., Henry, H.M., Dodds, K.G., Beattie, A.E., Wuliji, T. & Crawford, A.M. 1996. Mapping the Horns (Ho) locus in sheep: a further locus controlling horn development in domestic animals. *J. Hered*. 87: 358–63.

Moreno, C.R., Lantier, I., Lantier, F., Andreoletti, O., Vaiman, D., Sarradin, P., Echeynne, F., Cribiu, E.P., Cosseddu, G. & Elsen, J-M. 2002. *Proc. 7th Wrld. Cong. Genet. Appl. Livest. Prodn.* CD-ROM Communication No. 13-21. Montpellier, France.

Moreno, C.R., Lantier, F., Lantier, I., Sarradin, P. & Elsen, J-M. 2003. Detection of new quantitative trait loci for susceptibility to transmissible spongiform encephalopathies in mice. *Genetics* 165: 2085–2091.

Mulsant, P., Lecerf,, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., Pisselet, C., Riquet, J., Monniaux, D., Callebaut, I., Cribiu, E., Thimonier, J., Teyssier, J., Bodin, L., Cognie, Y., Chitour, N. & Elsen, J.M. 2001. Mutation in bone morphogenetic protein receptor-1B is associated with increased ovulation rate in Booroola Merino ewes. *Proc. Nat. Acad. Sci. USA* 98: 5104–5109.

Nimbkar, C., Pardeshi, V. & Ghalsasi, P. 2005. Evaluation of the utility of the FecB gene to improve the productivity of Deccani sheep in Maharashtra, India. *In* H.P.S, Makkar & G.J. Viljoen, eds.

Applications of gene-based technologies for improving animal production and health in developing countries, pp. 145–154. IAEA/FAO Internat. Symp., Netherlands, Springer.

Olayemi, M.E., Walkden-Brown, S.W., van der Werf, J.H.J. & Le Jambre, L.F. 2002. Genetic variability of resistance indicators for gastrointestinal nematode infection in Angora and Cashmere goats. *Proc. 7th Wrld. Cong. Genet. Appl. Livest. Prodn.* CD-ROM Communication No. 13-30, Montpellier, France.

Olivier, J.J., Cloete, S.W.P., Schoeman S.J. & Muller, C.J.C. 2005. Performance testing and recording in meat and dairy goats. *Small Rumin. Res.* 60: 83–93.

Pailhoux, E., Vigier, B., Schibler, L., Cribiu, E.P., Cotinot, C. & Vaiman, D. 2005. Positional cloning of the PIS mutation in goats and its impact on understanding mammalian sex-differentiation. *Genet. Sel. Evol*. 37 (Suppl 1): S55–S64.

Piper, L.R. & Bindon, B.M. 1982. Genetic segregation for fecundity in Booroola Merino sheep. In *Proc. Wrld. Cong. Sheep and Beef Cattle Breeding*, Vol. 1, pp. 395–400. Palmerston North, New Zealand, Dunmore Press.

Pollot, G.E. & Greeff, J.C. 2004. Genotype x environment interactions and genetic parameters for fecal egg count and production traits of Merino sheep. *J. Anim. Sci.* 82: 2840–2851.

Purvis, I.W. & Franklin, I.R. 2005. Major genes and QTL influencing wool production and quality: a review. *Genet. Sel. & Evol.* 37 (Suppl. 1): S97–S107.

Rogers, G.R., Hickford, J.G.H. & Bickerstaffe, R. 1994. Polymorphisms in two genes for B2 sulfur proteins of wool. *Anim. Genet.* 25: 407–415.

Safari, E., Fogarty, N.M. & Gilmour, A.R. 2005. A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. *Livest. Prodn. Sci.* 92: 271–289.

Schibler, L., Vaiman, D., Oustry, A., Giraud-Delville, C. & Cribiu, E.P. 1998. Comparative gene mapping a fine scale survey of chromosome rearrangements between ruminants and humans. *Genome Res.* 8: 901–915.

Smith, M.A., Shay, T.L., Beever, J.E., Notter, D.R. & Cockett, N.E. 2002. Identification of an Agouti-like locus in sheep. *Anim. Genet*. 33: 383–385.

Souza, C., MacDougall, J., Campbell, C., McNeilly, B.K. & Baird, D.T. 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1B (BMPR1B) gene. *J. Endocrin*. 169: R1–R6.

Sponenberg, D.P. 1997. Genetics of colour and hair texture. *In* L.R. Piper & A.Ruvinsky, eds. *The genetics of sheep,* pp. 51–86. Wallingford, UK, CABI.

Walling, G.A., Visscher, P.M., Wilson, A.D., McTeir, B.L., Simmand, G. & Bishop, S.C. 2004. Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. *J. Anim. Sci.* 82: 2234–2245.

Wilson, T., Wu, X., Juengel, J., Ross, I., Lumsden, J., Lord, E., Dodds, K., Walling, G., McEwan, J., O'Connell, A., McNatty, K. & Montgomery, G. 2001. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biology of Reprod*. 64: 1225–1235.

Yahyaoui, M.H., Angiolillo, A., Pilla, F., Sanchez, A., & Folch, J.M. 2003. Characterization and genotyping of the caprine kappa-casein variants. *J. Dairy Sci*. 86: 2715–2720.

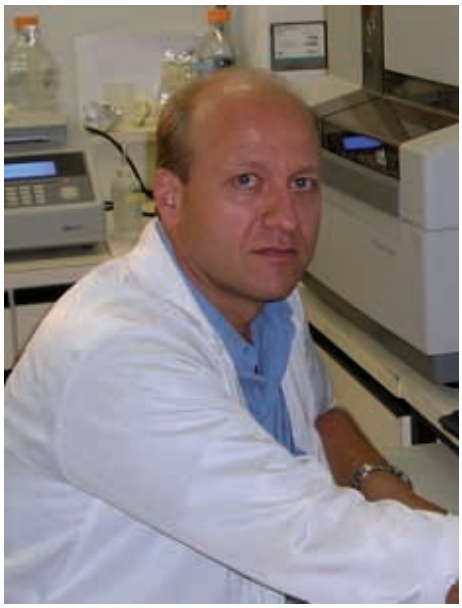
SECTION IV

Marker-assisted selection in forestry – case studies

CHAPTER 14

Marker-assisted selection in *Eucalyptus*

Dario Grattapaglia



SUMMARY

Planted *Eucalyptus* occupies globally more than 18 million hectares and has become the most widely planted hardwood tree in the world, supplying high-quality woody biomass for several industrial applications. In this chapter an overview is presented on the status and perspectives of marker-assisted selection (MAS) in species of *Eucalyptus*. After an introduction to the main features of modern eucalypt breeding and clonal forestry, some applications of molecular markers in support to operational breeding are presented. By reviewing the status of quantitative trait locus (QTL) mapping in *Eucalyptus*, the challenges and some realistic prospects for the application of MAS to improve relevant traits are outlined. With the expected availability of more powerful genomic tools, including a draft of the *Eucalyptus* genome, the main challenges in implementing MAS will be in phenotyping trees accurately, analysing the overwhelming amount of genomic data available and translating this into truly useful molecular tools for breeding.

INTRODUCTION

Planted *Eucalyptus* forests occupy globally more than 18 million hectares and have become the most widely planted hardwood forest tree in the world (FAO, 2001). *Eucalyptus* tree species used in production forestry are long-lived, evergreen species belonging to the angiosperm family Myrtaceae (Ladiges, Udovicic and Nelson, 2003). They are native to Australia and adjacent islands where they occur naturally from sea level to the alpine tree line, from high rainfall to semi-arid zones, and from the tropics to latitudes as high as 43° south (Eldridge *et al*., 1993; Ladiges, Udovicic and Nelson, 2003). Fast growth rates and a wide range of adaptability have contributed to the great interest that *Eucalyptus* species receive in many countries outside their native range. Besides the fast growth that allows for shorter rotations, many species display wood properties that make them very suitable for fuel and charcoal production, pulp and paper manufacture as well as sawn wood. While *E. globulus* is the premier species for temperate zone plantations in Australia, Chile, Portugal and Spain, elite hybrid clones involving *E. grandis* and *E. urophylla* are used extensively by the pulp and paper industry in tropical regions or Brazil, China, the Democratic Republic of the Congo and South Africa because of their wood quality, rapid growth, canker disease resistance and high volumetric yield.

Planted *Eucalyptus* stands supply in a rational and efficient way, high-quality woody raw material that would otherwise come from native tropical forests. In the decades to come, the expansion of these "fibre farms" will likely be limited by the growth of crop plantations and by public opinion pressure. Increased productivity of forests and refinements in the quality of wood products by selective breeding will become of increasing strategic importance to the forest industry. Molecular tools based on the direct identification of useful variation at the DNA level are expected to provide new opportunities for the genetic manipulation of growth, form and especially wood properties of planted trees by marker-assisted selection (MAS) approaches.

Almost fifteen years have passed since the first experiments in molecular breeding of forest trees. The development of linkage maps and quantitative trait loci (QTL) information in trees was greatly accelerated by the advent of more accessible DNA marker techniques, new concepts in linkage mapping and novel strategies for advanced generation tree breeding. From the outset, many expectations were generated for fast and accurate methods for early marker-based selection in trees. Significant progress has been made and the knowledge gathered led to some short-term opportunities for the incorporation of genomic analysis in tree genetics and breeding. However, it also became clear that several challenges remained before more refined and higher impact applications could be implemented.

In this chapter, an overview is presented on the status of MAS in species of *Eucalyptus*. The term MAS is used in *latu sensu*, i.e. encompassing the several molecular techniques and approaches that offer potential to contribute to eucalypt breeding. Some recent reviews have detailed several aspects of *Eucalyptus* genome research including gene discovery, candidate gene mapping, functional genomics and physical mapping (Moran *et al*., 2002; Grattapaglia, 2004; Poke *et al*., 2005; Shepherd and Jones, 2005; Myburg *et al*., 2006). The focus of this chapter is a more applied one, attempting to link the realities of current

eucalypt breeding practice and the molecular tools available or in development. To set the stage for a realistic appraisal of MAS for *Eucalyptus*, a brief introduction is presented of the main features of modern eucalypt breeding and clonal forestry in order to provide a better understanding of the challenges and opportunities that lie ahead for cost-efficient molecular breeding. Following this section, some current low technological input applications of molecular markers in support of operational breeding are presented, such as the quantification of genetic diversity and relationships, the analysis of mating patterns and paternity in seed orchards and fingerprinting for quality assurance and quality control of clonal propagation. Within the framework of MAS for trait advancement, after reviewing the status of QTL mapping in *Eucalyptus*, the challenges and some realistic prospects for the application of MAS to improve relevant traits are outlined. Finally, with the expected availability of a draft of the whole *Eucalyptus* genome within the next years, a succinct summary is presented on the prospects of advancing genomic approaches for gene identification and subsequent application of MAS.

EUCALYPTUS BREEDING AND PLANTATION FORESTRY

Eucalyptus domestication

Eucalypts spread rapidly around the world following their discovery by Europeans in the late eighteenth century (Eldridge *et al.,* 1993). They were introduced into countries such as Brazil, Chile, France, India, Portugal and South Africa in the first quarter of the 1800s (Doughty, 2000) and rapidly adopted in forest plantations as their fast growth and good adaptability became known. During the nineteenth and twentieth centuries, large quantities of seeds were collected and distributed directly from Australia through a number of seed collection expeditions carried out both by government organizations and private forestry companies throughout the world.

Eucalyptus species have a mixed mating system, but are predominantly outcrossers and animal pollinated. High levels of outcrossing are maintained by protandry and various incomplete pre- and post-zygotic barriers to self-fertilization including strong selection against the products of inbreeding (Pryor, 1976). Although the major eucalypt subgenera do not hybridize in nature, hybridization among species within the same subgenus has been detected, often making separation of species difficult (Pryor and Johnson, 1971). Hybridization becomes more frequent in exotic conditions outside the natural species range. In fact, this property has been widely exploited by eucalypt breeders who take advantage of the naturally occurring genetic variation for growth and wood properties among species (de Assis, 2000). Several artificial hybrid combinations have been produced, although hybrid inviability tends to increase with increasing taxonomic distance between the parents (Griffin, Burgess and Wolf, 1988; Potts and Dungey, 2004).

In several countries the continued plantation from local seed sources gave rise to landraces adapted to the specific environment of the country (Eldridge *et al.*, 1993). Seed collections from such local exotic plantings of multiple species became common and where plantings occurred, F_1 hybrids were derived (Potts and Dungey, 2004). While several of these F_1 hybrids performed well, especially when deployed as clones, seed collection from hybrid stands often resulted in plantations that performed poorly in subsequent generations and were

extremely variable. A textbook case is the Rio Claro hybrid swarm in Brazil (Campinhos and Ikemori, 1977; Brune and Zobel, 1981), a eucalypt arboretum where Navarro de Andrade, the "father of eucalypts" in Brazil first introduced and planted a collection of 144 different *Eucalyptus* species between 1904 and 1909. Several of these species hybridized once the natural barriers to introgression were removed in the exotic habitat so that seeds collected from these stands were largely interspecific hybrids. Large commercial plantations were established in Brazil with seeds from this arboretum following fiscal incentives for reforestation granted by the government starting in 1966. Although some of the resulting forests were on average economically inferior, in these very variable stands some outstanding trees for growth, form and disease resistance derived from chance events of recombination were found. The advent of operational cloning techniques at the beginning of the 1980s allowed capturing the superiority of such hybrids that are still used today in some of the most productive eucalypt clonal plantations in the world.

The history of eucalypt breeding, which is short when compared with crop species, was detailed by Eldridge *et al.* (1993) and more recently reviewed and updated by Potts (2004). Some of the earliest breeding was undertaken by French foresters in Morocco in 1954–55 (Eldridge *et al*., 1993). The advent of industrially-oriented eucalypt stands in the 1960s led to a more formal approach to breeding with, for example, the establishment of the Florida *E. grandis* breeding programme in 1961 (Franklin, 1986), *E. globulus* breeding in Portugal in 1965–66 (Potts *et al*., 2004) and large provenance tests of *E. camaldulensis* in many countries (Eldridge *et al.,* 1993). However, a major breakthrough in eucalypt plantation technology occurred in the 1970s with the establishment of the first commercial stands of selected clones derived from hardwood cuttings in the Democratic Republic of the Congo (Martin and Quillet, 1974) followed by Aracruz in Brazil (Campinhos and Ikemori, 1977). At the same time, in many tropical countries such as Brazil and South Africa, efforts were intensified to establish extensive provenance/progeny trials of species such as *E. urophylla*, *E. grandis* and some others that belonged to the same subgenus *Symphyomyrtus* (Eldridge *et al*., 1993). These trials were established from open pollinated seed lots collected from selected trees in the wild and constituted the base populations for subsequent selective breeding in many countries. This initial effort, which was carried out typically by government forestry research institutions, was followed during the 1980s by more intensive collections by private organizations, targeting elite provenances identified in earlier collections as being more adapted for species such as *E. grandis, E. tereticornis* and *E. viminalis* (Eldridge *et al*., 1993).

Eucalyptus breeding and plantation forestry

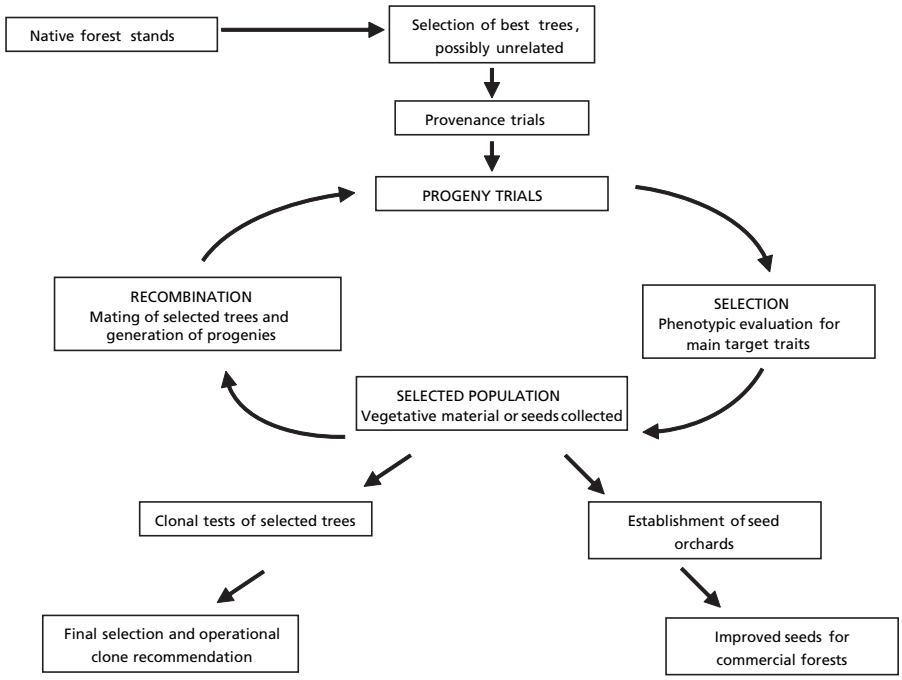
Eucalyptus plantation forestry species are well known for their fast growth, straight form, valuable wood properties, wide adaptability to soils and climates, and ease of management through coppicing (Eldridge *et al*., 1993; Potts, 2004). They are now planted in more than 90 countries where the various species are grown for products as diverse as sawn timber, poles, firewood, pulp, charcoal, essential oils, honey and tannin as well as for shade and shelter (Doughty, 2000). They are an important source of fuel and building material in rural communities in countries such

as China, Ethiopia, India, Peru and Viet Nam. However, it is the increasing global demand for short fibre pulp that has driven the massive expansion of eucalypt plantations and accompanying breeding practices throughout the world during the twentieth century (Turnbull, 1999). Their high fibre content relative to other wood components, coupled with the uniformity of fibres relative to other angiosperm species, has led to high demand for eucalypt pulp for coated and uncoated free-sheet paper, bleach board, sanitary products (fluff pulp), and to a lesser extent for top liners on cardboard boxes, corrugating medium, and as a filler in long fibre conifer products such as newsprint and containerboard (Kellison, 2001). In the last ten years, the development of new wood drying and sawing technologies has also increased interest in using plantation eucalypts for sawn wood, veneer, medium density fibreboard and as extenders in plastic and moulded timber (Kellison, 2001).

An FAO report estimated a total of 17.9 million ha of planted *Eucalyptus* worldwide with India as the largest planter with over 8 million ha followed by Brazil with 3 million (FAO, 2001). The majority of plantations consists of only a few eucalypt species and hybrids. The most important are *E. grandis, E. globulus, E. urophylla* and *E. camaldulensis*, which together with their hybrids account for about 80 percent of the plantation area, followed by *E. nitens, E. saligna, E. deglupta, E. pilularis, Corymbia citriodora* and *E. teriticornis* (Eldridge *et al.,* 1993; Waugh, 2004). Market favourites for pulpwood are *E. grandis, E. urophylla* and their hybrids in tropical and subtropical regions and *E. globulus* in temperate regions.

Although eucalypt breeding is currently a very dynamic and technically advanced operation carried out mainly by several private companies, eucalypts should be seen as still in their domestication infancy when compared with crop species, with most breeding programmes only one or two generations removed from the wild. However, with the combination of ample genetic variation both at the intra and interspecific levels and the ability to clone elite genotypes, eucalypts have quickly become among the most advanced genetic material in forestry. Breeding of eucalypts has moved faster in countries such as Brazil, Chile, Portugal and South Africa that adopted *Eucalyptus* for industrial plantation forestry. Most eucalypt breeding programmes worldwide are focused on genetically improving trees for industrial pulpwood production (Borralho, 2001; Kanowski and Borralho, 2004). The target traits of most breeding programmes include volumetric growth per hectare, wood density and pulp yield (Borralho, Cotterill and Kanowski, 1993). Traits such as pest and disease resistance and adaptability to abiotic stresses such as frost, drought or wind are usually secondary targets that become important when they have an impact on one or more of the main traits. Following the standard concepts in tree breeding, large genetic gains have been obtained in the early stages of eucalypt domestication, simply through species and provenance selection followed by individual selection and establishment of clonal or seedling seed orchards or clonal propagation of elite selections for direct deployment (Eldridge *et al.,* 1993; Kanowski and Borralho, 2004; Potts, 2004). Subsequent population improvement has also demonstrated significant genetic gain through recurrent selection in an open-pollinated breeding population coupled with open or controlled pollinated populations of the most elite selections or specialized

FIGURE 1
Basic breeding scheme used in *Eucalyptus* breeding involving recurrent selection cycles from which elite parent trees are derived for the establishment of seed orchards, as well as individual trees to be tested and later recommended as clones



breeds (Potts, 2004). For species that are easily propagated vegetatively, such as *E. grandis*, *E. urophylla* and several of their hybrids, clonally propagated breeding populations have enhanced gains by allowing the capture of additive and non-additive genetic effects (Figure 1).

Clonal forestry of *Eucalyptus*

After more than 25 years following the introduction of clonal forestry of *Eucalyptus* (Campinhos, 1980; Brandão, Campinhos and Ikemori, 1984), this forest production system is now perfectly integrated into the strategies and plans of advanced generation breeding programmes. Clonal propagation and hybrid breeding have constituted an extremely powerful combination of tools for the rapid improvement of the quality of wood and wood products. While the first hybrid clones were selected based on large-scale screening of high-yielding spontaneous hybrids resistant to diseases (such as the eucalypt canker), today clones are being derived increasingly from deliberate interspecific hybrid production strategies (Figure 2). Eucalypt hybrids, involving two or more species deployed as clones, currently make up a significant proportion of eucalypt plantation forestry, particularly in the tropics and subtropics. In a recent survey of clonal forestry in Brazil, for example, considering all the large and medium-sized companies, the area planted with clones corresponded

FIGURE 2
One-stop pollination of *Eucalyptus*

(A) Elite parent trees, kept as grafts in indoor insect-proof orchards, are induced to flower with growth regulators in approximately 12 to 15 months. (B) Flowers to be used are still closed; open protandric flowers are discarded. (C) Flowers are cut open before anthesis with a nail cutter. (D) Pollen from the other parent is deposited directly at the base of the style and no bag protection is needed as the greenhouse is kept free of insects. (Photographs courtesy of Teotônio F. de Assis)

to more than 1 008 000 ha, involving 362 different clones at a rate of 2 to 40 clones per company, and a range of 10 to 34 000 ha per clone (mean 4 150 ha). The annual introduction of new clonal plantations to support expansion of forest-based industrial production is in the order of 238 000 ha, with a mean of 1 820 ha per clone (de Assis, Rezende and Aguiar, 2005).

An important paradigm shift in eucalypt breeding for pulp and paper began in the 1990s with the increasing realization

that the actual "pulp factory" is the tree. Particularly in vertically integrated pulp production systems, as highly productive clonal forests with over 40 m^3/ha/yr became the standard (Binkley and Stape, 2004), the focus shifted quickly from volume growth to wood quality with the objective of improving pulp yield per hectare by reducing wood specific consumption (WSC), i.e. the amount of wood in cubic metres necessary to produce one tonne of pulp. Trees that yield more cellulose generate savings all the way from tree harvesting, transportation, chipping and pulping while mitigating the need for an accelerated expansion of the forest land base.

Clonal forestry of *E. grandis* x *E. urophylla* selected clones in the 1980s was able to reduce WSC from 4.9 to 4.0 m^3/tonne of pulp (Ikemori, Penchel and Bertolucci, 1994). However, it is now well known by breeders that *E. globulus* has the best combination of wood properties for pulp and paper among the commercially planted *Eucalyptus* species, resulting in a high pulp yield requiring approximately 25 percent less wood to produce the same tonne of cellulose. While only 3.0 m^3 of *E. globulus* wood are required per tonne of pulp, 4 m^3 are needed from selected *E. grandis. E. globulus* has a very adequate wood density in the range of 550 kg/m^3, the longest fibre length and the largest content of holocellulose and pentosans of any other intensively planted species (Sanchez, 2002). *E. globulus*, however, is much more demanding on soil fertility, is not adapted to tropical temperatures, is slower growing and more difficult to propagate clonally than *E. grandis*. In the last ten years, based on the very successful pioneering experiences in Brazil led by Teotônio de Assis, several breeding programmes in tropical countries have started an intensive effort to introgress superior *E. globulus* pulp traits into the tropical and subtropical high yielding genetic backgrounds of *E. grandis* and *E. urophylla*. Given the very high genetic diversity that segregates in such crosses together with the technical possibility of practising intensive within-family selection and clonal propagation, this effort has resulted in the development of exceptional trees that combine superior growth and adaptability to tropical conditions, higher pulp yielding wood and easy propagation using minicutting/hydroponics technology (de Assis, 2000, 2001; Figure 3). A new wave of clonal forestry is therefore starting that will most likely result in another significant jump in the quality of *Eucalyptus* forests.

It is therefore in the context of a highly specialized industrially-oriented breeding programme that fully exploits the power of hybrid breeding and clonal forestry that one needs to discuss the prospects of MAS in *Eucalyptus*. Understanding the fundamental differences between *E. grandis* and *E. globulus* at the molecular level to exploit better the natural allelic variation that exists in the genus has been the starting point.

MARKER-ASSISTED MANAGEMENT OF GENETIC VARIATION IN BREEDING POPULATIONS

The use of genome information for the practice of directional selection of superior genotypes still represents a challenge that depends on further and more refined experimental work (see below). Nevertheless, molecular markers can be used immediately to solve several questions related to the management and identification of genetic variation in breeding and production populations. These applications can be useful essentially to any breeding programme independently of its stage of develop-

FIGURE 3
Selection and clonal propagation of elite trees by the minicutting technology

(A) Elite trees are selected, juvenile sprouts are induced by partial bark stripping while keeping the tree alive. (B) Operational mother plants in hydroponic sand beds from where apical minicuttings are harvested for propagation. (C) Minicuttings are rooted without any use of growth regulators in controlled environment greenhouses. (D) High productivity clonal forest stands that reach over 60 m^3/ha/year.

ment. Although isozyme markers were initially used for these purposes (Moran and Bell, 1983), DNA polymorphisms provide an enhanced level of resolution both at the locus level with much higher expected heterozygosity values and at the genome level with greater coverage. DNA markers provide a powerful tool to quantify existing levels of genetic variation in breeding and production populations of forest trees. Molecular markers can be used to estimate the extent of genetic divergence between individuals selected to compose such populations and resolve several issues of individual identity even at high levels of relatedness, including varietal protection and the verification of alleged parentage in open pollinated breeding systems. Some operational applications of molecular markers for management of genetic variation in *Eucalyptus* are outlined below.

Identification of elite clones

The correct identification of clones is currently the most common application of molecular markers in *Eucalyptus* operational

breeding and production forestry. This application is nowadays routinely used by several forest companies in Australia, Brazil, Portugal, South Africa and Spain. Quality control and quality assurance of large-scale clonal plantation operations become crucial aspects in forestry, especially in vertically integrated production systems where the pulp mill plans on the availability of clones with specific wood properties at specific times. Given the scale of such operations that frequently have to feed plantation programmes of several thousand hectares per year, (i.e. several million seedlings), mislabellings can seriously affect the expected production. Correct clonal identity has also important implications in several breeding procedures such as seed orchard management or controlled pollination programmes affecting the expected gains of breeding cycles.

Several technologies are available today to resolve questions of clonal identity in *Eucalyptus.* Dominant markers such as random amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) have been used for clonal fingerprinting of eucalypts (Keil and Griffin, 1994; Nesbitt *et al.,* 1997; Costa e Silva and Grattapaglia, 1997). Dominant markers are, however, very limited in their ability to establish conclusively the identity of two redundant individual trees due to artefact polymorphisms. Dominant markers can be used to establish that two individuals are not the same, but the statement that two individuals are identical is usually only approximate and no formal test statistics can be attached to this assertion. The high degree of multi-allelism and the very clear and simple co-dominant Mendelian inheritance of microsatellites provide an extremely powerful system for the unique identification of individuals for fingerprinting purposes and parentage testing particularly when the individuals are expected to be related. Kirst *et al.* (2005a) demonstrated the high resolving power of this class of markers in *Eucalyptus*. A breeding population of 192 individuals of *E. grandis* was genotyped with a set of six highly polymorphic microsatellites. The number of alleles detected ranged from 6 to 33 with an average of 19.8±9.2 and the expected heterozygosity averaged 0.86±0.11. Using three loci all 192 genotypes could be readily discriminated. The combined probability of identity (i.e. the probability of two individuals having the same multilocus genotype) considering all six loci was less than one in 2 000 million. Similarity coefficients estimated from microsatellite data were much smaller, thus more discriminative, than those usually obtained in similar studies with RAPD and AFLP markers. In common with human forensic DNA analysis, the standard method for clonal identification in eucalypts today is based on multiplexed, multicolour fluorescent analysis of microsatellite markers sized in an automatic sequencer. The identity of samples is declared based on a maximum likelihood ratio where the likelihood of observing those genetic data conditional on the hypothesis of the two samples being derived from the same clone is compared with the alternative hypothesis, i.e. that the two samples are derived from different clones. Furthermore, the repeatability and precision of multilocus genotype determination allows correct comparisons across laboratories and at different times.

Varietal protection

Following publication of the varietal protection law in Brazil, specific instructions for protecting *Eucalyptus* clones were published in 2002 by the Ministry

of Agriculture of Brazil based on a set of validated morphological descriptors. To the best of the author's knowledge, this is the only country today that has formalized such descriptors, which include 36 morphological traits of leaves, flowers, bark and fruit as well as wood density. Although these descriptors generally satisfy the basic requirements of stability and low environmental influence, they are still difficult to evaluate, especially those related to mature traits in flowers and fruits. Furthermore, it is common that clones are related by common ancestry making their discrimination even more difficult. The high power of discrimination coupled with the general acceptance of DNA technology by eucalypt breeders in Brazil resulted in the inclusion of molecular markers as additional descriptors (Grattapaglia *et al*., 2003). The inclusion of DNA markers represented a remarkable advance that Brazil made in the international landscape of varietal protection of forest trees. Currently all requests for clonal protection are accompanied by a multilocus DNA profile (DNA fingerprint) of 15 to 20 microsatellite markers that were recommended based on several aspects such as robustness, polymorphic information content and general availability in the public domain. The perspective for the following years points to an increased number of applications for clone protection by forest companies in view of the outstanding value of elite eucalypt clones for the maintenance of competitiveness of the forestry-based industry. It can be expected that DNA markers will add a significant power of resolution for distinctness, uniformity and stability (DUS) tests in varietal protection of eucalypt clones, especially when closely related individuals are under scrutiny in legal disputes over clonal property.

Characterization of breeding populations

Breeding populations can be characterized by quantifying the levels and organization of genetic variation within and between breeding groups, sublines and progenies. These data can immediately be used to improve the structure of breeding populations, infuse new material and decide on selection, enrichment or elimination of germplasm entries. In the incomplete pedigree systems frequently used in eucalypts, marker-based systems have been used to monitor the levels of random genetic variation throughout the different cycles of a breeding programme thus allowing much greater flexibility and control over the rate of reduction of genetic variability. For example, RAPD markers were successfully used to characterize the wide range of genetic variation in a germplasm bank of *E. globulus* and thereby assist in the designing of further seed collections (Nesbitt *et al.,* 1995). Gaiotto and Grattapaglia (1997) estimated the distribution of genetic variability within and between open pollinated families of a long-term breeding population of *E. urophylla,* and proposed a selection strategy within and between families for incomplete pedigreed populations based on the incorporation of genetic diversity measures. Marcucci-Poltri *et al*. (2003) used AFLP and microsatellite markers to obtain quantitative estimates of genetic diversity in a *E. dunnii* breeding population selected for fitness to subtropical and cold environments. Molecular data were used to design a clonal seed orchard using the nine most divergent pairs of genotypes, thereby retaining 95.2 percent of the total number of alleles from the 140 polymorphic AFLP loci and the four microsatellite loci analysed. In a subsequent study, Zelener *et al.* (2005) selected *E. dunni* trees using trait selection index

and genetic diversity measures estimated from AFLP and microsatellites. Genetic differentiation estimates consistently showed low differentiation among provenances and great differentiation among families suggesting that orchard design should be based on individual or family selection rather than on provenance selection.

Mating and deployment designs based on genetic distance

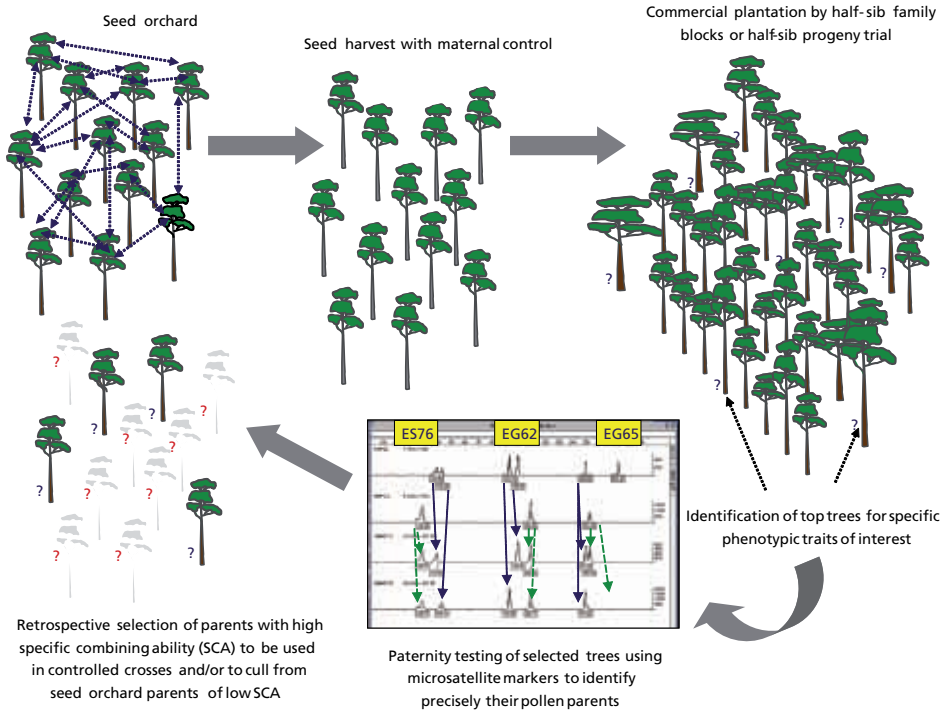
Given the wide genetic diversity and multiple sources of available germplasm for eucalypt breeding, choices typically have to be made as to which elite parents should be mated. Some selection based on the individual's own performance or on pedigree information is used before including it in a mating design. Any means of predicting tree performance deserves attention. One of the "holy grails" of molecular breeders has been the ability to predict progeny performance accurately based on distance estimates among parents from genetic marker data. Vaillancourt *et al.* (1995a) used genetic distances based on RAPD markers to predict heterosis in *E. globulus* progenies. The ability of genetic distance to predict heterosis was significant but accounted for less than 5 percent of the variation in specific combing ability. Baril *et al.* (1997) used the structure of RAPD genetic diversity within and between *E. grandis* and *E. urophylla* to work out prediction equations for the tree trunk volume of individual hybrids at 38 months. Surprisingly, this study showed that a genetic distance based on RAPD markers with similar frequencies in the two species successfully predicted the value of a cross. Through this model, the distance calculated between species explained the general combining ability and the specific combining ability of volume growth with a global coefficient of determination of 81.6 percent. RAPD markers were used to recommend more divergent crosses in a reciprocal recurrent selection programme for hybrid breeding in Brazil (Ribeiro, Bertolucci and Grattapaglia, 1997). A set with the 20 most and 20 least divergent crosses between populations was recommended. Matings between more divergent individuals will potentially allow segregation to be maximized in the resulting progenies and transgressive segregants to be recovered and used as clones.

RAPD data were used to quantify relatedness among elite eucalypt clones for deployment purposes. As the history of selective breeding in eucalypts is very recent, little, if any, pedigree information is typically available. Furthermore, clonal plantations of *Eucalyptus* generally involve only a few superior genotypes of unknown origin. Costa e Silva and Grattapaglia (1997) used RAPD markers to quantify the genetic relatedness among a group of 15 elite clones. Comparative similarity analyses showed that there was significantly more genomic variation in the group of clones than both within and between unrelated half-sib families from a single species. Data on genetic similarity among clones were also used to propose a deployment strategy in a "genetic mosaic", i.e. avoiding planting more genetically related clones side by side in contiguous forest blocks. This proposed strategy was based on the premise that related clones share a common origin and ancestry, have been subject to similar evolutionary selective pressures, and therefore share common susceptibility/tolerance alleles at pest and pathogen defence loci.

Mating system and paternity in breeding populations

Open pollinated breeding by controlling exclusively the maternal progenitor and

FIGURE 4
Retrospective selection strategy of elite parents based on paternity testing of progeny individuals displaying superior performance using microsatellite markers



half-sib progeny testing is still common practice in some eucalypt breeding programmes. It is a low cost option that allows good estimation of the breeding value of maternal parents. Nevertheless, a large amount of genetic variation is usually encountered within half-sib families and selection intensity within families is limited by the number of individuals usually deployed in a progeny test. Knowledge of outcrossing versus selfing rates is essential for maintaining adequate levels of genetic variability for continuous gains. A number of studies have shown that eucalypts are preferentially outcrossed both in natural populations as well as seed orchards. Isozyme markers were originally used (Moran, Bell and Griffin, 1989), but other types of markers now provide a much higher level of resolution. Outcrossing rate in an open pollinated breeding population of *E. urophylla* was estimated at 93 percent using RAPD markers, indicating predominant outcrossing and maintenance of adequate genetic variability within families (Gaiotto, Bramucci and Grattapaglia, 1997). A complex pattern of mating was described in a *E. regnans* seed orchard in Australia where gene dispersal was influenced by crop fecundity and orchard position of mother trees with approximately 50 percent of effective pollen gametes coming from males more than 40 metres away from mother trees (Burczyk *et al.*, 2002). In a

detailed mating system study in a *E. grandis* orchard in Madagascar, the outcrossing rate was found to be 96.7 percent but a pollination rate from outside the seed orchard of 39.2 percent was estimated based on six microsatellite markers (Chaix *et al*., 2003).

The ability to determine paternity precisely using DNA markers was recently proposed as a short-term breeding tactic for *Eucalyptus*. The conventional way to drive modifications in old forest tree seed orchards is to establish progeny trials involving each parent tree and then evaluate its contribution to the performance of the progeny by estimating its general and specific combining ability (GCA and SCA). Grattapaglia, Ribeiro and Rezende (2004) successfully applied an alternative retrospective parent selection tactic based on paternity testing of superior offspring. After identifying seed mixtures, selfed individuals and offspring sired by pollen parents outside the orchard, one particular pollen parent was found to have sired significantly more high-yielding progeny trees. Based on these results, low reproductive success parents were culled from the orchard and management procedures were adopted to minimize external pollen contamination. A significant difference ($p < 0.01$) in mean annual increment was observed between forest stands produced with seed from the orchard before and after selection of parents and revitalization of the orchard. An average realized gain of 24.3 percent in volume growth was obtained from the selection of parents as measured in forest stands at age two to four years. The marker-assisted tree breeding tactic efficiently identified top parents in a seed orchard and resulted in an improved seed variety. It should be applicable for rapidly improving the quality of output from seed orchards especially when the breeder is faced with an emergency demand for improved seeds (Figure 4).

MOLECULAR BREEDING

Molecular markers and maps for *Eucalyptus*

In the last ten years a number of studies have reported genetic maps for *Eucalyptus* built from combinations of several hundred RAPD, AFLP or RFLP markers (Grattapaglia and Sederoff, 1994; Verhaegen and Plomion, 1996; Marques *et al.,* 1998; Myburg *et al.,* 2003), together with RFLP, isozymes, EST, genes and some microsatellites (e.g. Byrne *et al*., 1995; Gion *et al.,* 2000; Bundock, Hayden and Vaillancourt, 2000; Thamarus *et al.,* 2002; Brondani *et al.,* 2002). In contrast to crop species where mapping populations are designed based on contrasting inbred lines, map construction in eucalypts has relied on available pedigrees drawn from operational breeding programmes. These pedigrees generally involve only the highly heterozygous parents and their F_1 progeny, either full-sibs of half-sibs. Genetic mapping has therefore been carried out using a pseudo-testcross strategy, analysing dominant markers present in one parent and absent in the other (Grattapaglia and Sederoff, 1994). Maps are therefore individual-specific and cannot be aligned or integrated as such unless other markers common to both maps are also used. Consequently, although some genome maps of eucalypts have been constructed, the use of the linkage information tends to remain restricted to the pedigree employed as the mapping population, limiting the interexperimental sharing of linkage mapping and QTL data generated.

It is now well accepted that true advancements in QTL validation across pedigrees for the eventual practice of MAS in *Eucalyptus*, will strongly depend on the availability of

higher-throughput, higher polymorphism typing systems such as microsatellites, organized in dense genetic maps (Brondani *et al.,* 1998; Thamarus *et al.,* 2002). Only 137 autosomal microsatellite markers have been published to date for species of *Eucalyptus*, including 67 from *E. globulus, E. nitens, E. sieberi* and *E. leucoxyon* (Byrne *et al.,* 1996; Steane *et al*., 2001; Glaubitz, Emebiri and Moran, 2001; www.ffp.csiro.au/tigr/molecular/eucmsps.html; Ottewell *et al.*, 2005) and 70 from *E. grandis* and *E. urophylla* (Brondani *et al*., 1998; Brondani, Brondani and Grattapaglia, 2002). Recently a set of 35 chloroplast DNA microsatellites was developed based on the full cp-DNA sequence of *E. globulus* (Steane, Jones and Vaillancourt, 2005). Microsatellite transferability across species of the subgenus Symphyomyrtus, which includes all the most widely planted species, varies between 80 and 100 percent depending on the section to which they belong. It still remains around 50 to 60 percent for species of different subgenera such as Idiogenes and Monocalyptus and goes down to 25 percent for the related genus Corymbia (Kirst *et al.,* 1997). Microsatellite comparative mapping data have also shown that genome homology across species of the same subgenus Symphyomyrtus is very high, not only in terms of microsatellite flanking sequence conservation, but also marker order along linkage maps (Marques *et al.,* 2002). Although some tens of microsatellites have been mapped on existing RAPD and AFLP framework maps (Brondani, Brondani and Grattapaglia, 2002; Marques *et al.,* 2002; Thamarus *et al.,* 2002), the genus *Eucalyptus* still lacks a more comprehensive genetic map widely useful for molecular breeding practice. To fill this gap, a novel set of 230 new microsatellites has recently been developed and a consensus map assembled covering at least 90 percent of the recombining genome of *Eucalyptus.* This map has 234 mapped loci on 11 linkage groups, an observed length of 1 568 cM and a mean distance between markers of 8.4 cM (Brondani *et al.,* 2006). This represents an important step forward for *Eucalyptus* comparative genomics, opening stimulating perspectives for evolutionary studies and molecular breeding applications. The generalized use of an increasingly larger set of interspecific transferable markers and consensus mapping information will allow faster and more detailed investigations of QTL synteny among species, validation of QTL and expression-QTL across variable genetic backgrounds, and positioning of a growing number of candidate genes co-localized with QTL, to be tested in association mapping experiments.

QTL mapping in *Eucalyptus*

Following the construction of linkage maps, several groups have reported the identification of genomic regions that have a significant effect on the expression of economically important traits in *Eucalyptus*. QTL mapping experiments have, without exception, found a few major effect QTL for all traits considered in spite of the limited experimental precision, the lack of pre-designed pedigree to maximize phenotypic segregation, and the relatively small segregating populations evaluated. This can be explained by the undomesticated nature and wide genetic heterogeneity of eucalypts added to the fact that most QTL mapping experiments were carried out in interspecific populations thus taking advantage of contrasting gene pools. QTL for juvenile traits such as seedling height, leaf area and seedling frost tolerance have been mapped (Vaillancourt *et al*., 1995b; Byrne *et al.,* 1997a, b), while traits related to vegetative

propagation ability such as adventitious rooting, stump sprouting and *in vitro* shoot multiplication have also been detected (Grattapaglia, Bertolucci and Sederoff, 1995; Marques *et al.,* 1999), as has a major QTL for early flowering (Missiaggia, Piacezzi and Grattapaglia, 2005). In addition, QTL for insect resistance and essential oil traits were mapped (Shepherd, Chaparro and Teasdale, 1999) and recently a major QTL for *Puccinia psidii* rust resistance with quasi Mendelian inheritance was found and mapped in *E. grandis* (Junghans *et al.,* 2003). Major QTL were also found for rotation age traits such as volume growth, wood specific gravity, bark thickness and stem form (Grattapaglia *et al*., 1996; Verhaegen *et al*., 1997; Thamarus *et al*., 2004; Kirst *et al.,* 2004, 2005b).

Although QTL of relatively large effects have been detected for growth traits, when it comes to potential application in MAS the best opportunities for QTL mapping are those related to specialized wood properties that have a direct impact on industrial processes. These traits are usually difficult to measure both because they require destructive whole stem sampling and because they are traits that are expressed late. Myburg (2001) demonstrated the application of indirect, high-throughput phenotyping of wood quality traits in *Eucalyptus* by near infrared reflectance spectroscopy (NIRS) for QTL mapping in a hybrid *E. grandis* x *E. globulus* backcross population. Approximately 300 individuals that had been previously genotyped with AFLP markers were analysed by NIRS, and predictions made for pulp yield, alkali consumption, basic density, fibre length and coarseness, and several wood chemical properties (lignin, cellulose and extractives). A variety of molecular marker classes and pedigree types were used in these experiments. QTL were detected in F_1, inbred or outbred F_2 and half-sib families with or without clonal replicates. Also looking at wood quality traits, Thamarus *et al.* (2004) used novel high-throughput and traditional methods to quantify wood density, fibre length, pulp yield and microfibril angle (MFA) in two full-sib families of *E. globulus* that shared a common parent. Pulp yield and cellulose content were determined by NIRS, and MFA was quantified by SilviScan. Except for fibre length, QTL for all traits could be detected in both populations, including three QTL in common genetic regions on both crosses for wood density, one for pulp yield and one for MFA. The proportion of phenotypic variation explained by the QTL identified in both crosses ranged from 3.2 to 15.8 percent.

Recently QTL analysis of transcript levels of lignin-related genes showed that their mRNA abundance is regulated by two genetic loci co-localized with QTL for growth, suggesting that the same genomic regions are regulating growth, lignin content and composition (Kirst *et al.,* 2004). In a subsequent study, Kirst *et al*. (2005b) showed that one identified expression QTL explained up to 70 percent of the transcript level variation for over 800 genes and that hotspots with co-localized expression QTL were identified on single tree AFLP typically containing genes associated with specific metabolic and regulatory pathways, suggesting coordinated genetic regulation. The correlation of gene expression profiles in segregating progeny can also extend knowledge about genes involved in these pathways. Complementary DNAs representing previously uncharacterized or hypothetical genes, whose transcript levels are strongly correlated with those of genes with known functions, may be associated with the same pathway or biological process.

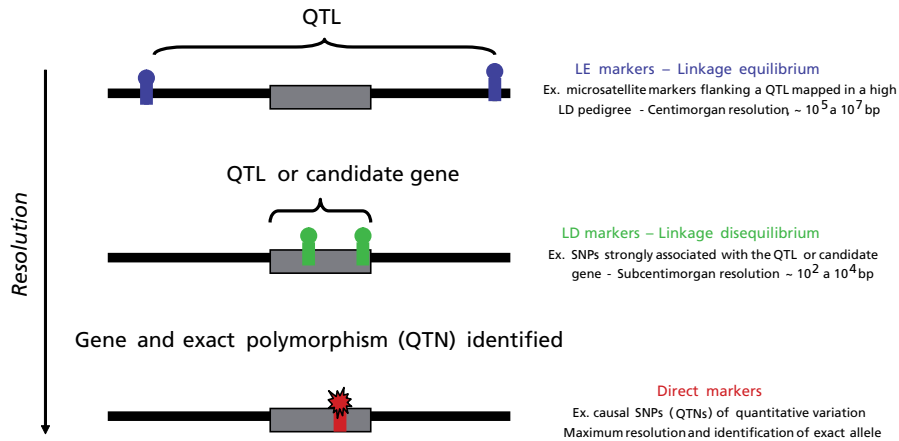
Similarly, new functions can tentatively be assigned to previously characterized genes that had not been described in the context of revealing pleiotropic action of these genes. However, a major limitation in this type of study in *Eucalyptus* is the lack of a completed genome sequence because, without this, the relative locations of large numbers of genes and their expression QTL cannot be determined. This information is required to assess whether the genetic control of gene expression variation is in the cis- or trans-form for each gene. In the context of MAS in *Eucalyptus*, a better understanding of such "master expression QTL" that apparently control cascades of gene expression of important biochemical pathways may be very promising targets for detailed characterization in association mapping experiments to uncover relevant polymorphisms to be used in molecular breeding practice.

In summary, although the number of reports detecting QTL in *Eucalyptus* has grown and these have become increasingly sophisticated, the large majority of mapped QTL have been localized on RAPD or AFLP maps. Consequently, it is impossible to compare positions of QTL for the same or correlated traits, seriously limiting the long-term value of such mapping for MAS. Exceptions are QTL studies where transferable markers such as a few microsatellites (Marques *et al*., 2002; Thamarus *et al.,* 2004) or candidate genes (Gion *et al.,* 2000; Thamarus *et al.,* 2004) were also mapped so that it is at least possible to make a rough preliminary comparison of QTL locations at the linkage group level. Especially in the genus *Eucalyptus* where breeders worldwide take advantage of interspecific genetic variation for wood properties and disease resistance through hybridization, the recent availability of a robust, genus-wide genetic map with highly transferable microsatellite markers (Brondani *et al*., 2006) should stimulate improved genomic undertakings including QTL validation across pedigrees, co-localization of QTL and candidate genes for guiding association mapping experiments, positional cloning of QTL and eventually MAS.

MAS in *Eucalyptus*

Twenty years have passed since the first demonstrations that QTL for major effects could be mapped with molecular markers (Stuber *et al*., 1980; Paterson *et al.,* 1988; Lander and Botstein, 1989), and several reviews have described the potential benefits and caveats of MAS in the plant genetics literature (e.g. Tanksley, 1993; Beavis, 1998; Young, 1999; Mauricio, 2001; Dekkers and Hospital, 2002). Yet, large-scale operational MAS is still largely restricted to very few crops and for very specific applications. Maize is probably the best example, where the financial returns on hybrid seed development coupled with the ability to control germplasm fully, has prompted large-scale investments in MAS by the private sector based on high-throughput single nucleotide polymorphism (SNP) genotyping platforms. Based on a detailed understanding of the molecular architecture of quantitative traits, current applications include yield oriented advanced backcross QTL (AB-QTL) systems as well as accelerated line conversion following trait introgression by marker-assisted backcrossing (MABC). In *Eucalyptus* and forest tree breeding in general, the application of molecular markers for directional selection is still an unfulfilled promise. This is largely due to: the recent domestication of tree crops and hence the wide genetic heterogeneity of breeding populations; the inability to develop inbred lines at least on a short-term basis to allow a

FIGURE 5
Classification of three different types of marker-trait associations relevant to *Eucalyptus* MAS (see text for details)

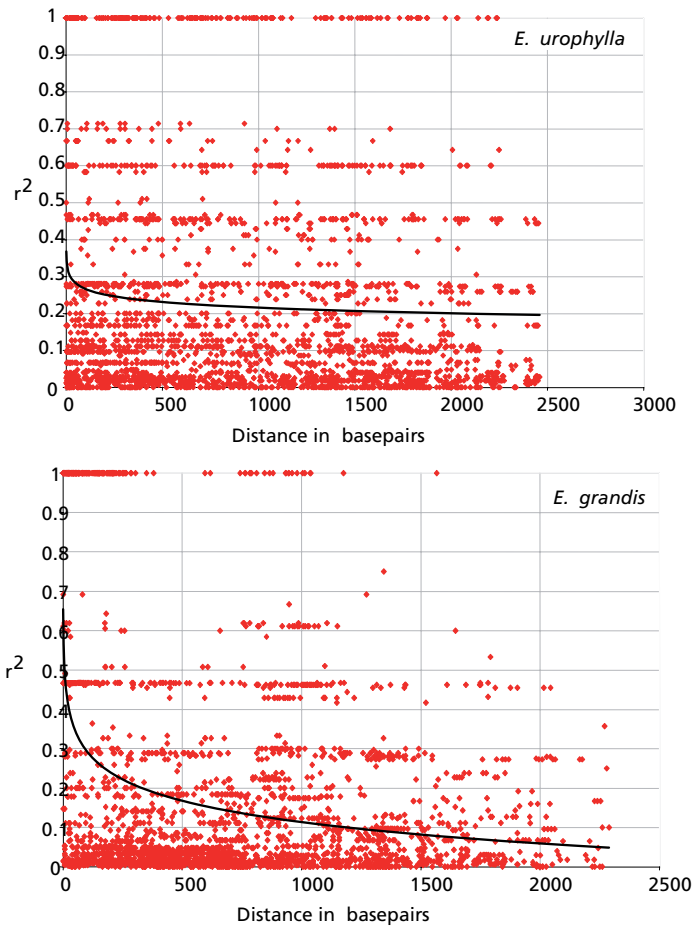


more precise understanding of genetic architecture of quantitative traits; the absence of simply inherited traits that could be immediately and more easily targeted; and finally to the very limited number of scientists actually working on forest trees.

If applying MAS in other intensively studied crops besides maize is already a significant challenge; this challenge is even more difficult and complex for *Eucalyptus* and forest trees in general as it pre-supposes: (i) the manipulation of polygenic traits with variable heritabilities in breeding populations with a heterogeneous genetic base and in linkage equilibrium; (ii) its incorporation in breeding schemes that involve altering the frequencies of favourable alleles through recurrent selection in large populations; and (iii) dealing with age x age trait correlations, and late expressing phenotypes (Grattapaglia, 2000). In applying MAS for forest trees, more will likely be learned from experiences in livestock (Dekkers, 2004; Chapter 10) than from annual crop plants, with the added advantage that gains can be quickly realized by large-scale cloning of selected individuals. In this context, the categorization of three different levels of marker-trait association described by Dekkers (2004) are relevant to trees: (a) direct markers, i.e. loci that code for the functional mutation; (b) linkage disequilibrium (LD) markers: loci that are in population-wide LD with the functional mutation; (c) linkage equilibrium (LE) markers: loci that are in population-wide LE with the functional mutation in outbred populations (Figure 5). In forest trees, besides the recent encouraging discovery of an LD marker for MFA in *Eucalyptus* (Thumma *et al.*, 2005), only LE marker-trait associations have been described. LE markers have been readily detected on a genome-wide basis by analysing large full-sib families with sparse marker maps allowing the detection of most QTL of moderate to large effects. For the other two types of marker-trait association, it is only now that the first association mapping experiments are being started to uncover

FIGURE 6

Rapid decay of intragenic linkage disequilibrium estimated as r^2 in the cinnamyl-alcohol dehydrogenase (cad) gene in two different *Eucalyptus* species



Source: Faria *et al.*, 2006.

LD markers, i.e. polymorphisms that are sufficiently close to the functional mutation (Neale and Savolainen, 2004). The challenge, however, is considerable, as LD in outcrossing forest trees such as pines decays very rapidly, in general within 1 500 to 2 000 bp (Neale and Savolainen, 2004), and similar behaviour has been seen in the few *Eucalyptus* genes analysed to date with significant LD extending for only a few hundred base pairs (Thumma *et al.*, 2005; Kirst, Marques and Sederoff, 2005; Faria *et al.*, 2006) (Figure 6). Genome-wide association studies for LD marker-trait discovery in tress will require very high SNP marker densities that are currently still impracticable (but see below), so that the only alternative left is a candidate gene approach. Finally, direct markers (i.e. polymorphisms that code for the functional mutations) would be the most valuable and

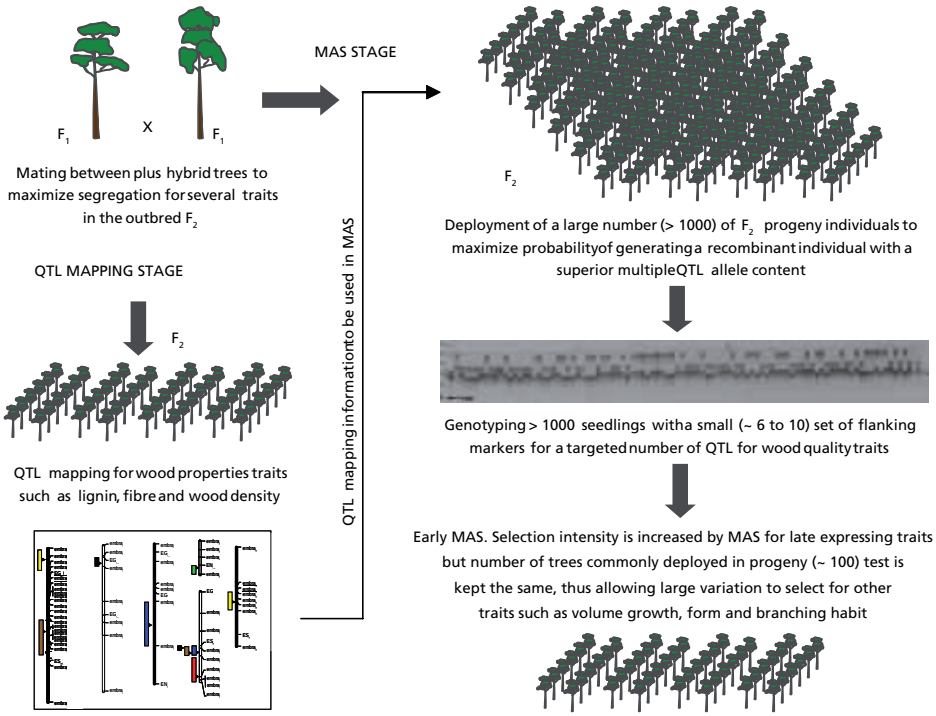
directly applicable in breeding. However, they are the most difficult to detect because causality is very difficult to prove unless very high penetrance Mendelian inheritances are tackled.

Prospects for using MAS in *Eucalyptus*

Eucalyptus breeding programmes vary broadly according to several aspects including the target species or hybrid, the possibility of deploying clones and the amount of resources available to the breeder. However, from the standpoint of integrating MAS, a reasonable premise is that this will only be a justifiable option when the breeding programme has already reached a relatively high level of sophistication, fully exploiting all the accessible breeding and propagation tools. Advanced breeding programmes that aim at elite clone selection involve a significant amount of time and effort being devoted to clonal testing before effective recommendations can be made concerning new clones for operational plantations. Small subline breeding for hybrid performance combined with clonal propagation of selected individuals is being used increasingly for extracting new elite clones (Potts, 2004). The recombination step of a breeding cycle involves the generation of several segregating progenies from selected parents derived from recurrent selection programmes for general combining ability, or reciprocal recurrent selection programmes for hybrid combining ability. This latter strategy has been adopted in tropical countries where the two reciprocal populations are actually two different species such as *E. grandis* and *E. urophylla*. Controlled crosses that were once an important obstacle for implementing pedigreed selection methods are now used routinely after the relatively recent advances made in controlled pollination methods for *Eucalyptus* (Harbard, Griffin and Espejo, 1999; de Assis, Warburton and Harwood, 2005) (see Figure 2). Progeny trials, together with expanded single family plots where larger numbers of full-sibs per family are deployed, are used to allow very intensive within-family selection based on all the available information, both at the family and individual level using BLUP-based selection indices. This selection is generally carried out at half-rotation age based on growth performance and on a preliminary assessment of wood specific gravity using indirect non-destructive techniques such as pilodyn penetration and/or NIRS and Raman spectroscopy (Schimleck, Michell and Vinden, 1996). Vegetative propagules are then rescued from selected trees either by coppicing, sequential grafting or *in vitro* techniques, multiplied and then used for the establishment of clonal tests.

This breeding scheme generates large amounts of linkage disequilibrium by hybridization and substantial amounts of non-additive genetic variation can be captured by vegetative propagation. These are favourable conditions for MAS in forest trees (Strauss, Lande and Namkoong, 1992). Favourable alleles at QTL segregating within-families could be efficiently tagged with microsatellite markers in linkage equilibrium with the actual functional polymorphisms and used for marker-assisted within-family selection for superior individuals. QTL linked markers could be used to carry out early selection thus reducing the time necessary to carry out the first selection especially for traits related to wood properties, and at the same time reducing the number of trees to be selected, propagated and advanced all the way to clonal trials (Figure 7). Therefore, in the context of molecular breeding, given their relatively short rotations and the possibility of

FIGURE 7
Proposed scheme for early marker-assisted selection of plus trees to be used as clones



QTL mapping is carried out and flanking markers in linkage disequilibrium with favourable alleles at major effect QTL alleles are identified. These markers are then used for early within-family selection in expanded progeny sizes at very early age to select for late expression traits such as wood density and lignin content.

deploying clones to capture non-additive genetic variation, it is reasonable to state that eucalypt is the forest tree crop for which MAS has the best prospects of application.

Quantitative theory as well as common sense suggest that MAS in forest trees should help, particularly in situations where trait heritability is low and selection occurs at the level of the individual tree. However, implementing MAS for such traits is a challenging task as extremely precise QTL mapping information is required and this can only be derived from experiments with large progeny sizes (in the order of several hundred individuals), clonal replicates for increased precision, representative and multiple genetic backgrounds and environments. To date, mapped QTL in forest trees still do not fall into this description although improved experiments are under way (Grattapaglia, 2004). Most experiments have mapped QTL for traits that display intermediate to high heritability and probably did not tag the top alleles that exist in the breeding populations as only a very limited sample of crosses were conducted. Furthermore, given the relatively small progeny sizes used for QTL detection (around 100 to 200 individuals), the estimated magnitude of the effects were largely overestimated following the well known "Beavis effect" (Beavis, 1998).

It is frequently stated that MAS for trees would be most useful for volume growth as this is a universal trait of interest and typically of low heritability at the individual tree level. However, in tropical conditions, it is most likely this will not be the target trait of first choice for MAS. Broad sense heritability at the clone mean level, which is the typical selection unit, is frequently above 0.8, allowing an almost perfect ranking and selection of clones even at very early ages (less than two years) under tropical conditions (Rezende, Bertolucci and Ramalho, 1994). Molecular markers for volume growth in these conditions will hardly make a significant contribution to increasing gain per unit time. The cost of scoring molecular markers dictates that the most likely application of MAS in *Eucalyptus* will be for traits that provide significant added value to the final product such as branching habit (for solid wood) and wood chemical traits, or allow clonal deployment such as adventitious rooting or somatic embryogenesis response. Within all possible quality traits, the option would be for those that display medium to high heritabilities but where phenotype assessment is difficult, expensive or requires waiting until the tree reaches maturity. Wood quality traits typically require the tree to start accumulating late wood and involve relatively lengthy procedures for phenotypic evaluation in the laboratory. These kinds of traits could be interesting targets for MAS in *Eucalyptus*, given that the costs of genotyping are sufficiently competitive and precision is high when compared with direct phenotype measurements. It is important to point out, however, that with the recent developments of fast sampling and indirect wood chemistry measurements based on NIRS (Schimleck, Michell and Vinden, 1996), the potential gain will only be realized on the basis of the time savings provided by very early selection. Selected individuals could be recombined more rapidly following flower induction (Griffin *et al.,* 1993) to produce the next generation, potentially increasing the genetic gain per unit time.

MAS for multiple traits will face many of the same difficulties faced by conventional multiple trait selection. Very large progeny sizes would have to be deployed to have a reasonable probability of recovering genotypes with a combination of favourable alleles at many QTL for many traits. When using MAS, priorities will have to be established not only for traits but also for specific QTL. This will require a very good understanding of the relative magnitude of each QTL, potential QTL x background interactions and pleiotropic effects of QTL. Linkage mapping, however, will allow the breeder to understand the basis of negative correlation between traits and possibly to break unwanted linkages by selecting specific recombinant genotypes.

Once the challenging issues related to the discovery of robust marker-trait associations, either within family (LE markers) or at the population level (LD or direct markers), are dealt with, a realistic strategy for the implementation of MAS in *Eucalyptus* might be to tackle only a few major QTL for a quality trait of significant added value. Theoretically, when the total proportion of the additive genetic variance explained by the marker loci exceeds the heritability of the character, selection on the basis of the markers alone is more efficient than selection on the individual phenotype. Such a goal might be achieved for a specific trait with just a few QTL alleles responsible for large effects. On the other hand, if no major gene is detected in an experiment of reasonable size, it might be wiser to dismiss

MAS for that particular trait. Estimates of heritability for a trait might be useful to give an initial clue. Intuitively, the probability of major genes existing for traits of low heritability is lower than for traits of high heritability. However, this should not be taken as a measure to discard possible QTL mapping experiments as, even with low heritabilities, traits might still display major QTL, and MAS would have the greatest impact particularly in such cases.

CONCLUSIONS AND PERSPECTIVES

The successful application of molecular breeding in *Eucalyptus* will depend heavily on first demonstrating and validating the clear-cut association between a DNA polymorphism and a quantitatively inherited phenotypic trait. In highly heterogeneous eucalypts, while conventional QTL mapping can reveal useful markers to be exploited in within-family selection practices, only a more direct LD mapping approach can uncover population wide applicable marker-trait associations. Such studies based on candidate genes have begun and the first candidate gene association for MFA was detected. However, this association explains only a small proportion (3.4 percent) of the variation to be really exciting news to breeders (Thumma *et al.,* 2005). One of the key issues when embarking on an association mapping experiment is the selection of candidate genes. Maximizing the probability of choosing the proper genes requires levels of knowledge of biochemistry, physiology and development that are generally not yet available even for well defined phenotypes and/or known metabolic pathways.

Following the path taken in human genetics, co-localization of candidate genes and QTL for relevant traits on linkage maps together with integrative expression-QTL mapping (Kirst *et al.,* 2004) could be a powerful way forward, although choosing the correct candidate depends heavily on the precision of the QTL localization. At the moment, there are two possibilities for circumventing the dilemma of choosing candidate genes correctly. The first is microarray-based genotyping with ultra-dense arrays of short (25 nt) oligonucleotides (Borevitz *et al*., 2003; Hazen and Kay, 2003; West *et al.,* 2006) that would allow sufficient throughput for association genetic analysis of thousands of genes at a time. Such an array format could later turn out to be a useful instrument for MAS once validated marker-trait associations have been established. The second would be to have access to a whole genome sequence so that candidate genes in a fine mapping interval delimited by markers flanking a QTL with centimorgan resolution could be mined, reannotated and then analysed in association mapping experiments.

A draft genome of *E. camaldulensis* is currently being sequenced at the Kazusa DNA Research Institute in Japan (T. Hibino, personal communication), and the possibility exists that a fully public 4X draft of the *E. grandis* genome will be sequenced by the Joint Genome Institute of the United States Department of Energy within the next years (J. Tuskan, personal communication) following a proposal recently submitted by an international group of *Eucalyptus* geneticists (www.ieugc.up.ac.za/DOE%20proposal%20-%20final%20-%2026%20July%202006.pdf) who recently formed the International *Eucalyptus* Genome Network (EUCAGEN) (www.ieugc.up.ac.za; Myburg, 2004). Such public collaborative efforts should contribute greatly to the advancement of *Eucalyptus* genetics, genomics and molecular breeding by bringing together existing private data-

bases and genomic resources and thereby expanding the value of such genome sequences. As such genome projects advance and new and more powerful analytical tools become accessible, the true challenge to dissecting the complexity of economically-important traits in *Eucalyptus* and implementing MAS will depend to a large extent on our ability to phenotype trees accurately, analyse the overwhelming amount of genomic data available and translate this into truly useful molecular tools for breeding. MAS should be considered on a case-by-case basis and without overstating the gains to be expected until hard experimental data are accumulated on the actual gains made from its application within industrial forests beyond those which can be attained by comparable investment in conventional phenotypic selection.

ACKNOWLEDGMENTS

I am very thankful to the Ministry of Science and Technology of Brazil and the participating forestry companies in the Genolyptus project for the continued competitive grant support for research and the Brazilian National Research Council (CNPq) for its support through a research fellowship. I am also very grateful to all my undergraduate and graduate students and research collaborators from several universities in Brazil and around the world and to breeders from private companies for discussions and scientific inputs. I wish to make a special acknowledgment to my good friend Teotônio Francisco de Assis, an icon of *Eucalyptus* breeding worldwide, for having introduced me to this wonderful world of *Eucalyptus* genetics and breeding.

REFERENCES

Baril, C.P., Verhaegen, D., Vigneron, P., Bouvet, J.M. & Kremer, A. 1997. Structure of the specific combining ability between two species of *Eucalyptus*. I. RAPD data. *Theor. Appl. Genet*. 94: 796–803.

Beavis, W.D. 1998. QTL analyses: power, precision, and accuracy. *In* H.A. Paterson, ed. *Molecular dissection of complex traits,* pp. 145–162. Boca Raton, FL, USA, CRC.

Binkley, D. & Stape, J.L. 2004 Sustainable management of eucalypt plantations in a changing world. *In* M. Tomé, ed. *Proc. IUFRO Conf. Eucalyptus in a Changing World*. CD-ROM. Aveiro, Portugal, Instituto Investigação de Floresta e Papel (RAIZ). Aveiro, Portugal,

Borevitz, J.O., Liang, D., Plouffe, D., Chang, H.S., Zhu, T., Weigel, D., Berry, C.C., Winzeler, E. & Chory, J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* 13: 513–523.

Borralho, N.M.G. 2001. The purpose of breeding is breeding for a purpose. *In* S. Barros, ed. *Proc. IUFRO Internat. Symp. Developing the Eucalypt of the Future*. CD-ROM. Valdivia, Chile, INFOR.

Borralho, N.M.G., Cotterill, P.P. & Kanowski, P.J. 1993. Breeding objectives for pulp production of *Eucalyptus globulus* under different industrial cost structures. *Can. J. For. Res.* 23: 648–656.

Brandão, L.G., Campinhos, E. & Ikemori, Y.K. 1984. Brazil's new forest soars to success. *Pulp Pap. Int.* 26: 38–40.

Brondani, R.P.V., Brondani, C. & Grattapaglia, D. 2002. Towards a genus-wide reference linkage map for Eucalyptus based exclusively on highly informative microsatellite markers. *Mol. Genet. Genomics* 267: 338–347.

Brondani, R.P.V., Brondani, C., Tarchini, R. & Grattapaglia, D. 1998. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor. Appl. Genet*. 97: 816–827.

Brondani, R.P.V., Williams, E.R., Brondani, C. & Grattapaglia, D. 2006. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* 6: 20 (available at www.biomedcentral.com/1471-2229/6/20).

Brune, A. & Zobel, B.J. 1981. Genetic base populations, gene pools and breeding populations for *Eucalyptus* in Brazil. *Silvae Genetica* 30: 146–149.

Bundock, P.C., Hayden, M. & Vaillancourt, R.E. 2000. Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. *Silvae Genet.* 49: 223–232.

Burczyk, J., Adams, W.T., Moran, G.F. & Griffin, A.R. 2002. Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighbourhood model. *Mol. Ecol*. 11: 2379–91.

Byrne, M., Murrell, J.C., Allen, B. & Moran, G.F. 1995. An integrated genetic linkage map for *Eucalyptus* using RFLP, RAPD and isozyme markers. *Theor. Appl. Genet.* 91: 869–875.

Byrne, M., Marquezgarcia, M.I., Uren, T., Smith, D.S. & Moran, G.F. 1996 Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Aust. J. Bot*. 44: 331–341.

Byrne, M., Murrell, J.C., Owen, J.V., Kriedemann, P., Williams, E.R. & Moran, G.F. 1997a. Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. *Theor. Appl. Genet*. 94: 674–681.

Byrne, M., Murrell, J.C., Owen, J.V., Williams, E.R. & Moran, G.F. 1997b. Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. *Theor. Appl. Genet.* 95: 975–979.

Campinhos, E. 1980. More wood of better quality through intensive silviculture with rapid growth improved Brazilian Eucalyptus. *Tappi* 63: 145–147.

Campinhos, E. & Ikemori, Y.K. 1977. Tree improvement program of *Eucalyptus* spp.: preliminary results. In *3rd Wrld. Consult. Forest Tree Breeding,* pp. 717–738. Canberra, CSIRO.

Chaix, G., Gerber, S., Razafimaharo, V., Vigneron, P., Verhaegen, D. & Hamon, S. 2003. Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. *Theor. Appl. Genet*. 107: 705–712.

Costa e Silva, C. & Grattapaglia, D. 1997. RAPD relatedness of elite clones, applications in breeding and operational clonal forestry, *Proc. Internat. IUFRO Conf. Eucalyptus Genet. & Silvicult.,* pp. 161–166. Salvador, Brazil.

de Assis, T.F. 2000. Production and use of *Eucalyptus* hybrids for industrial purposes. *In* D.G., Nikles, ed. *Proc. QFRI/CRC Workshop on Hybrid Breeding & Genetics of Forest Trees,* pp. 63–74. Brisbane, Australia, Department of Primary Industries.

de Assis, T.F. 2001. The evolution of technology for cloning *Eucalyptus* in a large scale. *Proc. IUFRO Internat. Symp. Developing the Eucalypt of the Future.* CD-ROM. Valdivia, Chile, INFOR.

de Assis, T.F., Rezende, G.D.S.P. & Aguiar, A.M. 2005. Current status of breeding and deployment for clonal forestry with tropical eucalypt hybrids in Brazil. *Proc. XXII IUFRO World Congress. Forests in the balance: linking tradition and technology. International Forestry Review* 7(5): 61. Brisbane, Australia.

de Assis, T.F., Warburton, P. & Harwood, C. 2005. Artificially induced protogyny: an advance in the controlled pollination of *Eucalyptus*. *Austr. Forestry* 68: 27–33.

Dekkers, J.C.M. 2004. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. *J. Anim Sci*. 82 (E supp.): 313–328.

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Revs. Genet.* 3: 22–32.

Doughty, R.W. 2000. *The* Eucalyptus*: a natural and commercial history of the gum tree*. Baltimore, MD, USA, and London, The Johns Hopkins University Press.

Eldridge, K., Davidson, J., Harwood, C. & van Wyk, G. 1993. *Eucalypt domestication and breeding.* Oxford, UK, Clarendon Press.

FAO. 2001. *Global forest resources assessment 2000: main report.* FAO Forestry Paper No. 140. Rome (also available at www.fao.org/forestry/fo/fra/main/index.jsp).

Faria, D.A., Alves, T.P.M., Pereira, R.W. & Grattapaglia, D. 2006. Frequência de SNPs e extensão do desequilíbrio de ligação ao longo dos genes CCR e CAD em *E. grandis, E. globulus* e *E. urophylla*. *Proc. 52nd Brazilian Genet. Congr.,* Abstract GP251. *Foz do Iguaçu.*

Franklin, E.C. 1986. Estimation of genetic parameters through four generations of selection in *Eucalyptus grandis*. *Proc. IUFRO Joint Mtg. Working Parties on Breeding Theory, Progeny Testing and Seed Orchards*, pp. 12–17. Williamsburg, VA, USA.

Gaiotto, F.A. & Grattapaglia, D. 1997. Estimation of genetic variability in a breeding population of *Eucalyptus urophylla* using AFLP (amplified fragment length polymorphism) markers. *Proc. Internat. IUFRO Conf. Eucalyptus Genet. & Silvicult.*, pp. 46–52. Salvador, Brazil, IUFRO

Gaiotto, F.A., Bramucci, M. & Grattapaglia, D. 1997. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* s.t. Blake with dominant RAPD and AFLP markers. *Theor. Appl. Genet*. 95: 842–849.

Gion, J-M., Rech, P., Grima-Pettenati, J., Verhaegen, D. & Plomion, C. 2000. Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Molec. Breeding* 6: 441–449.

Glaubitz, J.C., Emebiri, L.C. & Moran, G.F. 2001. Dinucleotide microsatellites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-base differences. *Genome* 44: 1014–1045.

Grattapaglia, D. 2000. Molecular breeding of *Eucalyptus* - State of the art, operational applications and technical challenges, *In* S.M. Jain & S.C. Minocha, eds. *Molecular biology of woody plants,* pp. 451–474. Dordrecht, Netherlands, Kluwer Academic Publishers.

Grattapaglia, D. 2004. Integrating genomics into *Eucalyptus* breeding. *Gen. Mol. Res*. 3: 369–379.

Grattapaglia, D. & Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.

Grattapaglia, D., Bertolucci, F.L.G. & Sederoff, R. 1995. Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Theor. Appl. Genet*. 90: 933–947.

Grattapaglia, D., Ribeiro, V.J. & Rezende, G.D. 2004. Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. *Theor. Appl. Genet*. 109: 192–199.

Grattapaglia, D., Bertolucci, F.L.G., Penchel, R. & Sederoff, R. 1996. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 144: 1205–1214.

Grattapaglia, D., Pimenta, D., Campinhos, E.N., Rezende, G.D.S. & de Assis, T.F. 2003. Marcadores moleculares na proteção varietal de *Eucalyptus. Proc. 8th Brazilian Forestry Congr.,* pp. 1–13. São Paulo, Brazilian Society of Silviculture.

Griffin, A.R., Burgess, I.P. & Wolf, L. 1988. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Hérit: a review. *Austr. J. Bot*. 36(1): 41–66.

Griffin, A.R., Whiteman, P., Rudge, T., Burgess, I.P. & Moncur, M. 1993. Effect of paclobutrazol on flower-bud production and vegetative growth in two species of Eucalyptus. *Can. J. For. Res*. 23: 640–647.

Harbard, J.L., Griffin, A.R. & Espejo, J. 1999. Mass controlled pollination of *Eucalyptus globulus*: a practical reality. *Can. J. For. Res.* 29: 1457–1463.

Hazen, S.P. & Kay, S.A. 2003. Gene arrays are not just for measuring gene expression. *Trends in Plant Sci.* 8: 413–416.

Ikemori, Y.K., Penchel, R.M. & Bertolucci, F.L.G. 1994. Integrating biotechnology into *Eucalyptus* breeding. In *Proc. Internat. Symp. Wood Biotech.*, pp. 79–84. Tokyo, Japan Wood Research Society and Nippon Paper Industries.

Junghans, D.T., Alfenas, A.C., Brommonschenkel, S.H., Oda, S., Mello, E.J. & Grattapaglia, D. 2003. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. *Theor. Appl. Genet*. 108: 175–80.

Kanowski, P.J. & Borralho, N.M.G. 2004. Economics of tree improvement. *In* J.A. Youngquist, ed. *Encyclopedia of forest science*, pp. 1561–1568. Oxford, UK, Elsevier Science.

Keil, M. & Griffin, A.R. 1994. Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. *Theor. Appl. Genet.* 89: 442–450.

Kellison, R.C. 2001. Present and future uses of eucalypts wood in the world. *In* S. Barros, ed. *Proc. IUFRO Internat. Symp. Developing the Eucalypt of the Future.* CD-ROM. Valdivia, Chile, INFOR.

Kirst, M., Marques, C.M. & Sederoff, R. 2005. Nucleotide diversity and linkage disequilibrium in three *Eucalyptus globulus* genes. *IUFRO Tree Biotech. Conf.,* Section 5, poster 28. 6–11 November 2005. Pretoria, South Africa.

Kirst, M., Brondani, R.P.V., Brondani, C. & Grattapaglia, D. 1997. Screening of designed primer pairs for recovery of microsatellite markers and their transferability among species of *Eucalyptus. Proc. IUFRO Conf. Eucalyptus Genet. & Silvicult.,* pp. 167–171. Salvador, Brazil, Embrapa.

Kirst, M., Myburg, A.A., De Leon, J.P., Kirst, M.E., Scott, J. & Sederoff, R. 2004. Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiol*. 135: 2368–2378.

Kirst, M., Cordeiro, C.M., Rezende, G.D. & Grattapaglia, D. 2005a. Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. *J. Hered*. 96: 161–166.

Kirst, M., Basten, C.J., Myburg, A., Zeng, Z-B. & Sederoff, R. 2005b. Genetic architecture of transcript level variation in differentiating xylem of *Eucalyptus* hybrids. *Genetics* 169: 2295–2303.

Ladiges, P.Y., Udovicic, F. & Nelson, G. 2003. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *J. Biogeogr*. 30: 989–998.

Lander, E.S. & Botstein, D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Marcucci-Poltri, S.N., Zelener, N., Rodriguez Traverso, J., Gelid, P. & Hopp, H. 2003. Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers. *Tree Physiol*. 23: 625–32.

Marques, C.M., Araujo, J.A., Ferreira, J.G., Whetten, R., O'Malley, D.M., Liu, B.H. & Sederoff, R. 1998. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theor. Appl. Genet*. 96: 727–737.

Marques, C.M., Vasquez-Kool, J., Carocha, V.J., Ferreira, J.G., O'Malley, D.M., Liu, B.H. & Sederoff, R. 1999. Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theor Appl Genet* 99: 936–946.

Marques, C.M., Brondani, R.P.V., Grattapaglia, D. & Sederoff, R. 2002. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theor. Appl. Genet*. 105: 474–478.

Martin, B. & Quillet, J. 1974. The propogation by cuttings of forest trees in the Congo. *Bois et Forêts des Tropiques* 155: 15–33.

Mauricio, R. 2001. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Revs. Genet*. 2: 370–381.

Missiaggia, A.A., Piacezzi, A.L. & Grattapaglia, D. 2005. Genetic mapping of *Eef1*, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genet. & Genomes* 1: 79–84.

Moran, G.F. & Bell, J.C. 1983. *Eucalyptus*. *In* S.D. Tanksley & T.J. Orton, eds. *Isozymes in plant genetics and breeding,* pp. 423–441. Amsterdam, Elsevier.

Moran, G., Bell, J.C. & Griffin, A.R. 1989. Reduction in levels of inbreeding in a seed orchard of *Eucalyptus regnans* F. Muell, compared with natural populations. *Silvae Genet.* 38: 32–36.

Moran, G.F., Thamarus, K.A., Raymond, C.A., Qiu, D., Uren, T. & Southerton, S.G. 2002. Genomics of *Eucalyptus* wood traits. *Ann. For. Sci*. 59: 645–650.

Myburg, A.A. 2001. *Genetic architecture of hybrid fitness and wood quality traits in a wide interspecific cross of Eucalyptus tree species*. North Carolina State University, Raleigh, NC (PhD thesis; available at www.lib.ncsu.edu/theses/available/etd-20010723-175234).

Myburg, A.A. 2004. The International *Eucalyptus* Genome Consortium (IEuGC): Opportunities and Resources for Collaborative Genome Research in *Eucalyptus*. *In* B. Li, & S. McKeand, eds. *Proc. IUFRO Conf. Forest Genet. & Tree Breeding in the Age of Genomics: Progress and Future,* pp. 154–155. Charleston, SC, USA.

Myburg, A.A., Griffin, A.R., Sederoff, R.R. & Whetten, R.W. 2003. Comparative genetic linkage maps of *Eucalyptus grandis, Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theor. Appl. Genet*. 107: 1028–42.

Myburg, A.A., Potts, B., Marques, C.M., Kirst, M., Gion, J.M., Grattapaglia, D. & Grima-Pettenati, J. 2006. *Eucalyptus. In* C. Kole, ed. *Genome mapping & molecular breeding in plants volume 7: forest trees*. Heidelberg, Berlin, New York and Tokyo, Springer (in press).

Neale, D.B. & Savolainen, O. 2004. Association genetics of complex traits in conifers. *Trends Plant Sci.* 9: 325–330.

Nesbitt, K.A., Potts, B.M., Vaillancourt, R.E., West, A.K. & Reid, J.B. 1995. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity* 74: 628–637.

Nesbitt, K.A., Potts, B.M., Vaillancourt, R.E. & Reid, J.B. 1997. Fingerprinting and pedigree analysis in *Eualyptus globulus* using RAPDs. *Silvae Genet.* 46: 6–11.

Ottewell, K.M., Donnellan, S.C., Moran, G.F. & Paton, D.C. 2005. Multiplexed microsatellite markers for the genetic analysis of *Eucalyptus leucoxylon*, Myrtaceae and their utility for ecological and breeding studies in other *Eucalyptus* species. *J. Hered*. 96: 445–451.

Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. & Tanksley, S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726.

Poke, F.S., Vaillancourt, R.E., Potts, B.M. & Reid, J.B. 2005. Genomic research in *Eucalyptus*. *Genetica* 125: 79–101.

Potts, B.M. 2004. Genetic improvement of eucalypts, pp 1480–1490. *In* J. Burley, J. Evans, & J.A. Youngquist, eds. *Encyclopedia of forest science*. Oxford, UK, Elsevier Science.

Potts, B.M. & Dungey, H.S. 2004. Hybridisation of *Eucalyptus*: key issues for breeders and geneticists. *New Forests* 27: 115–138.

Potts, B.M., Vaillancourt, R.E., Jordan, G.J., Dutkowski, G.W., Costa e Silva, J., McKinnon, G.E., Steane, D.A., Volker, P.W., Lopez, G.A., Apiolaza, L.A., Li, Y., Marques, C. & Borralho, N.M.G. 2004. Exploration of the *Eucalyptus globulus* gene pool. *In* M. Tomé ed. *Eucalyptus in a changing world*, pp. 46–61. Aveiro, Portugal, Instituto Investigação de Floresta e Papel (RAIZ).

Pryor, L.D. 1976. *The biology of eucalypts*. London, Edward Arnold.

Pryor, L.D. & Johnson, L.A.S. 1971. *A classification of the eucalypts*. Canberra, Australian National University Press.

Rezende, G.D.S.P., Bertolucci, F.L.G. & Ramalho, M.A.P. 1994. Eficiência da seleção precoce na recomendação de clones de eucalipto avaliados no norte do Espírito Santo e sul da Bahia. *Cerne* 1: 45–50.

Ribeiro, V.J., Bertolucci, F.L.G. & Grattapaglia, D. 1997. RAPD marker - guided matings in a reciprocal recurrent selection program of *Eucalyptus. Proc. Internat. IUFRO Conf. Eucalyptus Genet. & Silvicult.*, pp. 156–160. Salvador, Brazil.

Sanchez, J.R. 2002. Panel de tecnologia, pasta y papel. *In* Graficas Diumaró, ed. *Socioeconomia, patologia, tecnologia y sostenibilidade del eucalipto*, pp. 187–192. Brazil, Universidad de Vigo.

Schimleck, L.R., Michell, A.J. & Vinden, P. 1996. Eucalypt wood classification by NIR spectroscopy and principal components analysis. *Appita J*. 49: 319–324.

Shepherd, M., Chaparro, J.X. & Teasdale, R. 1999. Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. *Theor. Appl. Genet*. 99: 1207–1215.

Shepherd, M. & Jones, M. 2005. Molecular markers in tree improvement: Characterisation and use in *Eucalyptus*,. *In* H. Lörz & G. Wenzel, eds. *Molecular marker systems in plant breeding and crop improvement*, pp 399–409. Heidelberg, Germany, Springer-Verlag.

Steane, D.A., Jones, R.C. & Vaillancourt, R.E. 2005. A set of chloroplast microsatellite primers for *Eucalyptus*, Myrtaceae. *Mol. Ecol. Notes* 5: 538–541.

Steane, D.A., Vaillancourt, R.E., Russell, J., Powell, W., Marshall, D. & Potts, B.M. 2001. Development and characterization microsatellite loci in *Eucalyptus globulus* (Myrtacea). *Silvae Genet.* 50: 89.

Strauss, S.H., Lande, R. & Namkoong, G. 1992. Obstacles to molecular-marker-aided selection in forest trees. *Can. J. For. Res*. 22: 1050–1061.

Stuber, C.W., Moll, R.H., Goodman, M.M., Schaffer, H.E. & Weir, B.S. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize. *Genetics* 95: 225–236.

Tanksley, S.D. 1993. Mapping polygenes. *Ann. Rev. Genet.* 27: 205–233.

Thamarus, K.A., Groom, K., Murrell, J., Byrne, M. & Moran, G. 2002. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. *Theor. Appl. Genet.* 104: 379–387.

Thamarus, K.A., Groom, K., Bradley, A., Raymond, C.A., Schimleck, L.R., Williams, E.R. & Moran, G.F. 2004. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theor. Appl. Genet.* 109: 856–864.

Thumma, R.F., Nolan, M.F., Evans, R. & Moran, G.F. 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265.

Turnbull, J.W. 1999. Eucalypt plantations. *New Forests* 17: 37–52.

Vaillancourt, R.E., Potts, B.M., Watson, M., Volker, P.W., Hodge, G.R., Reid, J.B. & West, A.K. 1995a. Prediction of heterosis using RAPD generated distances between *Eucalyptus globulus* trees. In *Proc. CRCTHF-IUFRO Conf. Eucalypt Plantations: Improving Fibre Yield and Quality,* pp. 455–456*.* 19–24 February 1995, Hobart, Australia(available at www.forestry.crc.org.au/iufro95.htm).

Vaillancourt, R.E., Potts, B.M., Manson, A., Eldridge, T. & Reid, J.B. 1995b Using RAPD's to detect QTLs in an interspecific F_2 hybrid of *Eucalyptus.* In *Proc. CRCTHF-IUFRO Conf. Eucalypt Plantations: Improving Fibre Yield and Quality,* pp. 430–433. 19–24 February 1995, Hobart, Australia (available at www.forestry.crc.org.au/iufro95.htm).

Verhaegen, D. & Plomion, C. 1996. Genetic mapping in *Eucalyptus urophylla* and *E grandis*. RAPD markers. *Genome* 39: 1051–1061.

Verhaegen, D., Plomion, C., Gion, J.M., Poitel, M., Costa, P. & Kremer, A. 1997. Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor. Appl.Genet*. 95: 597–608.

Waugh, G. 2004. Growing *Eucalyptus globulus* for high-quality sawn products. *In* M. Tomé, ed. *Eucalyptus in a changing world,* pp. 79–84. Aveiro, Portugal, Instituto Investigação de Floresta e Papel (RAIZ).

West, M.A., van Leeuwen, H., Kozik, A., Kliebenstein, D.J., Doerge, R.W., St Clair, D.A. & Michelmore, R.W. 2006. High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Res.* 16: 787–795.

Young, N.D. 1999. A cautiously optimistic view for marker-assisted selection. *Mol. Breed.* 5: 505–510.

Zelener, N., Poltri, S.N., Bartoloni, N., Lopez, C.R. & Hopp, H.E. 2005. Selection strategy for a seedling seed orchard design based on trait selection index and genomic analysis by molecular markers: a case study for *Eucalyptus dunnii*. *Tree Physiol.* 25: 1457–1467.

CHAPTER 15

Marker-assisted selection in forestry species

Penny Butcher and Simon Southerton

SUMMARY

The primary goal of tree breeding is to increase the quantity and quality of wood products from plantations. Major gains have been achieved using recurrent selection in genetically diverse breeding populations to capture additive variation. However, the long generation times of trees, together with poor juvenile-mature trait correlations, have promoted interest in marker-assisted selection (MAS) to accelerate breeding through early selection. MAS relies on identifying DNA markers, which explain a high proportion of variation in phenotypic traits. Genetic linkage maps have been developed for most commercial tree species and these can be used to locate chromosomal regions where DNA markers co-segregate with quantitative traits (quantitative trait loci, QTL). MAS based on QTL is most likely to be used for within-family selection in a limited number of elite families that can be clonally propagated. Limitations of the approach include the low resolution of marker-trait associations, the small proportion of phenotypic variation explained by QTL and the low success rate in validating QTL in different genetic backgrounds and environments. This has led to a change in research focus towards association mapping to identify variation in the DNA sequence of genes directly controlling phenotypic variation (gene-assisted selection, GAS). The main advantages of GAS are the high resolution of marker-trait associations and the ability to transfer markers across families and even species. Association studies are being used to examine the adaptive significance of variation in genes controlling wood formation and quality, pathogen resistance, cold tolerance and drought tolerance. Single nucleotide polymorphisms (SNPs) in these gene sequences that are significantly associated with trait variation can then be used for early selection. Markers for SNPs can be transferred among individuals regardless of pedigree or family relationship, increasing opportunities for their application in tree breeding programmes in developing as well as developed countries. Significant reductions in genotyping costs and improved efficiencies in gene discovery will further enhance these opportunities.

INTRODUCTION

Tree breeding offers a unique set of challenges associated with long generation times, outcrossing breeding systems and a relatively short history of genetic improvement. Breeding populations are often only one or two generations from the wild state. This has the advantage over crop breeding of providing vast stores of genetic variation that can be utilized in tree improvement. Tree breeding programmes have generally relied on testing and selecting large numbers of genotypes derived from multiple genetic backgrounds, the maintenance of high genetic diversity in production forests, and sexual propagation and capture of additive genetic variation through recurrent selection (Strauss, Lande and Namkoong, 1992). Inbreeding depression and long generation intervals have precluded the use of inbred lines, although research into their development continues (Wu, Abarquez and Matheson, 2004). The greatest use of interspecific hybrids in operational tree breeding has been with introduced species; for example, *Pinus elliottii* x *P. caribaea* in Australia (Nikles, 1996), hybrid eucalypts in South Africa, Brazil and the Congo (Eldridge *et al.,* 1993), *Acacia mangium* x *A. auriculiformis* in Viet Nam (Kha, Hai and Vinh, 1998) and hybrid poplars in temperate regions. These programmes often rely on clonal propagation for deployment.

The goal of commercial tree breeding is to increase the quantity and quality of wood products from plantations. Production of industrial timber was estimated at 2.8 thousand million cubic metres in 2004 and has been increasing at an average annual rate of 2.4 percent since 1998 (FAOSTAT) with much of the recent increase being due to rapid economic growth in China. Consumption of fuelwood is increasing at a similar rate (Carson, Walter and Carson, 2004). Rising demand together with restrictions on the supply of timber from native forests mean that increases in forest productivity will be required. To date, increased production has been achieved by expanding the area of plantations, particularly in tropical regions where high growth rates can be achieved. Gains have also been made using conventional breeding, but further productivity increases are required to reduce pressure on native forests and limit the increases in land area required for plantations. MAS has the potential to enhance plantation productivity if the relationship between genetic variation in gene sequences and phenotypic variation in traits can be demonstrated.

The relatively long generation times and poor juvenile-mature trait correlations in forest trees have promoted interest in MAS to accelerate breeding through early selection. MAS relies on identifying DNA markers which explain a high proportion of additive variation in phenotypic traits. Initially, research focused on the use of DNA markers in genome-wide linkage analysis of progeny arrays (Lander and Botstein, 1989). By identifying patterns of co-segregation in complex traits and polymorphic markers (QTL), these studies aimed to reveal causative regions of the chromosome or gene that were inherited intact over a few generations. The QTL approach can be used for marker-aided breeding within families. The low success rate in validating QTL in different genetic backgrounds and environments (Neale, Sewell and Brown, 2002) led to a change in research focus towards population-level association mapping. This approach seeks to find alleles of genes that affect the phenotype directly (Neale and Savolainen, 2004), and relies on the retention of much

smaller regions of intact DNA over many generations. Candidate genes targeted in these studies can be identified by gene mapping, expressed sequence tag (EST) sequencing, gene expression profiling or functional studies (transgenics). If variation can be found in the sequence of these genes in different phenotypes, it allows MAS to be used for within- and between-family selection in forest trees.

The application of biotechnology in tree improvement research is currently taking different paths in developed and developing countries due to contrasting regulatory approval processes for genetically modified plants and differences in public acceptance of genetically modified organisms (GMOs). There is considerable resistance in developed countries towards transgenic trees, which has more to do with their possible effects on other plants and on the environment than with concerns about transgenic wood (Sedjo, 2004). Long-term field trials are needed to ensure the stability of any genetic modification and the absence of negative impacts on growth and resistance to environmental stresses before they can be incorporated into industrial plantations (Strauss *et al.,* 1998; Strauss *et al.,* 2004). Regulation costs, possible trade restrictions, lack of public acceptance of transgenics and lack of support by major forestry certification groups such as the Forest Stewardship Council (FSC) are currently barriers to the development of transgenics (Sedjo, 2004). Consequently, trials of transgenic trees in developed countries remain in the research phase, mostly conducted with young trees grown under glasshouse conditions (see Walter and Killerby, 2004 for review). Due to these problems, some research has shifted towards alternative methods of investigating gene function and incorporating desirable genes into breeding populations, mainly through association mapping.

Recent MAS research in forest trees has been greatly assisted by advances in our understanding of tree genomes. The complete sequencing of plant genomes such as *Arabidopsis* (*Arabidopsis* Genome Initiative, 2000) and rice (Yu *et al.,* 2002) is improving our understanding of the number of genes involved (25 000–55 000) in the development of different organs and the function of the genes.

The *Populus* genome was the first tree genome to be sequenced with 58 036 predicted genes (www.jgi.doe.gov/poplar) and efforts are under way to sequence the *Eucalyptus* genome (www.ieugc.up.ac.za), a genus of particular importance in countries with developing economies in Asia and South America. To date, partial coverage of the *E. camaldulensis* genome (600 Mb) has been completed by random shotgun sequencing, through collaboration between Oji Paper and the Kasuza DNA Research Institute in Chiba, Japan (S. Potter Ensis, personal communication). A draft sequence, based on four-fold coverage of the genome is expected to be available by mid-2007 (Poke *et al.,* 2005). The large genome size of conifers is currently a barrier to whole genome sequencing; however, comprehensive EST sequencing is likely to yield most genes expressed in target tissues.

Genomic resources and tools are now being established for important forest tree species. Rapidly growing numbers of ESTs are publicly available in a range of species including *Eucalyptus grandis, Pinus radiata*, *P. taeda*, *Picea abies*, *Populus trichocarpa*, *P. tremula x tremuloides* and *Cryptomeria japonica* (see Krutovskii and Neale, 2001 and Strabala, 2004 for reviews) and *Avicennia marina* (Mehta *et al.,* 2005).

Over 80 000 ESTs have been sequenced from pine (http://pinetree.ccgb.umn.edu/), over 130 000 from poplar (http://poppel.fysbot.umu.se/) and over 100 000 from spruce (www.arborea.ulaval.ca/en/; www.treenomix.ca/).

Comprehensive microarrays (Shena *et al.*, 1996) are now being used in many of these species, allowing transcription profiling of thousands of genes in contrasting phenotypes in a range of tissues under different environmental/stress/developmental regimes. Identification of candidate genes from expression profiling is based on the assumption that the genes showing genotype-specific differences in their level of expression are causing variation in that trait (Morgante and Salamini, 2003). Microarrays are being used to identify genes that are regulated differentially in individuals with contrasting wood traits in eucalypts (Moran *et al.*, 2002), symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza (Voiblet *et al.*, 2001) and genes involved in embryogenesis in pines (van Zyl *et al.*, 2003). Using microarrays, many genes involved in cell wall biosynthesis have been identified in loblolly pine (Whetten *et al.*, 2001; Pavy *et al.*, 2005), eucalypts (Paux *et al.*, 2004) and hybrid aspen (*Populus tremula x P. tremuloides*) (Hertzberg *et al.*, 2001). A combination of proteomics, which examines changes in protein expression in different tissue and developmental stages, and microarray technology is also being used to give a more complete picture of gene function, for example of drought tolerance in *Pinus pinaster* (Plomion *et al.*, 2004). This discovery work is uncovering large numbers of candidate genes that are excellent targets for both QTL mapping and association studies aimed at identifying markers for use in MAS.

FAMILY-BASED GENETIC LINKAGE MAPPING AND QTL ANALYSIS

Genetic linkage or recombination mapping relies on finding sufficient polymorphism using DNA markers in progeny arrays from full-sib pedigrees to identify associations between linked loci on a chromosome. Genetic linkage maps have been constructed for most of the commercially important forest tree genera (summarized in Table 1), and updated information on genetic linkage maps for forest trees is available at http://dendrome.ucdavis.edu/index.php. The number and location of chromosomal regions affecting a trait (QTL) and the magnitude of their effect can then be investigated by QTL mapping. QTL are identified by a statistical association between variation in a quantitative trait and segregation of alleles at a marker locus in a segregating population (mapping pedigree).

Most phenotypic traits of interest for tree breeding are characterized by continuous variation. Such traits are usually influenced by a number of genes with a small effect interacting with other genes and the environment. The main traits targeted for QTL mapping are wood properties and traits related to adaptation and growth (reviewed by Sewell and Neale, 2000). These include physical wood properties that affect the strength of sawn timber (e.g. wood density and microfibril angle), and properties that affect paper pulping, e.g. pulp yield, fibre length and the relative proportion of cellulose, hemicellulose and lignin, generally measured as percentage cellulose. In addition, QTL have been identified for disease resistance, growth, flowering, vegetative propagation, frost tolerance and leaf oil composition (Table 2).

The detection of QTL requires large sample sizes; the lower the heritability

TABLE 1
Genetic linkage maps constructed for forest trees, markers used and location of mapping pedigrees

Species	Markers[1]	Country	Reference
Acacia mangium	RFLP, SSR	Australia	Butcher and Moran, 2000
Cryptomeria japonica	RFLP, RAPD, isozyme	Japan	Mukai *et al.,* 1995
Eucalyptus camaldulensis	RAPD, RFLP, SSR	Egypt	Agrama, George and Salah, 2002
Eucalyptus globulus	RAPD, SSR	Australia	Bundock, Hayden and Vaillancourt, 2000
	Candidate genes, isozymes, ESTP, RFLP, SSR	Australia	Thamarus *et al.,* 2002
Eucalyptus grandis x *E. globulus*	AFLP	Uruguay	Myburg *et al.,* 2003
Eucalyptus grandis x *E. urophyllla*	RAPD	Brazil Congo	Grattapaglia and Sederoff, 1994; Verhaegen and Plomion, 1996
Eucalyptus nitens	Isozyme, RAPD, RFLP	Australia	Byrne *et al.,* 1995
Eucalyptus tereticornis x *E. globulus*	AFLP	Portugal	Marques *et al.,* 1998
Eucalyptus urophylla x *E. tereticornis*	RAPD	China	Gan *et al.,* 2003
Fagus sylvatica	AFLP, RAPD, SSR	Italy	Scalfi *et al.,* 2004
Hevea braziliensis x H. benthamiana (rubber tree)	AFLP, isozymes, RFLP, SSR	French Guyana	Lespinasse *et al.,* 1999
Larix decidua & L. kaempferi	AFLP, ISSR, RAPD	France	Arcade *et al.,* 2000
Picea abies	RAPD	Italy	Binelli and Bucci, 1994
	RAPD	Denmark	Skov and Wellendorf, 1998
	AFLP, SAMPL, SSR	Italy	Paglia, Olivieri and Morgante, 1998
Picea glauca	ESTP, RAPD, SCAR	Canada	Gosselin *et al.,* 2002
Pinus edulis	AFLP	USA	Travis *et al.,* 1998
Pinus elliottii var elliottii	RAPD	USA	Nelson, Nance and Doudrick, 1993
Pinus elliottii var. elliottii & P. caribaea var. hondurensis	AFLP, SSR	Australia	Shepherd *et al.,* 2003
Pinus palustris	RAPD	USA	Kubisiak *et al.,* 1996
Pinus pinaster	RAPD	France	Plomion, O'Malley and Durel, 1995
	AFLP, RAPD, protein	France	Costa *et al.,* 2000
	AFLP	France	Chagne *et al.,* 2002
Pinus radiata	RFLP, RAPD, SSR	Australia	Devey *et al.,* 1996
Pinus sylvestris	RAPD	Sweden	Yazdani, Yeh and Rimsha, 1995
Pinus taeda	Isozymes, RAPD, RFLP	USA	Devey *et al.,* 1994; Sewell, Sherman and Neale, 1999
	AFLP	USA	Remington *et al.,* 1999
Pinus thunbergii	AFLP, RAPD	Japan	Hayashi *et al.,* 2001
Populus	AFLP, candidate genes, isozymes, ISSR, RAPD, RFLP, STS, SSR	Belgium, France, USA	See review in Cervera *et al.,* 2004; Yin *et al.,* 2004
Pseudotsuga menziesii	RAPD, RFLP	USA	Jermstad *et al.,* 1998; Krutovskii *et al.,* 1998
Quercus robur	Isozyme, minisatellite, RAPD, SCAR, SSR. 5SrDNA	France	Barreneche *et al.,* 1998
Salix viminalis	AFLP, SSR	UK	Hanley *et al.,* 2002
Salix viminalis x S. schwerinii	AFLP, RFLP	Sweden	Tsarouhas, Gullberg and Lagercrantz, 2002

[1] AFLP = amplified fragment length polymorphism; ESTP = expressed sequence tagged polymorphism; ISSR = inter-simple sequence repeats; RAPD = random amplified polymorphic DNA; RFLP = restriction fragment length polymorphism; SAMPL = selective amplification of microsatellite polymorphic loci; SCAR = sequence characterized amplified regions; SSR = simple sequence repeat (microsatellite); STS = sequence-tagged sites

TABLE 2
Quantitative trait loci reported for forest tree species

Species	Markers[1]	QTL	Reference
Acacia mangium	RFLP, SSR	Disease resistance	Butcher, 2004
Cryptomeria japonica	Isozyme, RAPD, RFLP	Juvenile growth, flowering, vegetative propagation	Yoshimaru *et al.,* 1998
	RAPD	Wood quality	Kuramoto *et al.,* 2000
Eucalyptus globulus	Isozyme, RFLP, SSR	Wood density, pulp yield, microfibril angle	Thamarus *et al.,* 2004
Eucalyptus grandis	RAPD	Growth, wood density	Grattapaglia *et al.,* 1996;
	RAPD	Disease resistance	Junghaus *et al.,* 2003
E.grandis x *E. urophylla*	RAPD	Vegetative propagation	Grattapaglia, Bertolucci and Sederoff, 1995; Marques *et al.,* 1999
	RAPD	Growth, wood density	Verhaegen *et al.,* 1997
	RAPD	Leaf oil composition	Shepherd, Chaparro and Teasdale, 1999
Eucalyptus nitens	RFLP	Growth	Byrne *et al.,* 1997a
	RFLP	Frost tolerance	Byrne *et al.,* 1997b
Eucalyptus tereticornis x *E. globulus*	AFLP	Vegetative propagation	Marques *et al.,* 2002
Fagus sylvatica	AFLP, RAPD, SSR	Leaf traits, growth	Scalfi *et al.,* 2004
Hevea braziliensis x H. benthamiana	AFLP, isozyme, RFLP, SSR	Disease resistance	Lespinasse *et al.,* 2000
Pinus palustris x *P. elliottii*	RAPD	Juvenile growth	Weng *et al.,* 2002
Pinus pinaster	RAPD	Bud set, frost tolerance	Hurme *et al.,* 2000
	AFLP	Growth, water use efficiency	Brendel *et al.,* 2002
Pinus radiata	RAPD	Growth	Emebiri *et al.,* 1997,1998a,b
	AFLP, RAPD, SSR	Wood density	Kumar *et al.,* 2000
	RFLP, SSR	Growth, wood density, disease resistance	Devey *et al.,* 2004a,b
Pinus sylvestris	AFLP	Growth, cold acclimation	Lerceteau, Plomion and Andersson, 2000; Yazdani *et al.,* 2003
Pinus taeda	Isozymes, RAPD, RFLP	Growth	Kaya, Sewell and Neale, 1999
	RFLP	Physical wood properties	Groover *et al.,* 1994; Sewell *et al.,* 2000
	RFLP	Chemical wood properties	Sewell *et al.,* 2002
	ESTP, RFLP	Wood properties	Brown *et al.,* 2003
Populus	AFLP, ISSR, RAPD, RFLP, SCAR, SSR, STS	Growth, form, leaf architecture, leaf & bud phenology, disease resistance, wood quality	See review in Cervera *et al.,* 2004
Pseudotsuga menziesii	RAPD, RFLP	Vegetative bud flush	Jermstadt *et al.,* 2001a
	RAPD, RFLP	Cold hardiness	Jermstadt *et al.,* 2001b
	RAPD, RFLP	QTL x environment	Jermstadt *et al.,* 2003
Quercus robur	AFLP, RAPD, SCAR, SSR	Growth, bud burst	Saintagne *et al.,* 2004
Salix dasyclados x *S. viminalis*	AFLP	Growth, drought tolerance, bud flush	Tsarouhas, Gullberg and Lagercrantz, 2002, 2003; Rönnberg-Wästljung, Glynn and Weih, 2005

[1] AFLP = amplified fragment length polymorphism; ESTP = expressed sequence tagged polymorphism; ISSR = inter-simple sequence repeats; RAPD = random amplified polymorphic DNA; RFLP = restriction fragment length polymorphism; SCAR = sequence characterized amplified regions; SSR = simple sequence repeat (microsatellite); STS = sequence-tagged sites

of the trait and the larger the number of genes affecting the trait, the larger the sample size required (see Strauss, Lande and Namkoong, 1992). As shown by Brown *et al.* (2003), the use of small mapping populations of 100–200 segregating individuals, typical of most QTL studies in trees, is likely to cause an upward bias in the estimated phenotypic effect of QTL. Simulation and practical studies have shown that, in addition to sample size, QTL detection is affected by genetic background, environment and interactions among QTL. The location of QTL is imprecise as they can only be mapped to 5–10 cM. This may translate into a physical distance of several megabases, which may contain several hundred genes. The effect of a QTL is also likely to vary over time in perennial plants with changing biotic and abiotic factors (Brown *et al.,* 2003). This highlights the necessity of verifying QTL in different seasons, environments and genetic backgrounds (Sewell and Neale, 2000). The challenges of developing and genotyping the large progeny arrays required to locate QTL accurately in outbred pedigrees, and of verifying these QTL in different environments and ages, are such that MAS has not yet been applied in any commercial tree breeding programme.

In one of the most intensive studies on applying MAS to date, and based on data from over 1 300 individuals for wood density, 4 400 individuals for wood diameter from a single pedigree and using selective genotyping of the 50 highest and lowest scoring individuals for density and 100 of each for diameter, Devey *et al.* (2004a) were able to validate (in the same pedigree) two out of 13 QTL for diameter and eight out of 27 QTL for wood density in *Pinus radiata*. The effect of each QTL ranged from 0.8 to 3.6 percent of phenotypic variation, implying that these traits were controlled by a large number of genes, each of small effect. Using a different approach, Brown *et al.* (2003) used a verification population of 447 progeny (derived from re-mating the parents of the QTL pedigree) and an "unrelated population" of 445 progeny from the base pedigree to verify QTL in *Pinus taeda*. They found about half the QTL were detected in multiple seasons and fewer QTL were common to different populations.

An area where QTL mapping may assist breeders is in breaking linkages between negatively correlated traits. For example in *E. grandis* and *E. urophylla,* Verhaegen *et al.* (1997) reported co-location of QTL for the negatively correlated traits of wood density and growth. If these traits are controlled by tightly linked genes, markers could be used to select favourable recombinants.

Most markers used in QTL mapping have been anonymous markers that are unlikely to occur in a gene influencing a quantitative trait. In an attempt to increase the power of QTL mapping, candidate genes that may control the trait in question are being used as molecular markers. Candidate genes are typically sourced from the tissue of interest (e.g. xylem or leaves) and have either a known function intuitively related to the trait, or are of interest from studies of their expression using DNA microarrays. Comparative mapping and candidate gene approaches can utilize such information to search for homologous genes in different genomes. Candidate genes have been mapped to QTL for wood quality in *E. urophylla* and *E. grandis* (Gion *et al.,* 2000), *Pinus taeda* (Neale, Sewell and Brown, 2002), and *E. globulus* (Thamarus *et al.,* 2004). They have also been mapped to QTL for bud set and bud flush in *Populus deltoides* (Frewen *et al.,* 2000).

Kirst *et al.* (2004) measured transcript abundance in 2 608 genes in the differentiating xylem of 91 *E. grandis* backcross progeny. QTL analysis of lignin-related transcripts (expressed gene QTL [eQTL]) showed that their mRNA abundance is regulated by two genetic loci. Coordinated down-regulation of genes encoding lignin enzymes was observed in fast growing individuals, indicating that the same genomic regions are regulating growth and the lignin content and composition in the progeny. Comparative mapping has shown that gene content and gene order are conserved over long chromosomal regions among related species. Comparative maps are therefore likely to play an important role in enabling information on gene location and function to be transferred between species and genera. However, this will depend on orthologous genetic markers being mapped in each species (Krutovskii and Neale, 2001). To date, comparative maps have been published for *Populus* (Cervera *et al.,* 2001), *Pinus* (Devey *et al.,* 1999; Krutovsky *et al.,* 2004), *Quercus* and *Castanea* (Barreneche *et al.,* 2004).

MAS, based on QTL, is most likely to be used for within-family selection in a limited number of elite families that can be propagated clonally for deployment in large-scale industrial plantations. It is most suitable for traits that are expensive to measure or can only be detected after plants have been subjected to a particular stress or pathogen, and that have poor juvenile-mature correlations. Limitations of the approach include the low resolution of the marker-trait associations, the low proportion of phenotypic variation explained by QTL (generally less than 10 percent), and the low success rate in validating QTL in different genetic backgrounds and environments (Sewell and Neale, 2002). Recombination-based methodologies have been applied to inbred crop lines to positionally clone genes of interest in QTL regions (Salvi *et al.,* 2002); however, the use of this technique in forest trees is not practicable due to their outcrossing breeding system.

POPULATION-BASED ASSOCIATION STUDIES

Limitations of the QTL approach have led to a change in research focus towards identifying variations in the DNA sequence of genes directly controlling phenotypic variation, known by some as GAS. One of the main advantages of association genetics is the high resolution of marker-trait associations. As natural populations are used in association studies, recombinations that accumulate over many generations of the population break any long range associations between marker and trait leaving short stretches of the genome associated with the trait. If alleles or SNPs can be found that are strongly associated with superior phenotypes, they can be used for selection across breeding populations. This methodology is better suited to tree breeding programmes, which aim to maintain a broad genetic base, i.e. programmes with a large number of families. In contrast, the QTL approach is used for within-family selection. Spurious associations may be observed in association studies where there is undetected genetic structure in the breeding population that invalidates standard statistical tests. Strategies for dealing with population stratification have been developed to avoid these problems (Pritchard *et al.,* 2000; Wu and Zeng, 2001).

In the first association study published in forest trees, Thumma *et al.* (2005) identified 25 common SNP markers in the lignin biosynthesis gene *CCR* from *Eucalyptus nitens*. Using single-marker

and haplotype analyses in 290 trees from a natural population, they observed two haplotypes that were significantly associated with microfibril angle, a major determinant of timber strength. These results were confirmed in a full-sib family in *E. nitens* and in the related species *E. globulus.* In a powerful demonstration of the resolution of association genetics, Thumma *et al.* (2005) detected an alternatively-spliced variant of the *CCR* gene from the region of the significant haplotype, thereby revealing the probable molecular basis of the trait variation.

Association mapping is a particularly useful approach when genes are available that are likely to be functionally relevant to the trait of interest. Homologues of genes characterized in model species such as *Arabidopsis*, maize or rice, and poplar are excellent targets for association studies in forest species. In most cases, putative orthologues can be identified by comparing ESTs to gene sequences in public databases. In some cases, gene function may be determined by modulating the expression of selected genes using sense and antisense modification to up- and down-regulate gene expression, or intron RNA hairpin constructs to silence genes (Smith *et al.,* 2000). However, one of the advantages of association studies is the capacity to study a large number of genes simultaneously without the need for transformation (Peter and Neale, 2004).

There is considerable interest in understanding the genes controlling wood fibre cell wall development in forest trees as fibre microstructure is a major determinant of the commercial value of wood. For example, the angle of orientation of cellulose microfibrils (MFA) in fibre cell walls is known to affect timber strength and stiffness as well as fibre collapsibility, an important determinant of tensile strength in paper. Knowledge of cell wall biosynthesis would also assist in understanding and manipulating the development of abnormal wood, e.g. tension/compression wood (see Paux *et al.,* 2005; Pavy *et al.,* 2005), which is known to have an impact on wood stability, sawing patterns and pulpability. Wood is primarily composed of secondary xylem, and its properties are the product of sequential developmental processes from cambial cell division and expansion, to secondary wall formation and lignification. Genes expressed during xylogenesis determine the physical and chemical properties of wood. Important genes are now being identified that control the synthesis of the major constituents of the cell wall: cellulose, hemicellulose and lignin. Genes for cellulose synthesis (*CesA*) have been cloned in aspen (Joshi, Wu and Chiang, 1999; Wu, Joshi and Chiang, 2000), poplar (Djerbi *et al.,* 2005) and loblolly pine (Nairn and Haselkorn, 2005). Characterizing the *CesA* gene in aspen revealed strong similarity with a secondary cell wall protein in cotton, indicating that they serve similar functions in the two evolutionarily divergent genera. Transformation of cellulose synthase genes in aspen (Joshi, 2004) should further elucidate gene function. Each of the three loblolly *CesA* genes is orthologous to one of the three angiosperm secondary cell wall *CesA*s, suggesting functional conservation between angiosperms and gymnosperms. A search of the poplar genome revealed 18 distinct *CesA* gene sequences in *Populus trichocarpa* (Djerbi *et al.,* 2005). The *CesA* genes belong to a superfamily of *CesA*-like (*Csl*) genes, which includes a very large number of glycosyltransferases that are likely to be involved in the synthesis of the numerous non-cellulosic polysaccharides in plants (Liepman, Wilkerson and

Keegstra, 2005).

Lignin biosynthesis is well understood at the molecular level in plants (reviewed by Boerjan, Ralph and Baucher, 2003 and Peter and Neale, 2004) and is of particular interest in forest trees as removal of lignin for paper-making has major economic and environmental costs. In some cases, genetic modification of lignin structure has been shown to improve delignification in pulp and paper-making (Jouanin and Goujon, 2004), and down regulation of lignin pathway enzymes has also been shown to increase cellulose content (Hu *et al.*, 1999). Gymnosperms and angiosperms share a common set of enzymes that are responsible for the formation of guaiacyl lignin, while angiosperms have evolved at least two enzymes that catalyse the production of syringyl lignin. Association studies are now being carried out in loblolly to examine the adaptive significance of sequence variation in monolignol biosynthetic genes (Peter and Neale, 2004) and other genes controlling wood properties (Brown *et al.*, 2004). Similar research (Table 3) aimed at identifying genes controlling wood formation is being undertaken in Douglas fir (Krutovsky *et al.*, 2005), maritime pine (Pot *et al.*, 2004), radiata pine (S.G. Southerton and G.F. Moran, personal communication), spruce (MacKay *et al.*, 2005) and eucalypts (Moran *et al.*, 2002).

The availability of genes linked to a range of other traits in model plants opens up new areas of investigation in association genetics. For example, association studies are in progress to identify genes controlling pathogen resistance (Ersoz *et al.*, 2004; MacKay *et al.*, 2005), drought tolerance (Ersoz *et al.*, 2004), cold tolerance (Krutovsky *et al.*, 2005) and bud set (Paoli and Morgante, 2005) (Table 3). Further opportunities exist for association studies aimed at identifying SNPs linked to important traits. Flowering is particularly well understood at the molecular level (Zik and Irish, 2003), and increasing numbers of genes controlling flowering have been cloned in angiosperm tree species including eucalypts (Southerton *et al.*, 1998; Watson and Brill, 2004), silver birch (Elo *et al.*, 2001), poplar (Rottmann *et al.*, 2000) and gymnosperm tree species including spruce (Tandre *et al.*, 1995; Rutledge *et al.*,1998) and pines (Mouradov *et al.*, 1998, 1999).

Another important technological advance that is making large-scale association studies possible is the recent development of rapid, high-throughput

TABLE 3
Association studies in progress for forest tree species

Species	Trait	Reference
Eucalyptus nitens	Wood properties	Moran *et al.*, 2002; Thumma *et al.*, 2005
Populus	Wood properties	MacKay *et al.*, 2005
	Disease resistance	MacKay *et al.*, 2005
Picea glauca	Wood properties	MacKay *et al.*, 2005
	Disease resistance	MacKay *et al.*, 2005
Picea abies	Bud set	Paoli and Morgante, 2005
Pseudotsuga menziesii	Wood properties	Krutovsky *et al.*, 2005;
	Cold hardiness	Krutovsky *et al.*, 2005;
Pinus radiata	Wood properties	Southerton and Moran unpub. data
Pinus taeda	Wood properties	Peter and Neale, 2004; Brown *et al.*, 2004
	Disease resistance	Ersoz *et al.*, 2004
	Drought tolerance	Ersoz *et al.*, 2004
Pinus pinaster	Wood properties	Pot *et al.*, 2004

genotyping techniques that have drastically reduced the cost of genotyping SNPs in association populations (www.illumina.com/products/prod_snp.ilmn). It is now feasible to genotype SNPs in hundreds of genes potentially associated with a trait.

QTL mapping remains largely a research tool to improve our understanding of the number, distribution and mode of action of genes controlling quantitative traits. QTL can also play a role in GAS as a vehicle for validating significant SNP correlations identified in association populations (Thumma *et al.,* 2005). In the near future, association studies promise to yield numerous SNP markers that could be used in breeding programmes for early selection of superior alleles associated with a wide range of traits. As the efficiency of techniques for microarray analysis, SNP discovery, genotyping and other molecular procedures improve further, the opportunities to incorporate molecular technologies into breeding programmes for forest trees will increase.

USE OF MAS TO ENHANCE BREEDING PROGRAMMES IN DEVELOPING COUNTRIES

The adoption of molecular techniques varies widely, not only between developed and developing countries but also among the less developed economies (Chaix and Monteuuis, 2004). Countries such as China, India, Indonesia, Malaysia, Thailand and Viet Nam have established molecular laboratories for genotyping. Molecular markers are being used routinely to assess the level of genetic diversity in breeding programmes and to monitor any changes following selection (Butcher, 2003; Marcucci Poltri *et al.,* 2003). They are also being used to estimate levels of contamination and inbreeding in open-pollinated seed orchards (Chaix *et al.,* 2003; Harwood *et al.,* 2004), to validate intra- and interspecies crosses and to determine error rates in clonal propagation or trial establishment (see, for example, Bell *et al.,* 2004). This has identified relatively high error rates in several breeding programmes, affecting calculations of heritability, breeding value and genetic gain. Genetic linkage maps have been published for eucalypts in China (Gan *et al.,* 2003) and Brazil is prominent in eucalypt mapping and genomics (Grattapaglia, Chapter 14). EST libraries have been developed for mangroves in India as a first step towards characterizing genes associated with salinity tolerance (Mehta *et al.,* 2005), while DNA markers have been used for backward selection to identify superior male parents in eucalypt seed orchards in Brazil (Grattapaglia *et al.,* 2004). This approach has some potential for accelerating the breeding cycle in open-pollinated breeding programmes, particularly with species that are difficult to hand pollinate (Butcher, Moran and DeCroocq, 1998). The application of QTL-MAS in developing countries remains limited, exceptions being selection of coconut parents for breeding (FAO, 2003) and identification of QTL for rubber tree improvement (Lespinasse *et al.,* 2000). More widespread application may depend on economic considerations. Reports on the financial viability of MAS differ, with Johnson, Wheeler and Strauss (2000) indicating that large areas would need to be planted with MAS-improved germplasm to justify initial investment, while Wilcox *et al.* (2001) suggest significant financial gains are possible even when selection is based on DNA markers linked to a few loci each of relatively small effect. Association mapping has the potential for more widespread application in developing countries due, in part, to the ability to transfer markers among individuals, regardless of pedigree

or family relationships. The possibility of transferring SNP markers among species has already been demonstrated in eucalypts (Thumma *et al.,* 2005).

Forest trees, including many species in developing countries, are near their wild state, and significant improvement can still be made quite rapidly based on selection among existing genotypes (FAO, 2003). Association studies are ideally suited to exploiting variation in natural populations and do not rely on the existence of extensive pedigrees from controlled crosses. Suitable populations should include a large number of unrelated individuals of the same age growing on the one site. It has been estimated that 500 individuals would be necessary to detect association between a quantitative nucleotide responsible for 5 percent or more of the phenotypic variance (Long and Langley, 1999). The development of such populations would provide a good foundation for future GAS research in developing countries. While the markers developed using this approach are likely to be more easily transferred between breeding programmes, the application of GAS would require the subsequent development of advanced breeding programmes where the selection of superior alleles could take place. However, publicly funded forestry research is suboptimal in many developing countries and development priorities do not necessarily include genetic improvement programmes (FAO, 2003).

The major costs of GAS are associated with identifying candidate genes potentially linked to the relevant traits, and discovering SNPs. In some cases, public databases may contain large numbers of genes from the target or closely related species but, if not, there would be additional costs associated with EST sequencing of genes from the relevant tissue (i.e. xylem genes for wood traits). These additional costs may be offset partially by EST sequencing clones from mixed cDNA libraries derived from a number of unrelated trees from the association population.

Previously, the cost of genotyping SNPs was prohibitive, but this has fallen dramatically in recent years as high-throughput technologies have been developed for the human HapMap project (International HapMap Consortium, 2003). The Illumina Beadstation technology (www.illumina.com) is particularly suited to smaller-scale genotyping projects such as those being undertaken in forest trees. Cost is certainly a limitation in many developing countries including much of Africa and some South American countries; however, in most Asian countries and countries such as Brazil where molecular genetic laboratories are already well established, costs would not be prohibitive. The full benefits of GAS would require development of efficient clonal propagation and deployment systems before it was applied routinely.

Less stringent regulatory approval processes and greater public acceptance of genetically modified plants have allowed Brazil and China to take a lead role in commercializing transgenic tree technology. China is the only country to announce the commercial release of transgenics (poplar) with 300–500 hectares being planted in 2002 (Wang, 2004). Regulatory approval for the release of *Bacillus thuriengensis* (Bt) insect resistant eucalypts in Brazil is pending (Sedjo, 2004). Given the difficulty of carrying out long-term transgenic field trials with long rotation conifers, transgenic approaches are likely to be restricted to modification of high-value traits such as wood fibre properties in short rotation species grown on a large scale. Conventional breeding, using either open or controlled

pollination in seed orchards, will continue to be the most important mechanism for developing new genotypes for increased genetic gain (Carson, Walter and Carson, 2004).

ACKNOWLEDGEMENTS

We thank Simon Potter and Reddy Thumma (ENSISjv) for constructive comments on the manuscript.

REFERENCES

Agrama, H.A., George, T.L. & Salah, S.F. 2002. Construction of genome map for *Eucalyptus camaldulensis* Dehn. *Silv. Genet*. 51: 201–206.

***Arabidopsis* Genome Initiative.** 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Arcade, A., Anselin, F., Faivre Rampant, P., Lesage, M.C., Paques, L.E. & Prat, D. 2000. Application of AFLP, RAPD and ISSR markers to genetic mapping of European and Japanese larch. *Theor. Appl. Genet*. 100: 299–307.

Barreneche, T., Bodénès, C., Lexer, C., Trontin, J.F., Fluch, S., Streiff, R., Plomion, C., Roussel, G., Steinkellner, H., Burg, K., Favre, J.M., Glössl, J. & Kremer, A. 1998. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme, and 5S rDNA markers. *Theor. Appl. Genet.* 97: 1090–1103.

Barreneche, T., Casosoli, M., Russell, K., Akkak, A., Meddour, H., Plomion, C., Villai, F. & Kremer, A. 2004. Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). *Theor. Appl. Genet.* 108: 558–566.

Bell, J.C., Powell, M., Devey, M.E. & Moran, G.F. 2004. DNA profiling, pedigree lineage analysis and monitoring in the Australian breeding program of radiata pine. *Silv. Genet*. 53: 130–134.

Binelli, G. & Buchi, G. 1994. A genetic linkage map of *Picea abies* based on RAPD markers, as a tool in population genetics. *Theor. Appl. Genet.* 88: 283–288.

Boerjan, W., Ralph, J. & Baucher, M. 2003. Lignin biosynthesis. *Ann. Rev. Plant Biol*. 54: 519–46.

Brendel, O., Pot, D., Plomion, C., Rozenberg, P. & Guehl, J.M. 2002. Genetic parameters and QTL analysis of d13C and ring width in maritime pine. *Plant Cell Environ.* 25: 945–953.

Brown, G.R., Bassoni, D.L., Gill, G.P., Fontana, J.R., Wheeler, N.C., McGraw, R.A., Davis, M.F., Sewell, M.M., Tuskan, G.A. & Neale, D.B. 2003. Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.) III. QTL verification and candidate gene mapping. *Genetics* 164: 1537–1546.

Brown, G.R., Gill, G.P., Kuntz, R.J., Beal, J.A., Nelson, D., Wheeler, N.C., Penttila, B., Roers, J. & Neale, D.B. 2004. Associations of candidate gene single nucleotide polymorphisms with wood property phenotypes in loblolly pine, Abstract W98. In *Plant & Animal Genomes XII Conference,* San Diego, CA, USA.

Bundock, P.C., Hayden, M. & Vaillancourt, R. 2000. Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. *Silv. Genet*. 49: 223–232.

Butcher, P.A. 2003. Molecular breeding of tropical trees. *In* A. Rimbawanto, & M. Susanto, eds. *Proc. Internat. Conf. on Advances in Genetic Improvement of Tropical Tree Species,* pp. 59–66. Centre for Forest Biotechnology and Tree Improvement, Yogyakarta, Indonesia.

Butcher, P.A. 2004. Genetic mapping in acacias. *In* S. Kumar & M. Fladung, eds. *Molecular genetics and breeding of forest trees*, pp. 411–427. New York, USA, Haworth Press.

Butcher, P.A. & Moran, G.F. 2000. Genetic linkage mapping in *Acacia mangium* 2. Development of an integrated map from two outbred pedigrees using RFLP and microsatellite loci. *Theor. Appl. Genet.* 101: 594–605.

Butcher, P.A., Moran, G.F. & DeCroocq, S. 1998. Use of molecular markers in domestication and breeding programs for acacias. *In* J.W.Turnbull, H.R. Crompton, & K. Pinyopusarek, eds. *Recent developments in acacia planting*, pp. 211–215. ACIAR Proc. No. 82, Canberra. ACIAR.

Byrne, M., Murrell, J.C., Allen, B. & Moran, G.F. 1995. An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. *Theor. Appl. Genet*. 91: 869–875.

Byrne, M., Murrell, J.C., Owen, J.V., Kriedemann, P., Williams, E.R. & Moran, G.F. 1997a. Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. *Theor. Appl. Genet*. 94: 674–681.

Byrne, M., Murrell, J.C., Owen, J.V., Williams, E.R. & Moran, G.F. 1997b Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. *Theor. Appl. Genet*. 95: 975–979.

Carson, M., Walter, C. & Carson, S. 2004. The future of forest biotechnology. *In* R. Kellison, S. McCord, & K.M.A Gartland, eds. *Forest biotechnology in Latin America*, pp. 13–38. Proceedings, Workshop Biotecnolegia Forestal, Biotechnology Forum, Conceptión, Chile (available at www.forestbiotech.org/pdffiles/ChlePDFfinal31Jan2005.pdf).

Cervera, M.T., Sewell, M.M., Rampant, P.-F., Storme, V. & Boejan, W. 2004. Genome mapping in *Populus. In* S. Kumar, & M. Fladung, eds. *Molecular genetics and breeding of forest trees*, pp. 387–410. New York, USA, Haworth Press.

Cervera, M.T., Stormel, V., Ivens, B., Gusmão, J., Liu, B.H., Hostyn, V., Van Slyckenc, J., Van Montagu, M. & Boerjan, W. 2001. Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa* based on AFLP and microsatellite markers. *Genetics* 158: 787–809.

Chagne, D., Lalanne, C., Madur, D., Kumar, S., Frigerio, J.-M., Krier, C., Decroocq, S., Savoure, Al, Bou-Dagher-Karrat, M., Bertocchi, E., Brach, J. & Plomion, C. 2002. A high density linkage map for maritime pine based on AFLPs. *Ann. For. Sci*. 59: 627–636.

Chaix, G., Gerber, S., Razafimaharo, V., Vigneron, P., Verhaegen, D. & Hamon, S. 2003. Gene flow estimation in a Malagasy seed orchard of *Eucalyptus grandis*. *Theor. Appl. Genet.* 107: 705–712.

Chaix, G. & Monteuuis, O. 2004. Biotechnology in the forestry sector. In *Preliminary review of biotechnology in forestry including genetic modification*. FAO Forest Genetic Resources Working Paper FGR/59E. Rome (available at www.fao.org/docrep/008/ae574e/ae574e00.htm).

Costa, P., Pot, D., Dubos, C., Frigerio, J.M., Pionneau, C., Bodenes, C., Bertocchi, E., Cervera, M.-T., Remington, D.L. & Plomion, C. 2000. A genetic map of Maritime pine based on AFLP, RAPD and protein markers. *Theor. Appl. Genet*. 100: 39–48.

Devey, M.E., Fiddler, T.A, Liu, B.-H., Knapp, J. & Neale, D.B. 1994. An RFLP linkage map for loblolly pine based on a three-generation pedigree. *Theor. Appl. Genet*. 88: 273–278.

Devey, M.E., Bell, J.C., Smith, D.N., Neale, D.B. & Moran, G.F. 1996. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD and microsatellite markers. *Theor. Appl. Genet*. 92: 673–679.

Devey, M.E., Sewell, M.M., Uren, T.L. & Neale, D. 1999. Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. *Theor. Appl. Genet*. 99: 656–662.

Devey, M.E., Carson, S.D., Nolan, M.F., Matheson, A.C., Te Riini, C. & Hohepa, J. 2004a. QTL association for density and diameter in *Pinus radiata* and the potential for marker-aided selection. *Theor. Appl. Genet*. 108: 516–524.

Devey, M.E., Groom, K.A., Nolan, M.F., Bell, J.C., Dudzinski, M.J., Old, K.M., Matheson, A.C. & Moran, G.F. 2004b. Detection and verification of quantitative trait loci for resistance to *Dothistroma* needle blight in *Pinus radiata*. *Theor. Appl. Genet*. 108: 1056–1063.

Djerbi, S., Lindskog, M., Arvestad, L., Sterky, F. & Teeri, T.T. 2005. The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes. *Planta* 221: 739–746.

Eldridge, K.G., Davidson, J., Harwood, C.E. & van Wyk, G. 1993. *Eucalypt Domestication and Breeding*. Oxford, Clarendon Press.

Elo, A., Lemmetyinen, J., Turunen, M.-L., Tikka, T. & Sopanen, T. 2001. Three MADS-box genes similar to *APETALA* 1 and *FRUITFULL* from silver birch (*Betula pendula*). *Physiol. Plant* 112: 95–103.

Emebiri, L.C., Devey, M.E., Matheson, A.C. & Slee, M.U. 1997. Linkage of RAPD markers to NESTUR, a stem growth index in radiata pine seedlings. *Theor. Appl. Genet.* 95: 119–124.

Emebiri, L.C., Devey, M.E., Matheson, A.C. & Slee, M.U. 1998a. Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theor. Appl. Genet.* 97: 1053–1061.

Emebiri, L.C., Devey, M.E., Matheson, A.C. & Slee, M.U. 1998b. Interval mapping of quantitative trait loci affecting NESTUR, a stem growth efficiency index of radiata pine seedlings. *Theor. Appl. Genet.* 97: 1062–1068.

Ersoz, E.S., Gonzalez-Martinez, S., Gill, G., Brown G, Morse, A., Davis, J., White, T. & Neale, D.B. 2004. SNP discovery in candidate genes for disease resistance (pitch canker and fusiform rust) and drought tolerance in loblolly pine, Abstract W95. In *Plant & Animal Genomes XII Conference*, San Diego, CA, USA.

FAO. FAOSTAT interactive database (available at http://faostat.fao.org/default.aspx).

FAO. 2003. *Molecular marker-assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries*. Summary Document to Conference 10 (17 November to 14 December 2003) of the FAO Biotechnology Forum (available at www.fao.org/biotech/logs/C10/summary.htm).

Frewen, B.E., Chen, T.H.H., Howe, G.T., Davis, J., Rohde, A., Boerjan, W. & Bradshaw, H.D.Jr. 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* 154: 837–845.

Gan, S., Shi, J., Li, M., Wu, K., Wu, J. & Bai, J. 2003. Moderate density molecular maps of *Eucalyptus urophylla* S.T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. *Genetica* 118: 59–67.

Gion, J.-M., Rech, P., Grima-Pettenati, J., Verhaegen, D. & Plomion, C. 2000. Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Mol. Breed.* 6: 441–449.

Gosselin, I., Zhou, Y., Bousquet, J. & Isabel, N. 2002. Megagametophyte-derived linkage maps of white spruce (*Picea glauca*) based on RAPD, SCAR and ESTP markers. *Theor. Appl. Genet.* 104: 987–997.

Grattapaglia, D. & Sederoff, R.R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.

Grattapaglia, D., Bertolucci, L.G. & Sederoff, R.R. 1995. Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. *Theor. Appl. Genet*. 90: 933–947.

Grattapaglia, D., Bertolucci, L.G., Penchel, R. & Sederoff, R.R. 1996. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 144: 1205–1214.

Grattapaglia, D., Ribeiro, V.J. & Rezende, G.D.S.P. 2004. Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. *Theor. Appl. Genet*. 109: 192–199.

Groover, A., Devey, M., Fiddler, T., Lee, J., Megraw, R., Mitchel-Olds, T., Sherman, B., Vujcic, S., Williams, C. & Neale, D. 1994. Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* 138: 1293–1300.

Hanley, S., Barker, J.H.A., van Oijen, J.W., Aldam, C., Harris, S.L., Ahman, I., Larsson, S. & Karp, A. 2002. A genetic linkage map of willow (*Salix viminalis*) based on AFLP and microsatellite markers. *Theor. Appl. Genet*. 105: 1087–1096.

Harwood, C.E., Ha Huy Thinh, Tran Ho Quang, Butcher, P.A. & Williams, E.R. 2004. The effect of inbreeding on early growth of *Acacia mangium* in Vietnam. *Silv. Genet*. 53: 65–69

Hayashi, E., Kondo, T., Terada, K., Kuramoto, N. Goto, Y., Okamura, M. & Kawasaki, H. 2001. Linkage map of Japanese black pine based on AFLP and RAPD markers including markers linked to resistance against the pine needle gall midge. *Theor. Appl. Genet*. 102: 871–875.

Hertzberg, M., Aspeborg, H., Schrader, J., Andersson, A., Erlandsson, R., Blomqvist, K., Bhalerao, R., Uhlén, M., Teeri, T.T., Lundeberg, J., Sundberg, B., Nilsson, P. & Sandberg, G. 2001. A transcriptional roadmap to wood formation. *Proc. Nat Acad. Sci. USA* 98: 14732–14737.

Hu, W.-J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.-J. & Chiang, V.L. 1999. Repression of lignin biosynthesis promotes cellulose accumulation and growth. *Nature Biotech.* 17: 808–812.

Hurme, P., Sillanpaa, M.J., Arjas, E., Repo, T. & Savolainen, O. 2000. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics* 156: 1309–1322.

International HapMap Consortium. 2003. The International HapMap Project. *Nature* 426: 789–796.

Jermstad, K.D., Bassoni, D.L., Jech, K.S., Ritchie, G.A., Wheeler, N.C. & Neale, D.B. 2003. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. QTL by environment interactions. *Genetics* 165: 1489–1506.

Jermstad, K.D., Bassoni, D.L., Jech, K.S., Wheeler, N.C. & Neale, D.B. 2001a. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. I. Timing of vegetative bud flush. *Theor. Appl. Genet*. 102: 1142–1151.

Jermstad, K.D., Bassoni, D.L., Wheeler, N.C., Anekonda, T.S., Aitken, S.N., Adams, W.T. & Neale, D.B. 2001b. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. II. Spring and fall cold hardiness. *Theor. Appl. Genet*. 102: 1152–1158.

Jermstad, K.D., Bassoni, D.L., Wheeler, N.C. & Neale, D.B. 1998. A sex-averaged linkage map in coastal Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) based on RFLP and RAPD markers. *Theor. Appl. Genet*. 97: 762–770.

Johnson, G.R., Wheeler, N.C. & Strauss, S.H. 2000. Financial feasibility of marker-aided selection in Douglas fir. *Can. J. For. Res.* 30: 1942–1952.

Joshi, C.P. 2004. Molecular genetics of cellulose biosynthesis in trees. *In* S. Kumar & M. Fladung, eds. *Molecular genetics and breeding of forest trees,* pp. 141–165. New York, USA, Haworth Press.

Joshi, C.P., Wu, L. & Chiang, V.L. 1999. A novel xylem specific and tension stress responsive cellulose synthase gene from quaking aspen, Abstract p.19. *In* M. Campbell, ed. *Forest Biotechnology 99.* Oxford, UK, Oxford University Press.

Jouanin, L. & Goujon, T. 2004. Tuning lignin metabolism through genetic engineering in trees. *In* S. Kumar & M. Fladung, eds. *Molecular genetics and breeding of forest trees,* pp. 167–191. New York, USA, Haworth Press.

Junghaus, D.T., Alfenas, A.C., Brommonschenkel, S.H., Oda,S., Mello, E.J. & Grattapaglia, D. 2003. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. *Theor. Appl. Genet*. 108: 175–180.

Kaya, Z., Sewell, M.M. & Neale, D.B. 1999. Identification of quantitative trait loci influencing annual height- and diameter-increment growth in loblolly pine. *Theor. Appl. Genet*. 98: 586–592.

Kha, L.D., Hai, N.D. & Vinh, H.Q. 1998. Clonal tests and propagation options for natural hybrids between *Acacia manguim* and *A. auriculiformis. In* J.W. Turnbull H.R. Crompton & K. Pinyopusarek, eds. *Recent developments in acacia planting,* pp. 203–210. ACIAR Proc. No. 82, Canberra, ACIAR.

Kirst, M., Myburg, A.A., De León, J.P.G., Kirst, M.E., Scott, J. & Sederoff, R. 2004. Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiol*. 135: 2368–2378.

Krutovskii, K.V. & Neale, D.B. 2001. Forest genomics for conserving adaptive genetic diversity. FAO Forest Genetic Resources Working Paper FGR/3 (E). Rome (available at www.fao.org/DOCREP/003/X6884E/X6884E00.HTM).

Krutovskii, K.V., Vollmer, S.S., Sorensen, F.C., Adams, W.T., Knapp, S.J. & Strauss, S.H. 1998. RAPD genome maps of Douglas fir. *J. Hered*. 89: 197–205.

Krutovsky, K.V., Troggio, M., Brown, G., Jermstad, K.D. & Neale, D.B. 2004. Comparative mapping in the Pinaceae. *Genetics* 168: 447–461.

Krutovsky, K.V., Pande, P., Jermstad, K.D., Howe, G.T., St. Clair, J.B., Wheeler, N.C. & Neale, D.B. 2005. Nucleotide diversity and linkage disequilibrium in adaptive trait related candidate genes in Douglas fir, Abstract W110. In *Plant & Animal Genomes XII Conference,* San Diego, CA, USA.

Kubisiak T.L., Nelson, C.D., Nance, W.L. & Stine, M. 1996. Comparison of RAPD linkage maps constructed for a single longleaf pine from both haploid and diploid mapping populations. *For. Genet*. 3: 203–211.

Kumar, S., Spelman, R.J., Garrick, D.J., Richardson, T.E., Lausberg, M. & Wilcox, P.L. 2000. Multiple marker mapping of wood density loci in an outbred pedigree of radiata pine. *Theor. Appl. Genet*. 100: 926–933.

Kuramoto, N., Kondo, T., Fujisawa, Y., Nakata, R., Hayashi, E. & Goto, Y. 2000. Detection of quantitative trait loci for wood strength in *Cryptomeria japonica*. *Can. J. For. Res.* 30: 1525–1533.

Lander, E.S. & Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Lerceteau, E., Plomion, C. & Andersson, B. 2000. AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study. *Mol. Breed.* 6: 451–458.

Lespinasse, D., Grivet, L., Troispoux, V., Rodier-Goud, M., Pinard, F. & Seguin, M. 2000. Identification of QTLs involved in resistance to South American leaf blight (*Microcyclus ulei*) in the rubber tree. *Theor. Appl. Genet*. 100: 975–984.

Lespinasse, D., Rodier-Goud, M., Grivet, L., Leconte, A., Legnate, H. & Seguin, M. 1999. A saturated linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite and isozyme markers. *Theor. Appl. Genet*. 100: 127–138.

Liepman, A.H., Wilkerson, C.G. & Keegstra, K. 2005. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proc. Nat. Acad. Sci. USA* 102: 2221–2226.

Long, A.D. & Langley, C.H. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Gen. Res.* 9: 720–731.

MacKay, J., Bousquet, J., Cooke, J., Isabel, N., Pavy, N., Regan, S., Retzel, E. & Séguin, A. 2005. Functional genomic investigations of wood formation and defense response in trees, and beyond, Abstract W101. In *Plant & Animal Genomes XII Conference*, San Diego, CA, USA.

Marcucci Poltri, S.N., Zelena, N., Rodriguez Traverso, J., Gelid, P. & Hopp, H.E. 2003. Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers. *Tree Physiol.* 23: 625–632.

Marques, C.M., Araújo, J.A., Ferreira, J.G., Whetten, R., O'Malley, D.M., Liu, B.-H. & Sederoff, R. 1998. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theor. Appl. Genet*. 96: 727–737.

Marques, C.M., Vasquez-Kool, J., Carocha, V.J., Ferreira, J.G., O'Malley, D.M., Liu, B.-H. & Sederoff, R. 1999. Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theor. Appl. Genet*. 99: 936–946.

Marques, C.M., Brondani, R.P.V., Grattapaglia, D. & Sederoff, R. 2002. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theor. Appl. Genet*. 105: 474–478.

Mehta, P., Sivaprakash, K., Parani, M., Venkataraman, G. & Parida, A.K. 2005. Generation and analysis of expressed sequence tags from the salt tolerant mangrove species *Avicennia marina* (Forsk.) Vierh. *Theor. Appl. Genet*. 110: 416–424.

Moran, G.F., Thamarus, K.A., Raymond, C.A., Qiu, D., Uren, T. & Southerton, S. 2002. Genomics of *Eucalyptus* wood traits. *Ann. For. Sci.* 59: 645–650.

Morgante, M. & Salamini, F. 2003. From plant genomics to breeding practice. *Curr. Opin. Biotech*. 14: 214–219.

Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S. & Teasdale, R.D. 1998. *NEEDLY*, a *Pinus radiata* ortholog of *FLORICAULA/LEAFY* genes, expressed in both reproductive and vegetative meristems. *Proc. Nat. Acad. Sci. USA* 95: 6537–6542.

Mouradov, A., Hamdorf, B., Teasdale, R.D., Kim, J.T. Winter, K.U. & Theißen, G.A. 1999. *DEF/GLO*-like MADS-Box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Dev. Genet.* 25: 245–252.

Mukai, Y., Suyama, Y., Tsumura, Y., Kawahara, T., Yoshimaru, H., Kondo, T., Tomaru, N., Kuramoto, N. & Murai, M. 1995. A linkage map for sugi (*Cryptomeria japonica*) based on RFLP, RAPD and isozyme loci. *Theor. Appl. Genet*. 90: 835–840.

Myburg, A.A., Griffin, A.R., Sederoff, R.R. & Whetten, R.W. 2003. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theor. Appl. Genet*. 107: 1028–1042.

Nairn, C.J. & Haselkorn, T. 2005. Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms. *New Phytologist* 166: 907–915.

Neale, D.B. & Savolainen, O. 2004. Association genetics of complex traits in conifers. *Trends Plant Sci.* 9: 325–330.

Neale, D.B., Sewell, M.M. & Brown, G.R. 2002. Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. *Ann. For. Sci.* 59: 595–605.

Nelson, C.D., Nance, W.L. & Doudrick, R.L. 1993. A partial genetic linkage map of Slash pine (*Pinus elliottii* Englem var. *elliottii*) based on random amplified polymorphic DNAs. *Theor. Appl. Genet.* 87: 145–151.

Nikles, D.G. 1996. The first 50 years of the evolution of forest tree improvement in Queensland. *In* M.J. Dieters, A.C. Matheson, D.G. Nikles, C.E. Harwood, & S.M. Walker, eds. *Tree improvement for sustainable tropical forestry*, pp. 51–64. Proc. QFRI-IUFRO Conference, Caloundra, Queensland, Australia.

Paglia, G.P., Olivieri, A.M. & Morgante, M. 1998. Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K.). *Mol. Gen. Genet*. 258: 466–478.

Paoli, E.D. & Morgante, M. 2005. Nucleotide diversity and linkage disequilibrium in a set of candidate genes for time of bud set in *Picea abies,* Abstract W106. In *Plant & Animal Genomes XII Conference*, San Diego, CA, USA.

Paux, E., Tamasloukht, M., Ladouce, N., Sivadon, P. & Grima-Pettenati, J. 2004. Identification of genes preferentially expressed during wood formation in *Eucalyptus. Plant Mol. Biol.* 55: 263–280.

Paux, E., Carocha, V., Marques, C., Mendes de Sousa, A., Borralho, N., Sivadon, P. & Grima-Pettenati, J. 2005. Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytol*. 167: 89–100.

Pavy, N., Laroche, J., Bousquet, J. & MacKay, J. 2005. Large-scale statistical analysis of secondary xylem ESTs in pine. *Plant Mol. Biol.* 57: 203–224.

Peter, G. & Neale, D. 2004. Molecular basis for the evolution of xylem lignification. *Curr. Opin. Plant Biol.* 7: 737–742.

Plomion, C., O'Malley, D.M. & Durel, C.E. 1995. Genomic analysis in maritime pine (*Pinus pinaster*). Comparison of two RAPD maps using selfed and open-pollinated seeds of the same individual. *Theor. Appl. Genet*. 90: 1028–1034.

Plomion, C., Bahrman, N., Costa, P., Dubos, C., Frigerio, J.-M., Lalanne, C., Madur, D., Pionneau, C. & Gerber, S. 2004. Proteomics for genetic and physiological studies in forest trees: application in maritime pine. *In* S. Kumar & M. Fladung, eds. *Molecular genetics and breeding of forest trees*, pp. 53–79. New York, USA, Haworth Press.

Poke, F.S., Vaillancourt, R.E., Potts, B.M. & Reid, J.B. 2005. Genomic research in *Eucalyptus*. *Genetica* 125: 79–101.

Pot, D., Eveno, E., Garnier-Géré, P. & Plomion, C. 2004. Association studies for wood properties in maritime pine, Abstract W88. In *Plant & Animal Genomes XII Conference*, San Diego, CA, USA.

Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. 2000. Association mapping in structured populations. *Am. J. Hum. Genet.* 67: 170–181.

Remington, D.L., Whetten, R.W., Liu, B.-H. & O'Malley, D.M. 1999. Construction of an AFLP genetic map with nearly complete coverage in *Pinus taeda*. *Theor. Appl. Genet*. 98: 1279–1292.

Rönnberg-Wästljung, A.C., Glynn, C. & Weih, M. 2005. QTL analysis of drought tolerance and growth for a *Salix dasyclados* x *Salix viminalis* hybrid in contrasting water regimes. *Theor. Appl. Genet*. 110: 537–549.

Rottmann, W.H., Meilan, R., Sheppard, L.A. Brunner, A.M., Skinner, J.S., Ma, C., Cheng, S., Jouanin, L., Pilate, G. & Strauss, S.H. 2000. Diverse effects of overexpression of *LEAFY* and *PTFL*, a poplar (*Populus*) homolog of *LEAFY/FLORICAULA*, in transgenic poplar and *Arabidopsis. Plant J.* 22: 235–245.

Rutledge, R., Regan, S., Nicolas, O., Fobert, P., Côté, C., Bosnich, W., Kauffeldt, C., Sunohara, G., Séguin, A. & Stewart, D. 1998. Characterization of an *AGAMOUS* homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis. Plant J.* 15: 625–634.

Saintagne, C.-S., Bodenes, C., Barreneche, T., Bertocchi, E., Plomion, C. & Kremer, A. 2004. Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. *Theor. Appl. Genet.* 109: 1648–1659.

Salvi, S., Tuberosa, R., Chiapparrino, E., Maccaferri, M., Veillet, S., van Beuningen, L. Isaac, P., Edwards, K. & Phillips, R.L. 2002. Towards positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol*. 48: 601–613.

Scalfi, M., Troggio, M., Piovani, P., Leonardi, S., Magnaschi, G., Vendramin, G.G. & Menozzi, P. 2004. A RAPD, AFLP and SSR linkage map, and QTL analysis in European beech (*Fagus sylvatica* L.). *Theor. Appl. Genet.* 108: 433–441.

Sedjo, R.A. 2004. Transgenic trees and trade: problems on the horizon? *Resources* 155: 8–13.

Sewell, M.M. & Neale, D.B. 2000. Mapping quantitative traits in forest trees. *In* S.M. Jain & S.C. Minocha, eds. *Molecular biology of woody plants*, Vol. 1, pp. 407–423. Dordrecht, Netherlands, Kluwer.

Sewell, M.M., Sherman, B.K. & Neale, D.B. 1999. A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* 151: 321–330.

Sewell, M.M., Bassoni, D.L., Megraw, R.A., Wheeler, N.C. & Neale, D.B. 2000. Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.) I. Physical wood properties. *Theor. Appl. Genet.* 101: 1273–1281.

Sewell, M.M., Davis, M.S., Tuskan, G.A., Wheeler, N.C., Elam, C.C., Bassoni, D.L. & Neale, D.B. 2002. Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.) II. Chemical wood properties. *Theor. Appl. Genet.* 104: 214–222.

Shena, M., Shalon, D., Davis, R.W. & Brown, P.O. 1996. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470.

Shepherd, M., Chaparro, J.X. & Teasdale, R. 1999. Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. *Theor. Appl. Genet*. 99: 1207–1215.

Shepherd, M., Cross, M., Dieters, M.J. & Henry, R. 2003. Genetic maps for *Pinus elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* using AFLP and microsatellite markers. *Theor. Appl. Genet*. 106: 1409–1419.

Skov, E. & Wellendorf, H. 1998. A partial linkage map of *Picea abies* clone V6470 based on recombination of RAPD-markers in haploid megagametophytes. *Silv. Genet*. 47: 273–282.

Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G. & Waterhouse, P.M. 2000. Total silencing by intron-spliced hairpin RNAs. *Nature* 407: 319–320.

Southerton, S.G., Strauss, S.H., Olive, M.R., Harcourt, R.L., Decroocq, V., Zhu, X., Llewellyn, D.J., Peacock, W.J. & Dennis, E.S. 1998. *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY. Plant Mol. Biol.* 37: 897–910.

Strabala, T.J. 2004. Expressed sequence tag databases from forestry tree species. *In* S. Kumar & M. Fladung, eds. *Molecular genetics and breeding of forest trees,* pp. 19–51. New York, USA, Haworth Press.

Strauss, S.H., Lande, R. & Namkoong, G. 1992. Limitations of molecular-marker-aided selection in forest tree breeding. *Can. J. For. Res.* 22: 1050–1061.

Strauss, S.H., Brunner, A.M., DiFazio, S.P. & James, R.R. 1998. Environmental effects of genetically engineered woody biomass crops. *Biomass & Bioenergy* 14: 403–414.

Strauss, S.H., Brunner, A.M., Busov, V.B., Caiping, M.A. & Meilon R. 2004. Ten lessons from 15 years of transgenic *Populus* research. *Forestry* 77: 455–465.

Tandre, K., Albert, V.A., Sundas, A. & Engstrom, P. 1995. Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* 27: 69–78.

Thamarus, K.A., Groom, K., Murrell, J., Byrne, M. & Moran G.F. 2002. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. *Theor. Appl. Genet*. 104: 379–387.

Thamarus, K.A., Groom, K., Bradley, A., Raymond, C.A., Shimleck, L.R., Williams, E.R. & Moran G.F. 2004. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theor. Appl. Genet*. 109: 856–864.

Thumma, B.R., Nolan, M.F., Evans R. & Moran G.F. 2005. Polymorphisms in *cinnamoyl CoA reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265.

Travis, S.E., Ritland, K., Whitham, T.G. & Keim, P. 1998. A genetic linkage map of pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. *Theor. Appl. Genet*. 97: 871–880.

Tsarouhas, V., Gullberg, U. & Lagercrantz, U. 2002. An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. *Theor. Appl. Genet.* 105: 277–288.

Tsarouhas, V., Gullberg, U. & Lagercrantz, U. 2003. Mapping of quantitative trait loci controlling timing of bud flush in *Salix*. *Hereditas* 138: 172–178.

van Zyl, L., Bozhkov, P.V., Clapham, D.H., Sederoff, R.R. & von Arnold, S. 2003. Up, down and up again is a signal global gene expression pattern at the beginning of gymnosperm embryogenesis. *Gene Expr. Patterns* 3: 83–91.

Verhaegen, D. & Plomion, C. 1996. Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. *Genome* 39: 1051–1061.

Verhaegen, D., Plomion, C., Gion, J.-M., Poitel, M., Costa, P. & Kremer, A. 1997. Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers: 1.Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor. Appl. Genet,* 95: 597–608.

Voiblet, C., Duplessis, S., Encelot, N. & Martin, F. 2001. Identification of symbiosis-regulated genes in *Eucalyptus globulus – Pisolithus tinctorius* ectormycorrhiza by differential hybridization of arrayed cDNAs. *Plant J.* 25: 181–191.

Walter, C. & Killerby, S. 2004. A global study on the state of forest tree genetic modification. In *Preliminary review of biotechnology in forestry including genetic modification*. FAO Forest Genetic Resources Working Paper FGR/59E. Rome (available at www.fao.org/docrep/008/ae574e/ae574e00.htm).

Wang, H. 2004. The state of genetically modified trees in China. In *Preliminary review of biotechnology in forestry, including genetic modification*, pp. 96–105. FAO Forest Genetic Resources Working Paper FGR/59E. Rome (available at www.fao.org/docrep/008/ae574e/ae574e00.htm).

Watson, J.M. & Brill, E.M. 2004. *Eucalyptus grandis* has at least two functional *SOC 1*-like floral activator genes. *Funct. Plant Biol.*, 31: 225–234.

Weng, C., Kubisiak, T.L., Nelson, C.D. & Stine, M. 2002. Mapping quantitative trait loci controlling early growth in a (longleaf pine x slash pine) x slash pine BC1 family. *Theor. Appl. Genet.* 104: 852–859.

Whetten, R.W., Sun, Y.H., Zhang, Y. & Sederoff, R. 2001. Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Mol. Biol.* 47: 275–291.

Wilcox, P.L., Carson, S.D., Richardson, T.E., Ball, R.D., Horgan, G.P. & Carter, P. 2001. Benefit-cost analysis of DNA marker-based selection in progenies of *Pinus radiata* seed orchard parents. *Can. J. For. Res*. 31: 2213–2224.

Wu, R. & Zeng, Z-B. 2001. Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* **157:** 899–909.

Wu, H.X., Abarquez, A. & Matheson, A.C. 2004. Inbreeding in *Pinus radiata* V. The effects of inbreeding on fecundity. *Silv. Genet*. 53: 80–87.

Wu, L., Joshi, C.P. & Chiang, V.L. 2000. A xylem specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. *Plant. J.* 22: 495–502.

Yazdani, R., Nilsson, J.E., Plomion, C. & Mathur, G. 2003. Marker trait association for cold acclimation and growth rhythm in *Pinus sylvestris*. *Scand. J. For. Res*. 18: 29–38.

Yazdani, R., Yeh, F.C. & Rimsha, J. 1995. Genomic mapping of *Pinus sylvestris* (L.) using random amplified polymorphic markers. *For. Genet*. 2: 109–116.

Yin, T.M., DiFazio, S.P., Gunter, L.E., Riemenshneider, D. & Tuskan, G.A. 2004. Large scale heterospecific segregation distortion in *Populus* revealed by a dense genetic linkage map. *Theor. Appl. Genet*. 109: 451–463.

Yoshimaru, H., Ohba, K., Tsurumi, K., Tomaru, N., Murai, M., Mukai, Y., Suyama, Y., Tsumura, Y., Kawahara, T. & Sakamaki, Y. 1998. Detection of quantitative trait loci for juvenile growth, flower bearing and rooting ability based on a linkage map of sugi (*Cryptomeria japonica*). *Theor. Appl. Genet*. 97: 45–50.

Yu, J. Hu, S., Wang, J., Wong, G.K-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, C., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X, Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, O., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, O., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou,C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L. & Yang, H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79–92.

Zik, M. & Irish, V.F. 2003. Flower development: initiation, differentiation, and diversification. *Annu. Rev. Cell Dev. Biol.* 19: 119–140.

SECTION V

Marker-assisted selection in fish – case studies

CHAPTER 16

Possibilities for marker-assisted selection in aquaculture breeding schemes

Anna K. Sonesson

SUMMARY

FAO estimates that there are around 200 species in aquaculture. However, only a few species have ongoing selective breeding programmes. Marker-assisted selection (MAS) is not used in any aquaculture breeding scheme today. The aim of this chapter, therefore, is to review briefly the current status of aquaculture breeding schemes and to evaluate the possibilities for MAS of aquaculture species. Genetic marker maps have been published for some species in culture. The marker density of these maps is, in general, rather low and the maps are composed of many amplified fragment length polymorphism (AFLP) markers anchored to few microsatellites. Some quantitative trait loci (QTL) have been identified for economically important traits, but they are not yet mapped at a high density. Computer simulations of within-family MAS schemes show a very high increase in genetic gain compared with conventional family-based breeding schemes, mainly due to the large family sizes that are typical for aquaculture breeding schemes. The use of genetic markers to identify individuals and their implications for breeding schemes with control of inbreeding are discussed.

INTRODUCTION

Aquaculture is an expanding industry, with a total global value of US$61 billion (FAO, 2003). FAO estimates that there are around 200 species in culture, of which carps and oysters have the largest worldwide production. However, only a few species have ongoing selective breeding programmes.

MAS is not used in any aquaculture breeding scheme today. The aim of this chapter, therefore, is to review briefly the current status of aquaculture breeding schemes and to evaluate the possibilities for MAS of aquaculture species.

TRAITS OF BREEDING INTEREST

Growth rate is the most important trait for most aquaculture species under selection. It is recorded on the selection candidates, and can easily be improved using mass selection. Sexual maturation is a trait that leads to reduced growth, reduced feed conversion efficiency and reduced fillet quality in several aquaculture species. Therefore, selection against early maturation is often performed, i.e. the individuals that become sexually mature before market size are discarded as selection candidates. In tilapia, late maturity is desirable because of excessive spawning that results in overcrowding of ponds and reduced size of the fish.

For many other traits, however, accurate measurement techniques for live individuals are inadequate. Hence, selection must be based on information from other family members, e.g. on siblings. MAS would be especially valuable for traits that are difficult and/or costly to measure on the selection candidate or for traits that are measured late in life or at slaughter (Lande and Thompson, 1990; Poompuang and Hallerman, 1997). Examples of these important traits are:

- *Disease resistance*. Challenge tests exist for viral (e.g. white spot syndrome in shrimps and infectious pancreatic necrosis in most marine fishes) and bacterial (e.g. furunculosis and vibriosis) diseases, as well as for parasites (e.g. sea lice). When challenge tests are used in breeding programmes, however, surviving individuals cannot enter the breeding nucleus because of the risk that they will introduce the disease to the nucleus. Therefore, these individuals cannot be selection candidates and only their sibs, who have no records for these traits, are candidates.
- *Fillet quality traits*. To this group of traits belong colour, texture, gaping, different fat-related traits (e.g. fat percentage and distribution) and dressing percentage. Accurate measurements of these traits are available only for slaughtered individuals. Techniques for measuring fillet colour on live fish are under development.
- *Feed conversion efficiency* is a trait that can be measured practically only at the family level at a young age in the breeding nucleus, but not at the individual level or in grow-out operations. The value of such early records is rather limited because of the unknown correlation with feed conversion efficiency at a later age. Feed intake is, in general, a difficult trait to measure in aquaculture species due to unequal feed intake over days. No active selection programme for aquaculture species selects directly for feed conversion efficiency; rather, indirect selection is practised by selecting for growth.
- *Salinity and low temperature tolerance* are two traits of interest for tilapia breeding programmes. Today, tilapias are produced in freshwater in tropical and subtropical areas. The purpose of selecting for salinity and temperature tolerance

is to develop fish that could reproduce and grow in areas of higher salinity and lower temperature, i.e. to increase the production area for tilapias. Although these traits could be measured early in the life of the fish and therefore could be selected, sexually mature fish may respond differently to the temperature and salinity conditions.

Gjedrem and Olesen (2005) provide a more complete list of aquaculturally relevant traits and their heritabilities and correlations.

STRUCTURE OF BREEDING SCHEMES

Most aquaculture species are currently bred in mono- or polyculture systems (i.e. with one or several species reared together) using mass selection for growth rate. Only about 30 family-based breeding programmes worldwide utilize sib information in the estimation of breeding values (B. Gjerde, personal communication). The main part of a family-based breeding programme is a closed breeding nucleus, with trait information from sibs coming from test stations. Breeding programmes for species with limited reproductive ability (e.g. salmonids as opposed to several highly fecund marine species such as Atlantic cod) have a multiplier unit, where genetic material from the nucleus is used to produce eggs or fry for the grow-out producers. The limiting factor for the breeding nucleus is often the number of tanks, where the fry of a certain full-sib family are kept until individuals are large enough to be physically tagged. The number of offspring per full-sib family is large, such that a very high selection intensity can be achieved. Generally, each male is mated to two females in order that the tank effect can be estimated separately from the additive genetic effects.

The high intensity of selection within the nucleus can easily result in high rates of inbreeding. Introduction of unrelated wild stock is often practised to reduce the rates of inbreeding. However, introduction of wild stock also leads to reduced genetic gain, and should generally be avoided in ongoing breeding schemes. Optimum contribution is an approach that maximizes genetic gain while restricting the rates of inbreeding for schemes with discrete (Meuwissen, 1997; Grundy, Villanueva and Woolliams, 1998) or overlapping (Meuwissen and Sonesson, 1998; Grundy, Villanueva and Woolliams, 2000) generation structures or for traits with a polygenic effect and the effect of a known single gene (Meuwissen and Sonesson, 2004). The key determination is the number of matings (full-sib families) per selected individual. One practical constraint in some marine species is that matings are volitional (natural mating in, for example, a tank) and thus depend on the availability of individuals ready to spawn at a certain moment. Hence, for these species, the number of matings per male or female is restricted for each spawning. The use of frozen milt makes the use of males more flexible. Milt from many species including salmonids, carp and shrimps can be frozen (Stoss, 1983), but the practical use of cryopreserved sperm in aquaculture breeding programmes has not been fully utilized.

GENETIC MARKER MAPS

A genetic marker map is an ordered listing of the genes or molecular markers occurring along the length of the chromosomes in the genome. Distances between genes or markers are estimated in terms of how frequently recombination occurs between them. Genetic marker maps are available for some aquaculture species (Table 1). Most of these genetic maps are constructed using amplified fragment length polymorphism

(AFLP) markers (Vos *et al.*, 1995), which are generally anchored to a smaller collection of microsatellites. The marker density of the maps is currently rather low, and the markers are spread unevenly over the genome, which may explain why the number of linkage groups found in, for example, channel catfish (Waldbieser *et al.*, 2001) or white shrimp (Pérez *et al.*, 2004) does not correspond to the number of chromosomes. In rainbow trout, tetraploidy has been found for 20 chromosome arms (Sakamoto *et al.*, 2000). Recombination rates can differ between males and females, and hence marker map lengths can differ considerably between males and females. In salmonids, females have the higher recombination rate, which implies that most information comes from the females when constructing the marker map. The ratio between female and male recombination rates was 3.25:1.00 for rainbow trout (Sakamoto *et al.*, 2000), 1.69:1.00 for Arctic char (Woram *et al.*, 2004) and 8.25:1.00 for Atlantic salmon (Moen *et al.*, 2004a). However, in other aquaculture species, males have the higher recombination rate. For example, the ratio between male and female recombination rates was 7.4:1.00 in Japanese flounder (Coimbra *et al.*, 2003). There are also differences in recombination rate over the length of the chromosomes in males, i.e. recombination rate was higher in telomeric regions than in proximal regions in rainbow trout (Sakamoto *et al.*, 2000).

TABLE 1
Aquaculture species for which there are genetic marker maps

Species	Reference
Scallop	Li *et al.* (2005)
	Wang *et al.* (2004)
Pacific oyster	Hubert and Hedgecock (2004)
Eastern oyster	Yu and Guo (2003)
White shrimp	Pérez *et al.* (2004)
Kuruma prawn	Li *et al.* (2003)
Black tiger shrimp	Wilson *et al.* (2002)
Kuruma prawn	Moore *et al.* (1999)
Atlantic salmon	Moen *et al.* (2004a)
	Gilbey *et al.* (2004)
Arctic char	Woram *et al.* (2004)
Rainbow trout	Nichols *et al.* (2003)
	Sakamoto *et al.* (2000)
	Young *et al.* (1998)
Salmonids	May and Johnson (1990)
Common carp	Sun and Liang (2004)
European sea bass	Chistiakov *et al.* (2005)
Channel catfish	Waldbieser *et al.* (2001)
	Liu *et al.* (2003)
Tilapia	Lee *et al.* (2005)
	McConnell *et al.* (2000)
	Agresti *et al.* (2000)
	Kocher *et al.* (1998)
Japanese flounder	Coimbra *et al.* (2003)

MAPPING OF QTL

QTL are loci whose variability underlies variation in expression of a quantitative character (Geldermann, 1975). Detection of QTL would help in understanding the genetic architecture of the trait, i.e. the numbers and relative effects of genes that determine expression of a trait. A small, but growing, number of QTL for important traits have been identified in farmed aquatic species (Table 2). In tilapias, QTL have been identified for cold tolerance (Cnaani *et al.*, 2003; Moen *et al.*, 2004b). QTL for upper temperature tolerance (Jackson *et al.*, 1998; Danzmann, Jackson and Ferguson, 1999; Perry, Ferguson and Danzmann, 2003; Somorjai, Danzmann and Ferguson 2003), and for resistance to different disease traits have been found in salmonids, e.g. for infectious hematopoietic necrosis virus (Rodriguez *et al.*, 2005), infectious pancreatic necrosis virus (Ozaki *et al.*, 2001), *Ceratomyxa shasta* (Nichols, Bartholomew and Thorgaard, 2003) and infectious salmon anemia (Moen *et al.*, 2004c, 2006). QTL for general disease resistance and immune response have been found in tilapias (Cnaani *et al.*, 2004) and

TABLE 2
Known marker-QTL linkages in aquaculture species

Trait	Reference
Salmonids	
Spawning time	Leder, Danzmann and Ferguson (2006)
Early development	Martinez *et al.* (2005)
Pyloric caeca number	Zimmerman *et al.* (2005)
Natural killer cell-like activity	Zimmerman *et al.* (2004)
Hematopoetic necrosis resistance	Rodriguez *et al.* (2005)
Development rate	Sundin *et al.* (2005)
Infectious salmon anemia resistance	Moen *et al.* (2004c, 2006)
Ceratomyxa shasta resistance	Nichols, Bartholomew and Thorgaard (2003)
Infectious pancreatic necrosis resistance	Ozaki *et al.* (2001)
Infectious hematopoietic necrosis resistance	Khoo *et al.* (2004)
Body weight and condition factor	Reid *et al.* (2005)
Spawning date and body weight	O'Malley *et al.* (2003)
Growth and maturation	Martyniuk *et al.* (2003)
Temperature tolerance	Somorjai, Danzmann and Ferguson (2003)
Meristic traits	Nichols, Wheeler and Thorgaard (2004)
Embryonic development	Robison *et al.* (2001)
Albinism	Nakamura *et al.* (2001)
Development rate	Nichols *et al.* (2000)
Spawning time	Sakamoto *et al.* (1999)
Upper temperature tolerance, size	Perry *et al.* (2001, 2003)
Upper temperature tolerance	Danzmann, Jackson and Ferguson (1999)
Upper temperature tolerance	Jackson *et al.* (1998)
Tilapia	
Cold tolerance	Moen *et al.* (2004b)
Cold tolerance and fish size	Cnaani *et al.* (2003)
Stress and immune response	Cnaani *et al.* (2004)
Colour pattern	Streelman, Albertson and Kocher (2003)
Early survival	Palti *et al.* (2002)
Sex determination	Lee, Penman and Kocher (2003); Lee, Hulata and Kocher (2004)

salmonids (Zimmerman *et al.*, 2004). The data used for quantifying disease resistance and temperature tolerance traits are often based on challenge tests, for which models accounting for non-normality of data and special algorithms that take account of censored data (survival models) are used in combination with the QTL mapping methods (e.g. Moen *et al.*, 2006). In salmonids, QTL have been found related to body weight and size (Martyniuk *et al.*, 2003; O'Malley *et al.*, 2003; Reid *et al.*, 2005), for colouration pattern (Streelman, Albertson and Kocher, 2003) and for one form of albinism (Nakamura *et al.*, 2001). Zimmerman *et al.* (2005) found QTL for pyloric caeca number, a trait related to feed conversion efficiency. Epistasis has been found for upper temperature tolerance and body length in rainbow trout (Danzmann, Jackson and Ferguson, 1999; Perry, Ferguson and Danzmann, 2003); the epistasis depended on the genetic background, which would result in reduced effectiveness of MAS (Danzmann, Jackson and Ferguson, 1999).

The genetic diversity of wild and cultured populations, high fecundity, and the possibility of interspecific hybridization and reproductive manipulation of aquaculture species can be exploited in QTL mapping studies. These features have resulted in a

wide diversity of experimental populations being used in such studies. Double haploids (see below) have been used for QTL mapping in salmonids (by androgenesis; Robison *et al.,* 2001; Zimmerman *et al.,* 2005) and tilapias (by gynogenesis; Palti *et al*., 2002). Backcross populations have been set up where strains with large phenotypic differences in the trait of interest are crossed, e.g. for temperature tolerance in rainbow trout (Danzmann, Jackson and Ferguson, 1999). The strains can come from one species, but crosses between species also have been used (Streelman, Albertson and Kocher, 2003 for tilapia; Rodriguez *et al.,* 2005 for salmonids). Finally, families derived from a breeding nucleus have been used for QTL mapping. (e.g. Moen *et al*., 2006 for Atlantic salmon).

Most QTL mapping is based on marker association studies (e.g. Sakamoto *et al*., 1999) or on a marker association study followed by interval mapping (Moen *et al.,* 2006). Combined linkage/linkage disequilibrium methods (e.g. Meuwissen *et al*., 2002) have high precision for mapping QTL in outbred populations, but have not been used for QTL mapping in aquaculture species. This lack could be explained by the sparcity of genetic marker maps for the species under study and by many of the studies having used special crosses as mentioned above, where linkage is the main source of information for mapping the QTL.

Various reproductive manipulations may be applied to aquaculture species. One interesting reproductive manipulation technique for outbred populations for marker and QTL mapping is gyno- and androgenetic double haploids (Chourrout, 1984). In gynogenesis, the sperm's chromosomes are inactivated by irradiation and following fertilization, diploidy is restored by applying a temperature or hydrostatic pressure shock. The result is an individual that is double haploid with only the female's chromosomes. Depending on when diploidy is restored, two types of double haploid individuals can be produced. If an early shock is applied, extrusion of the second polar body is suppressed and, when the two maternal chromosome sets fuse, some heterozygosity is retained. If a late shock is applied, the first mitotic cleavage of the zygote is suppressed and, when the two maternal chromosome sets fuse, the resulting individual is virtually homozygous. In androgenesis, the egg is irradiated and, after "fertilization", the egg is shocked to suppress the first mitosis (Parsons and Thorgaard, 1984). The result is an individual that is a double haploid with only the male's chromosomes and that is virtually homozygous.

The power of an experiment to detect QTL depends on the effect of QTL alleles, the recombination rate among the marker and QTL loci, and the sample size of the mapping population. The effect of QTL genotypes is higher for double haploid than for full- or half-sib family designs in an outbred population because the QTL genotypes occur only in a homozygous form in double haploids (i.e. in their most extreme form). The relative advantage of a population of mitotic double haploids, where the two chromosome sets are copies of each other, is largest when the QTL has a small effect (Martinez, Hill and Knott, 2002). In meiotic haploid individuals, the two chromosome sets in an individual have been recombined. The power of QTL detection in these meiotic double haploid populations is therefore expected to resemble that of selfed populations (Martinez, Hill and Knott, 2002). Double haploids have been used for genetic marker and QTL mapping, as noted above.

On the other hand, the extremely large full-sib family size that is possible for aquaculture species may make use of special reproductive techniques for marker and QTL mapping studies unnecessary. For example, in Atlantic cod, both males and females produce millions of gametes at spawning. Also, Atlantic cod and halibut are examples of repeat spawners, in contrast to salmonids that normally die after a single spawning. Repeated spawning makes the mating structure more flexible, e.g. a certain pairing can be repeated or an individual may be used in multiple pairings. Another disadvantage of using double haploids is that they are fully inbred individuals and may therefore express the trait of interest differently than non-inbred animals. An example of this is the environmental variance for wing length in *Drosophila melanogaster*, which has been shown to be larger for inbred individuals than non-inbred individuals (Falconer and Mackay, 1996). Traits with dominant expression may be expressed differently because of inbreeding depression.

MAS SCHEMES

After QTL detection experiments, breeders will have knowledge of marker-QTL linkages and an estimate of the respective effects of QTL alleles on the trait in the population. This knowledge may be applied using MAS of spawners for producing the next generation. Generally, QTL detection would be carried out in one experiment and MAS in another (Poompuang and Hallerman, 1997). For within-family MAS schemes, the phase of marker and QTL alleles needs to be established for all families on which selection will be practised. Two within-family MAS schemes have been well studied for livestock populations. The first is a three-generation scheme, which is suitable for breeding schemes with progeny testing (Kashi, Hallerman and Solomon, 1990). Progeny tests are not, however, currently performed upon aquaculture species. The second is a two-generation scheme (Mackinnon and Georges, 1998), which may be modified for application to fish populations.

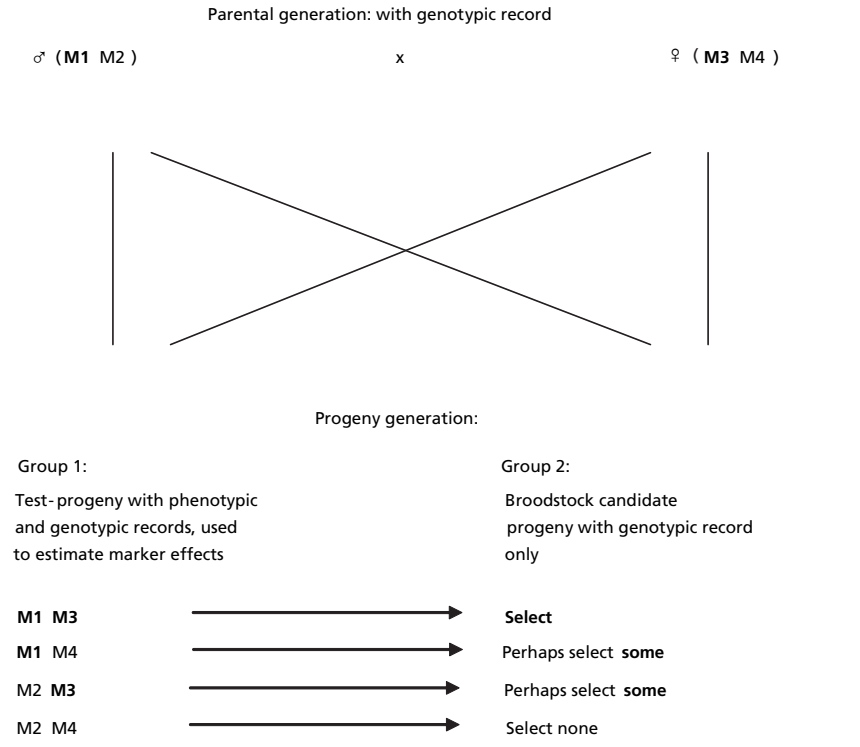
In the two-generation scheme (Figure 1), it is assumed that both sires and dams have genotypic records for markers linked to the QTL and that there are two groups of progeny from these parents. The group of test progeny has both phenotypic performance and genotypic records, while the group of selection candidates only has genotypic records. Phenotypic evaluation often implies that these individuals cannot be used later for breeding, perhaps because they were used in a challenge test or were slaughtered to obtain sib information for carcass traits. The genetic markers of the sire are denoted M1 and M2, and those of the dam M3 and M4, with M1 and M3 linked with the performance-increasing QTL alleles. With these linkage relationships, M1M3-bearing progeny would be selected while some M1M4- and M2M3-, and no M2M4-bearing progeny would be selected. The proportions of each genotype selected would vary depending upon selection intensity.

When QTL are mapped densely (up to 5 cM between markers), both linkage disequilibrium within families and population-wide linkage disequilibrium can be used in the MAS scheme (Smith and Smith, 1993; Dekkers and Hospital, 2002).

Simulation of two-generation within-family MAS schemes

The attractive feature of MAS is the potential for increasing the genetic gain in a selective breeding programme. Lande and

FIGURE 1
Two-generation marker evaluation and selection scheme for aquaculture species



Individuals with marker genotypes shown in **bold** are selected. The proportion of each genotype selected will vary with selection intensity.

Thompson (1990) showed that the efficiency of MAS relative to conventional selection alone depends upon the heritability of the trait under selection, the proportion of genetic variance associated with marker loci and the particular selection scheme at issue.

Stochastic simulations of two-generation within-family MAS schemes in a typical aquaculture breeding programme were carried out by Sonesson (2006) for selection for one trait. Selection was truncated based upon best linear unbiased prediction (BLUP) breeding values including information of the genetic marker (Fernando and Grossman, 1989). The heritability, h^2, of the trait under selection was 0.06 or 0.12, and 20 percent of the genetic variance was accounted for by the QTL.

Genetic gain was 0.202 for MAS and 0.176 for conventional breeding, after selection in generation 1 (Table 3), i.e. MAS had 15 percent higher genetic gain than conventional breeding. After selection in generation 2, genetic gain was 68 percent higher for MAS than conventional breeding. The performance-increasing QTL allele was then almost fixed (i.e. its frequency approached 1). The increase in genetic gain is mainly due to the increased frequency of the positive QTL allele for the MAS scheme, where a higher QTL frequency implies more

TABLE 3
Genetic gain in schemes with heritability (h^2) of 0.06 or 0.12

	Generation number			
	1	2	3	4
		$h^2 = 0.06$		
Conventional	0.176	0.121	0.134	0.117
MAS	0.202	0.203	0.135	0.114
		$h^2 = 0.12$		
Conventional	0.337	0.206	0.196	0.191
MAS	0.361	0.318	0.206	0.177

genetic variance (as long as the frequency of the positive allele is less than 0.5). For situations with a higher h^2 of 0.12, genetic gain was 7 percent higher after selection in generation 1 and 54 percent higher after selection in generation 2, i.e. the superiority of MAS was somewhat lower for schemes with the higher heritability.

IDENTIFICATION OF INDIVIDUALS USING GENETIC MARKERS

In traditional family-based breeding programmes, individuals from the same full-sib families are reared separately (e.g. in tanks) until they are large enough to be tagged physically. This mode of rearing is very costly, and the number of full-sib families therefore limits the size of the breeding nucleus. In addition, separate rearing of full-sib families results in common environmental (tank) effects that need to be estimated, which in turn affects other parts of the design and analysis of the breeding programme. That is, in the mating design, a sire has to be mated to several dams (or *vice versa*) in order to be able to separate analytically these common environmental effects from additive genetic effects. The tank effect is generally higher for newly domesticated species, where feed and other environmental effects are not yet standardized. For example, the tank effect was 3–12 percent for juvenile Atlantic cod (Gjerde *et al.*, 2004), and the nursery pond effect was 32 percent for rohu carp (Gjerde *et al.*, 2003) compared with, for example, a tank effect of 2–6 percent for Atlantic salmon and rainbow trout (Rye and Mao, 1998; Pante *et al.*, 2002). Were it possible to pool progeny groups into one tank, tank effects would be eliminated, a smaller number of tanks would be needed per spawner and more pairs could be spawned.

Parentage assignment using molecular markers is useful for tracing pedigrees in breeding programmes, and can be used to identify parents in breeding schemes where progeny groups are pooled. Parentage assignment implies that parents and offspring are all genotyped for a number of genetic markers that are well spread over the genome and that the information on genotypes is used to assign individual progeny to the correct parental pair. The parent-offspring relationship is such that each offspring inherits one allele from each parent, making it possible to exclude possible parents when this condition is not fulfilled.

Exclusion of parents is the basic method of assigning parents. The exclusion probability per locus (E_l) can be calculated according to the formulae of Hanset (1975) and Dodds *et al.* (1996). The global exclusion probability over loci (E_g) is:

$$E_g = 1 - \prod_{l=1}^{l=L} (1 - E_l)$$

where L is the total number of marker loci screened. Genotyping errors can result in the true parents being excluded, because uncertainty is not accounted for with this method. Even a low error rate reduces the correct assignment rate, which increases with the number of loci and number of alleles. SanCristobal and Chevalet (1997) derived a likelihood and a Bayesian-based method that could take account of genotyping errors. They used the likelihood method on data for 50 parental pairs and their offspring and a genotyping error rate of 2 percent, but with no error rate accounted for in the likelihood calculation. The correct parentage assignment rate was only 88 percent using a system of five loci with five alleles, and 83 percent using a system of eight loci with five alleles. After including a genotyping error rate of 10^{-3} in the same calculations, the correct assignment rate increased to nearly 100 percent.

The number of loci and alleles per loci needed for correct parent assignment was estimated deterministically and validated stochastically by Villanueva, Verspoor and Visscher (2002). They showed that nine loci with five alleles per locus or six loci with ten alleles would assign 99 percent of offspring to the correct parents from 100 or 400 crosses. Similar results were found by Bernatchez and Duchesne (2000). These results agree well with the results from empirical studies of aquaculture populations. For example, Herbinger *et al.* (1995) assigned 66 percent of the offspring to correct parental pairs in a complete factorial cross between ten males x ten females (i.e. 100 parental pairs) of rainbow trout using four loci with four to ten alleles per locus. Perez-Enriquez, Takagi and Taniguchi (1999) reported 73 percent correct assignment of parental pairs using five microsatellites when 7 800 parental pairs were possible for a population of red sea bream. Norris, Bradley and Cunningham (2000) assigned over 95 percent of the parental pairs correctly using eight polymorphic loci with 10–29 alleles per locus in Atlantic salmon when the number of possible parental pairs was over 12 000. Jackson, Martin-Robichaud and Reith (2003) assigned 96–100 percent of the progeny to the correct parental pair in F_1 Atlantic halibut populations using five microsatellite loci. Castro *et al.* (2004) assigned over 99 percent of all parental pairs correctly with six microsatellites for 176 full-sib families of turbot. Vandeputte *et al.* (2004) assigned 95.3 percent of all parental pairs in a complete factorial cross of 10 dams x 24 sires of common carp using eight microsatellites. In addition to assessing parentage, full- and half-sib relationships have also been estimated using genetic markers in aquaculture stocks, e.g. Atlantic salmon (Norris, Bradley and Cunningham, 2000) and rainbow trout (McDonald, Danzmann and Ferguson, 2004).

There are underlying assumptions that affect experimental power for assigning parents. Some of these assumptions are:

- Hardy-Weinberg equilibrium (HWE). Small effective population sizes, non-random mating and unequal family size will lead to deviations from HWE. HWE is not an issue in assigning parental pairs to offspring, but affects the assignment of offspring genotypes to the parents (Estoup *et al.*, 1998), i.e. some genotypes of parents might be more difficult than others from which to assign offspring. If these genotypes are present in large families, parental assignment rate will be reduced relative to what would occur if they are present in small families.
- Zero mutation rate and no scoring errors. Castro *et al.* (2004) reported a mutation

rate of 6.7×10^{-4} when 13 464 gametes and marker information from 12 loci were screened in a turbot population. Mutations and scoring error have the same effect on excluding potential parents, giving rise to incorrect assignments.

- Unlinked loci and linkage equilibrium. Linkage and linkage disequilibrium between the loci will reduce the effective number of loci used for the parental assignment. Note that the power to assign parental pairs correctly can thereby differ between different sets of microsatellites. Estoup *et al.* (1998) quantified the difference in the power of microsatellite marker sets used to assign parents correctly in turbot (eight loci, eight alleles per loci) and rainbow trout (eight loci, four alleles per loci) populations by calculating the frequency of good and unique parent assignments (f_{gu}). The more variable set of microsatellites resulted in higher f_{gu} for larger maximal mating schemes for turbot than the less variable set of microsatellites for rainbow trout. In general, a set of loci with an equal number of alleles has the highest exclusion probability (Weir, 1996; Jamieson and Taylor, 1997).

WALK-BACK SELECTION SCHEMES

Practical breeding schemes using molecular marker-based parental assignment have been reported. Doyle and Herbinger (1994) proposed carrying out parentage assignment for individuals using genetic markers, such that full-sib families would not have to be kept separately until tagging but, rather, would be held in one large tank. Note that individuals that are genotyped also need to be physically tagged, so that genotyping results can be traced back to particular individuals. At the time for selection, fish in the tank were first ranked on their phenotypic value, assuming that selection was for only one trait that could be measured on the selection candidates (e.g. weight). Then the individual with the highest phenotypic value was selected and genotyped for family identification. Thereafter, the individual with the second highest phenotypic value was genotyped and selected if it was not a full- or half-sib of other, already-selected individuals, such that within-family selection was performed. This procedure continued until the desired number of brood stock was selected. This approach of progressing through the performance ranking, genotyping and selecting the best-performing individuals within families was termed "walk-back" selection. Matings subsequently would be made between families, a strategy preferred because it would keep the rate of inbreeding low (Falconer and Mackay, 1996). Herbinger *et al.* (1995) reported setting up a rainbow trout breeding programme using the walk-back selection programme of Doyle and Herbinger (1994) and genetic markers to estimate full/half relationships among the candidates using a likelihood ratio method and thereafter performing within-family selection.

Using stochastic simulations, Sonesson (2005) studied a combination of optimum contribution selection and walk-back selection. Optimum contribution is a selection method that maximizes genetic gain with a restriction on the rate of inbreeding (see earlier in this chapter). Hence, the combination of optimum contribution and walk-back selection ensures that the rate of inbreeding is under control, while the genetic gain is higher than in the within-family selection schemes of Doyle and Herbinger (1994) because selection is both within and between families. In the study by Sonesson (2005), batches of candidates were pre-selected from a single tank on

their phenotypic values and the batch size varied from 50 to all (1 000, 5 000 or 10 000) candidates. Relatively small batches of fish were genotyped at any one time to minimize genotyping costs. Thereafter, BLUP breeding values (Henderson, 1984) were estimated and the optimum contributions of the candidates calculated using the method of Meuwissen (1997). If the constraint on the rate of inbreeding could not be achieved, another batch of fish was genotyped and included in the total number of candidates. Results showed that with a batch size of 100, 76–92 percent of the genetic gain was achieved compared with schemes where all 1 000, 5 000 or 10 000 fish were genotyped to provide candidates for the optimum contribution selection algorithm. Hence, high genetic gain was achieved at a fixed rate of inbreeding with low genotyping costs.

The main practical advantage of these marker-assisted breeding schemes is that expenses associated with separate rearing of full-sib families are not incurred, which decreases start-up and operational costs for the breeding scheme. The most important trait at the start of a breeding programme is probably growth, which can easily be measured on the candidate. Use of the optimum contribution selection algorithm keeps the rate of inbreeding under control, which is especially important in breeding programmes for aquaculture species where selection intensity can be very high due to the large family sizes. In combination with BLUP estimated breeding values, which have a high within-family correlation such that individuals with the highest breeding values will tend to come from only a few families, high rates of inbreeding can result (Sonesson, Gjerde and Meuwissen, 2005). However, there remain unsolved issues with the combined optimum contribution and walk-back selection method:

- Biased BLUP breeding values lead to a reduction in accuracy of selection, because not all selection candidates are included in the estimation of breeding values.
- Low and unequal survival of families may lead to reduced genetic variation and thus increased rate of inbreeding. However, the optimum contribution selection will correct for some of this loss by selecting from more families to keep the genetic base broader than when selecting only for the BLUP estimated breeding values. One option for reducing this problem of unequal and low survival is to pool an equal number of individuals from each family after the main period of early mortality is over. In general, it is possible to use more parents in these programmes compared with conventional family-based selection programmes, which could compensate for some of the loss of families contributing to the next generation due to low and unequal survival.
- Multitrait selection is probably the largest practical problem to solve. One alternative could be to base the pre-selection on only one or two traits that are inexpensive to measure on the candidate. Techniques for measuring more traits on live selection candidates are steadily evolving (e.g. fat content in Atlantic salmon, Solberg *et al*., 2003), such that the sib-testing system might be unnecessary for these traits in the future.

INTROGRESSION SCHEMES

Many fish breeding schemes have been started with a relatively narrow genetic base, selecting for only one or two traits in relatively few animals. However, because all farmed aquatic species still have wild ancestors, introgression of genes (i.e. identified genes or QTL) from these wild

ancestors into breeding populations is possible. However, one assumes that all other traits of the wild fish are unwanted in the breeding population, such that only the particular gene of interest should be introgressed, leaving the genome of the breeding population otherwise intact (Hospital and Charcosset, 1997). In white shrimp, for example, wild stocks have been found to have higher disease resistance but lower growth rates than stocks in culture. In this example then, only the genes for disease resistance should be introgressed (taking account of the possible effect of these genes on growth). The problem of actually implementing marker-assisted introgression in populations is to find the trait of interest in wild populations at a reasonable cost, and then to identify genes or marked QTL for the trait to be introgressed (Visscher, Haley and Thompson, 1992; Koudande *et al*., 2000). This is a costly and time-consuming process, especially for species with long generation intervals. Methods for simultaneous QTL mapping and introgression would be useful.

ACKNOWLEDGEMENTS

I am grateful to Bjarne Gjerde, Eric Hallerman and Thomas Moen for their help with the manuscript.

REFERENCES

Agresti, J.J., Seki, S., Cnaani, A., Poompuang, S., Hallerman, E.M., Umiel, N., Hulata, G., Gall, G.A.E. & May, B. 2000. Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. *Aquaculture* 185: 43–56.

Bernatchez, L. & Duchesne, P. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Can. J. Fish Aquat. Sci.* 57: 1–12.

Castro, J., Bouza, C., Presa, P., Pino-Querido, A., Riaza, A., Ferreiro, I., Sánchez, L. & Martínez, P. 2004. Potential sources of error in parentage assessment of turbot (*Scophthalamus maximus*) using microsatellite loci. *Aquaculture* 242: 119–135.

Chistiakov, D.A., Hellemans, B., Haley, C., Law, A.S., Tsigenopoulos, C.S., Kotulas, G., Bertotto, D., Libertini, A. & Volckaert, F.A.M. 2005. A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. *Genetics* 170: 1821–1826.

Chourrout, D. 1984. Pressure-induced retention of second polar body and suppression of first cleavage in rainbow trout: production of all-triploids, all-tetraploids, and heterozygous and homozygous gynogenetics. *Aquaculture* 36:111–126.

Cnaani, A., Hallerman, E.M., Ron, M., Weller, J.I., Indelman, M., Kashi, Y., Gall, G.A.E. & Hulata, G. 2003. Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F2 tilapia hybrid. *Aquaculture* 223: 117–128.

Cnaani, A., Zilberman, N., Tinman, S. & Hulata, G. 2004. Genome-scan analysis for quantitative traits in an F_2 tilapia hybrid. *Mol. Gen. Genomics* 272: 162–172.

Coimbra, M.R.M., Kabayashi, K., Koretsugu, S., Hasegawa, O., Ohara, E., Ozaki, S., Sakamoto, T., Naruse, K. & Okamoto, N. 2003. A genetic map of the Japanese flounder *Paralichthys olivaceus*. *Aquaculture* 220: 203–218.

Danzmann, R.G., Jackson, T.R. & Ferguson, M.M. 1999. Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. *Aquaculture* 173: 45–58.

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Revs. Genet.* 3: 22–32.

Dodds, K.G., Tate, M.L., McEwan, J.C. & Crawford, A.M. 1996. Exclusion probabilities for pedigree testing farm animals. *Theor. Appl. Genet*. 92: 966–975.

Doyle, R.W. & Herbinger, C.M. 1994. The use of DNA fingerprinting for high-intensity within-family selection in fish breeding. *Proc. 5th Wrld. Congr. Genetics Appl. Livest. Prodn.* pp. 364–371. Guelph, Ontario. Department of Animal and Poultry Science, University of Guelph.

Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P. & Guyomard, R. 1998 Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Can. J. Fish. Aquat. Sci.* 55: 715–725.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to quantitative genetics, 4th edition*. Harlow, UK, Longman Publishing.

FAO. 2003. FAOSTAT statistical database. (available at http://faostat.fao.org).

Fernando, R.L. & Grossman, M. 1989. Marker-assisted selection using best linear biased prediction. *Genet. Sel. Evol*. 21: 467–477.

Geldermann, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor. Appl. Genet*. 46: 319–330.

Gilbey, A., Verspoor, E., McLay, A. & Houlihan, D. 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Anim. Genet*. 35: 98–105.

Gjedrem, T. & Olesen, I. 2005. Basic statistical parameters. *In* T. Gjedrem, ed. *Selection and breeding programs in aquaculture*, pp. 45–72. Doordrecht, Netherlands, Springer.

Gjerde, B., Mahapatra, K.D., Reddy, P.V.G.K., Saha, J.N., Jana, R.K. & Rye, M. 2003. Genetic and phenotypic parameters for growth in rohu (*Labeo rohita*) in mono- and polyculture production systems. In *Proc. 8th Internat. Symp. Genet. in Aquaculture*, p. 85. Puerto Varas, Chile.

Gjerde, B., Terjesen, B.F., Barr, Y., Lein, I. & Thorland, I. 2004. Genetic variation for juvenile growth and survival in Atlantic cod (*Gadus morhua*). *Aquaculture* 236: 167–177.

Grundy, B., Villanueva, B. & Woolliams, J.A. 1998. Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet. Res*. 72: 159–168.

Grundy B., Villanueva B. & Woolliams J. 2000. Dynamic selection for maximizing response with constrained inbreeding in schemes with overlapping generations. *Anim. Sci*. 70 373–382.

Hanset, R. 1975. Probabilité d'exclusion de paternité et de monozygotie, probabilité de similtude. Généralisation à N alleles co-dominants. *Ann. Med. Vet*. 119: 71–80.

Henderson, C.R. 1984. *Applications of linear models in animal breeding.* Guelph, Ontario, Canada,Guelph Univ. Press.

Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M. & Cook, D. 1995. DNA fingerprint-based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137: 245–256.

Hubert, S. & Hedgecock, D. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* 168: 351–362.

Hospital, F. & Charcosset, A. 1997. Marker-assisted introgression of quantitative trait loci. *Genetics* 147: 1469–1485.

Jackson, T.R., Ferguson, M.M., Danzmann, R.G., Fishback, A.G., Ihssen, P.E., O'Connell, M. & Crease, T.J. 1998. Identification of two QTL influencing upper temperature tolerance in three rainbow trout (*Onchorhynchus mykiss*) half-sib families. *Heredity* 80: 143–151.

Jackson, T.R., Martin-Robichaud, D.J. & Reith, M.E. 2003. Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. *Aquaculture* 220: 245–259.

Jamieson, A. & Taylor, S.C.S. 1997. Comparisons of three probability formulae for parentage exclusion. *Anim. Genet*. 28: 397–400.

Kashi, Y., Hallerman, E. & Soller, M. 1990. Marker-assisted selection of candidate bulls for progeny testing programs. *Anim. Prod*. 51: 63–74.

Khoo, S.K., Ozaki, A., Nakamura, F., Arakawa, T., Ishimoto, S., Nikolov, R., Sakamoto, T., Akatsu, T., Mochuzuki, M., Denda, I. & Okamoto, N. 2004. Identification of a novel chromosomal region associated with infectious hematopoietic necrosis (IHN) resistance in rainbow trout, Oncorhynchus mykiss. *Fish Path*. 39: 95–101.

Kocher, T.D., Lee, W.J., Sobolewska, H., Penman, D. & McAndrew, B. 1998 A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* 148: 1225–1232.

Koudandé, O.D., Iraqi, F., Thomson, P.C., Teale, A.J. & van Arendonk, J.A.M. 2000. Strategies to optimize marker-assisted introgression of multiple linked QTL. *Mamm. Genome* 11:145–150.

Lande, R. & Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.

Leder, E.H., Danzmann, R.G. & Ferguson, M.M. 2006. The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *J. Heredity* 97: 74–80.

Lee, B.-Y., Penman, D.J. & Kocher, T.D. 2003. Identification of the sex-determining region in tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Anim. Genet*. 34: 379–383.

Lee, B.-Y., Hulata, G. & Kocher, T.D. 2004. Two unlinked loci control the sex of blue tilapia (*Oreochromis aureus*). *Heredity* 92: 543–549.

Lee, B.-Y., Lee, W.J., Streelman, T., Carleton, K.L., Howe, A.E., Hulata, G., Slettan, A., Stern, J.E., Terai, Y. & Kocher, T.D. 2005. A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* 170: 237–244.

Li, Y., Byrne, K., Miggiano, E., Whan, V., Moore, S., Keys, S., Crocos, P., Preston, N. & Lehnert, S. 2003. Genetic mapping of the kuruma prawn (*Penaeus japonicus*) using AFLP markers. *Aquaculture* 219: 143–156.

Li, L., Xiang, J., Liu, X., Zhang, Y., Dong, B. & Zhang, X. 2005. Construction of AFLP-based genetic linkage map for Zhikong scallop, *Chlamys farreri* Jones et Preston and mapping of sex-linked markers. *Aquaculture* 245: 63–73.

Liu, Z., Karsi, A., Li, P., Cao, D. & Dunham, R. 2003. An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* 165: 687–694.

Mackinnon, M.J. & Georges, M.A.J. 1998. Marker-assisted preselection of young dairy sires prior to progeny-testing. *Livest. Prod. Sci*. 54: 229–250.

Martinez, V.A., Hill, W.G. & Knott, S.A. 2002. On the use of double haploids for detecting QTL in outbred populations. *Heredity* 88: 423–431.

Martinez, V., Thorgaard, G., Robison, B. & Sillanpää, M.J. 2005. An application of Bayesian QTL mapping to early development in double haploid lines of rainbow trout including environmental effects. *Genet. Res. Camb*. 86: 209–221.

Martyniuk, C.J., Perry, G.M.L., Mogahadam, H.K., Ferguson, M.M. & Danzmann, R.G. 2003. The genetic architecture of correlations among growth-related traits and male age at maturation in rainbow trout. *J. Fish Biol.* 63: 746–764.

May, B. & Johnson, J.E. 1990. Composite linkage map of salmonid fishes. *In* S.J. O'Brien, ed. *Genetic maps,* 5th edition, pp. 4.151–4.159. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press.

McConnell, S.K.J., Beynon, C., Leamon, J. & Skibinski, D.O.F. 2000. Microsatellite marker-based genetic linkage maps of *Oreochromis aureus* and *O. niloticus* (Cichlidae): extensive linkage group segment homologies revealed. *Anim. Genet*. 31: 214–216.

McDonald, G.J., Danzmann, R.G. & Ferguson, M.M. 2004. Relatedness determination in the absence of pedigree information in three cultured strains of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 233: 65–78.

Meuwissen, T.H.E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci*. 75: 934–940.

Meuwissen, T.H.E. & Sonesson, A.K. 1998. Maximizing the response of selection with a predefined rate of inbreeding – overlapping generations. *J. Anim. Sci.* 76: 2575–2583.

Meuwissen, T.H.E. & Sonesson, A.K. 2004. Genotype-assisted optimum contribution selection to maximise selection response over a specified time period. *Genet. Res*. 84: 109–116.

Meuwissen, T.H.E., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M.E. 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161: 373–379.

Moen, T., Hoyheim, B., Munck, H. & Gomez-Raya, L. 2004a. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Anim. Genet.* 35: 81–92.

Moen, T., Agresti, J.J., Cnaani, A., Moses, H., Famula, T.R., Hulata, G., Gall, G.A.E. & May, B. 2004b. A genome scan of a four-way tilapia cross supports the existence of a quantitative trait locus for cold tolerance on linkage group 23. *Aquaculture* 35: 893–904.

Moen, T., Fjalestad, K.T., Munck, H. & Gomez-Raya, L. 2004c. A multistage testing strategy for detection of quantitative trait loci affecting disease resistance in Atlantic salmon. *Genetics* 167: 851–858.

Moen, T., Hayes, B., Sonesson, A.K., Lien, S., Høyheim, B., Munck, H. & Meuwissen, T.H.E. 2006. Mapping of a quantitative trait loci for resistance against infectious salmon anemia in Atlantic salmon (*Salmo salar*): comparing survival models with binary models. *Genetics*. Submitted.

Moore, S.S., Whan, V., Davis, G.P., Byrne, K., Hetzel, D.J.S. & Preston, N. 1999. The development and application of genetic markers for the kuruma prawn (*Penaeus japonicus*). *Aquaculture* 173: 19–32.

Nakamura, K., Ozaki, A., Akutsu, T., Iwai., K., Sakamoto, T., Toshizaki, G. & Okamoto, N. 2001. Genetic mapping of the dominant albino locus in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genom*. 265: 687–93.

Nichols, K.M., Bartholomew, J. & Thorgaard, G.H. 2003. Mapping multiple genetic loci associated with *Ceratomyxa shasta* resistance in *Oncorhynchus mykiss*. *Dis. Aquat. Org*. 56: 145–154.

Nichols, K.M., Wheeler, P.A. & Thorgaard, G.H. 2004. Quantitative trait loci analyses for meristic traits in *Oncorhynchus mykiss*. *Env. Biol. Fish*. 69: 317–331.

Nichols, K.M., Robison, B.D., Wheeler, P.A. & Thorgaard, G.H. 2000. Quantitative trait loci (QTL) associated with developmental rate in clonal *Oncorhynchus mykiss* strains. *Aquaculture* 209:233.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies, I., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A. & Thorgaard, G.H. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet*. 34: 102–115.

Norris, A.T., Bradley, D.G. & Cunningham, E.P. 2000. Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. *Aquaculture* 182: 73–83.

O'Malley, K.G., Sakamoto, T., Danzmann, R.G. & Ferguson, M.M. 2003. Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *J. Hered.* 94: 273–84.

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M.R.M., Akutsu, T. & Okamoto, N. 2001. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genom*. 265: 23–31.

Palti, Y., Shirak, A., Cnaani, A., Hulata, G., Avtalion, R.R. & Ron, M. 2002. Detection of genes with deleterious alleles in an inbred line of tilapia (*Oreochromis aureus*). *Aquaculture* 206: 151–164.

Pante, M.J.R., Gjerde, B., McMillan, I. & Misztal, I. 2002. Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 204: 383–392.

Parsons, J.E. & Thorgaard, G.H. 1984. Induced androgenesis in rainbow trout. *J. Exp. Zool*. 231: 407–412.

Pérez, F., Erazo, C., Zhinaula, M., Volckaert, F. & Calderon, J. 2004. A sex-specific linkage map of the white shrimp *Penaeus* (*Litopenaeus*) *vannamei* based on AFLP markers. *Aquaculture* 242: 105–118.

Perez-Enriquez, R., Takagi, M. & Taniguchi, N. 1999. Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. *Aquaculture* 173: 413–423.

Perry, G.M., Ferguson, M.M. & Danzmann, R.G. 2003. Effects of genetic sex and genomic background on epistasis in rainbow trout (*Oncorhynchus mykiss*). *Genetica* 119: 35–50.

Perry, G.M., Danzmann, R.G., Ferguson, M.M. & Gibson, J.P. 2001. Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity* 86: 333–341.

Poompuang, S. & Hallerman, E.M. 1997. Toward detection of quantitative trait loci and marker-assisted selection in fish. *Rev. Fisheries Sci*. 5: 253–277.

Reid, D.P., Szanto, A., Glebe, B., Danzmann, R.G. & Ferguson, M.M. 2005. QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity* 94: 166–172.

Robison, B.D., Wheeler, P.A., Sundin, K., Sikka, P. & Thorgaard, G.H. 2001. Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *J. Hered*. 92: 16–22.

Rodriguez, M.F., LaPatra, S., Williams, S., Famula, T. & May, B. 2005. Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* 241: 93–115.

Rye, M. & Mao, I.L. 1998. Nonadditive genetic effects and inbreeding depression for body weight in Atlantic salmon (*Salmo salar*). *Livest. Prodn. Sci*. 57: 15–22.

Sakamoto, T., Danzmann, R.G., Okamoto, N., Ferguson, M.M. & Ihssen, P.E. 1999. Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173: 33–43.

Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., Okamoto, N., Ferguson, M.M., Holm, L.-E., Guyomard, R. & Hoyheim, B. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155: 1331–1345.

SanCristobal, M. & Chevalet, C. 1997. Error tolerant parent identification from a finite set of individuals. *Genet. Res. Camb*. 70: 53–62.

Smith, C. & Smith, D.B. 1993. The need for close linkages in marker-assisted selection for economic merit in livestock. *Anim. Breed. Abstr*. 61: 197–204.

Solberg, C., Saugen, E., Swenson, L.P., Bruun, L. & Isaksson, T. 2003. Determination of fat in live farmed Atlantic salmon using non-invasive NIR techniques. *J. Sci. Food Agr*. 83: 692–696.

Somorjai, I.M., Danzmann, R.G. & Ferguson, M.M. 2003. Distribution of temperature tolerance quantitative trait loci in Arctic charr (*Salvelinus alpinus*) and inferred homologies in rainbow trout (*Oncorhynchus mykiss*). *Genetics* 165: 1443–1456.

Sonesson, A.K. 2005. A combination of walk-back and optimum contribution selection for fish – a simulation study. *Genet. Sel. Evol.* 37: 587–599.

Sonesson, A.K. 2006. Marker-assisted selection in breeding schemes for aquaculture species. *Genet. Sel. Evol*. (in press*).*

Sonesson, A.K., Gjerde, B. & Meuwissen, T.H.E. 2005. Truncation selection for BLUP-EBV and phenotypic values in fish breeding schemes. *Aquaculture* 243: 61–68.

Stoss, J. 1983. Fish gamete preservation and spermatozoan physiology. *In* W.S. Hoar, D.J. Randall, & E.M. Donaldson, eds. *Fish physiology, vol. 9, part B*. pp. 305–350. New York, NY, USA, Academic Press.

Streelman, J.T., Albertson, R.C. & Kocher, T.D. 2003. Genome mapping of the orange blotch colour pattern in cichlid fishes. *Molec. Ecol*. 12: 2465–2471.

Sun, X. & Liang, L. 2004. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* 238: 165–172.

Sundin, K., Brown, K.H., Drew, R.E., Nichols, K.M., Wheeler, P.A. & Thorgaard, G.H. 2005. Genetic analysis of development rate QTL in backcrosses of clonal rainbow trout. *Aquaculture* 247: 75–83.

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Rodina, M., Gela, D., Vallod, D., Chevassus, B. & Linhart, O. 2004. Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). *Aquaculture* 235: 233–236.

Villanueva, B., Verspoor, E. & Visscher, P.M. 2002. Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. *Anim. Genet*. 33: 33–41.

Visscher, P.M., Haley, C.S. & Thomson, R. 1992. Marker-assisted introgression in backcross breeding programs. *Genetics* 144: 1923–1932.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., vanden Lee, T., Hornes, M., Fritters, A., Pot, J., Paleman, J., Kuiper, M. & Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res*. 23: 4407–4414.

Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J. & Wolters, W.R. 2001. A microsatellite-based genetic linkage map for channel catfish (*Ictalurus punctatus*). *Genetics* 158: 727–734.

Wang, S., Bao, Z., Pan, J., Zhang, L., Yao, B., Zhan, A., Bi, K. & Zhang, Q. 2004. AFLP linkage map of an intraspecific cross in *Chlamys farreri*. *J. Shellfish Res*. 23: 491–499.

Weir, B.S. 1996. *Genetic data analysis, 2nd edition*. Sunderland, MA, USA., Sinauer Associates, Inc.

Wilson, K., Li, Y., Whan, V., Lehnert, S., Byrne, K., Moore, S., Pongsomboon, S., Tassanakajon, A., Rosenberg, G., Ballment, E., Fayazi, A., Swan, J., Kenway, M. & Benzie, J. 2002. Genetic mapping of the black tiger shrimp (*Penaeus monodon*) with amplified fragment length polymorphism. *Aquaculture* 204: 297–309.

Woram, R.A., McGowan, C., Stout, J.A., Gharbi, K., Ferguson, M.M., Hoyheim, B., Sakamoto, J., Davidson, W., Rexroad, C. & Danzmann, R.G. 2004. A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome* 47: 304–315.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P. & Thorgaard, G.H. 1998. A detailed map of rainbow trout produced using double haploids. *Genetics* 148: 839–850.

Yu, Z.N. & Guo, X.M. 2003. Genetic linkage map of the eastern oyster (*Crassostrea virginica* Gmelin). *Biol. Bull.* 204: 327–338.

Zimmerman, A.M., Evenhuis, J.P., Thorgaard, G.H. & Ristow, S.S. 2004. A single major chromosomal region controls natural killer cell-like activity in rainbow trout. *Immunogen.* 55: 825–35.

Zimmerman, A.M., Wheeler, P.A., Ristow, S.S. & Thorgaard, G.H. 2005. Composite interval mapping reveals three QTL associated with pyloric caeca number in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 247: 85–95.

CHAPTER 17

Marker-assisted selection in fish and shellfish breeding schemes

Victor Martinez

SUMMARY

The main goals of breeding programmes for fish and shellfish are to increase the profitability and sustainability of aquaculture. Traditionally, these have been carried out successfully using pedigree information by selecting individuals based on breeding values predicted for traits measured on candidates using an "animal model". This methodology assumes that phenotypes are explained by a large number of genes with small effects and random environmental deviations. However, information on individual genes with medium or large effects cannot be used in this manner. In selective breeding programmes using pedigree information, molecular markers have been used primarily for parentage assignment when tagging individual fish is difficult and to avoid causing common environmental effects from rearing families in separate tanks. The use of these techniques in such conventional breeding programmes is discussed in detail.

Exploiting the great biological diversity of many fish and shellfish species, different experimental designs may use either chromosomal manipulations or large family sizes to increase the likelihood of finding the loci affecting quantitative traits, the so-called QTL, by screening the segregation of molecular markers. Using information on identified loci in breeding schemes in aquaculture is expected to be cost-effective compared with traditional breeding methods only when the accuracy of predicting breeding values is rather low, e.g. for traits with low heritability such as disease resistance or carcass quality. One of the problems facing aquaculture is that some of the resources required to locate QTL accurately, such as dense linkage maps, are not yet available for the many species. Recently, however, information from expressed sequence tag (EST) databases has been used for developing molecular markers such as microsatellites and single nucleotide polymorphisms (SNPs). Marker-assisted selection (MAS) or genome-wide marker-assisted selection (G-MAS) using linkage disequilibrium within families or across populations are not widely used in aquaculture, but their application in actual breeding programmes is expected to be a fertile area of research. This chapter describes how genomic tools can be used jointly with pedigree-based breeding strategies and the potential and real value of molecular markers in fish and shellfish breeding schemes.

INTRODUCTION

The main goals of fish and shellfish breeding programmes are to increase the profitability and sustainability of production enterprises, while maintaining genetic variability in the cultured stock. Traditionally, selective breeding has targeted traits such as body weight that can be easily improved using mass selection. Relatively few studies have analysed other traits that are economically important. However, disease resistance and carcass quality are traits that are difficult to measure on candidates for selection, but have major effects on the production efficiency and profitability of many species in aquaculture.

When developing efficient breeding programmes, pedigree information is required to maximize effective population sizes and to use information from relatives to increase the accuracy of predicting breeding values for all traits included in the breeding objective. In most commercial applications, pedigree information is lacking; therefore, markers can be used to infer relatedness between individuals, with or without parental information. Several issues need to be considered on a case-by-case basis when applying such molecular information for increasing the profitability of breeding programmes in practice.

For traits that are difficult to measure on candidates for selection, prediction of breeding value has to rely on measurements on relatives. Under these circumstances, the accuracy of predicted breeding values (and thus, response) is lower than when records are obtained directly on candidates for selection. In addition, there is an increased probability of co-selecting relatives. It is especially for these traits that molecular markers that directly affect or are linked to quantitative trait loci (QTL) have been regarded as useful for marker-assisted selection (MAS) or gene-assisted selection (GAS) programmes.

This chapter begins by discussing the status of "conventional" breeding programmes, the challenges involved when starting such programmes for new species and the possibilities of incorporating marker information in "conventional" programmes. An outline is then provided of the molecular markers developed for aquaculture species and of their use for genetic analysis. The main features of designs for QTL mapping, including the use of chromosomal manipulations, are described, followed by a discussion of the prospects and challenges of GAS or MAS for disease or carcass traits. Finally, new genomic tools are considered briefly.

BREEDING PROGRAMMES AND RESPONSE TO SELECTION

Management of modern breeding programmes in aquaculture requires using pedigree information to carry out sound and efficient statistical evaluations (using best linear unbiased prediction [BLUP] methodology). This approach enables breeders to maximize genetic gain while limiting rates of inbreeding to acceptable levels (Meuwissen, 1997; Toro and Mäki-Tanila, 1999).

Most of the genetic improvement in fish and shellfish species to date has been made through the use of traditional selective breeding (reviewed by Hulata, 2001). Well-designed breeding programmes have shown substantial response to selection for body weight, e.g. Atlantic salmon, 10–14 percent (Gjøen and Bentsen, 1997). In rainbow trout, rates of genetic gain varied from 8 percent for indirect selection for body weight at sea (Kause *et al.*, 2005) to 13 percent for direct selection (Gjerde, 1986). The response to selection was about

10 percent for body weight in a breeding programme for coho salmon (CMG-IFOP) initially funded by FAO (Martinez and Hidalgo, unpublished data), and a similar response was obtained for this species in the United States of America (Hershberger *et al.,* 1990). Estimates for tilapia follow largely the same trend, with a response of about 10 percent (Ponzoni *et al.,* 2005). In common carp, responses to selection for body weight were inconsistent between up-selected and down-selected lines, although exhaustion of additive genetic variation for increased growth rate, genotype-by-environment interaction, or competition effects could not be ruled out (Moav and Wohlfarth, 1976). In oysters, asymmetrical response to selection for body weight was found (Toro *et al.,* 1995; Ward, English and McGoldrick, 2000).

Although responses to selection have not been well documented, significant estimates of genetic parameters have been obtained for carcass traits (Gjerde and Schaeffer, 1989; Kause *et al.,* 2002; Quinton, McMillan and Glebe, 2005) and disease resistance (Gjøen *et al.,* 1997; Henryon *et al.,* 2002, 2005). Rates of genetic gain are expected to be lower for these traits than for body weight because breeding value predictions rely solely on measurements from relatives.

Several breeding programmes have been initiated recently for new aquaculture species, such as mussels, scallops, *Artemia* and shrimp. The biology of these species poses interesting avenues for the design of conventional breeding programmes, taking into account factors such as self-fertilization, intrafamily competition, cannibalism, lack of methods for physical tagging, and mating preferences. For example, competition can affect the expression of quantitative traits due to co-variances among members of a group managed together in a pond or tank and, if not considered properly, this effect can seriously affect the rates of response to selection (Muir, 2005). However, this effect can be included explicitly in the model of analysis using the co-variance among members of a group, the so-called "associative effects" from other genotypes in the group. The theory of Griffing (1967) for BLUP evaluation was developed in the context of tree breeding, but deserves further investigation in the analysis of fish and shellfish breeding. This may be especially true for species taken recently from the wild or those that show cannibalistic behaviour.

Another recent example is the development of scallop breeding programmes. *Argopecten purpuratus* is a simultaneously hermaphroditic species. In the first breeding phase, the scallop liberates sperm, after which the eggs are expelled. To decrease the level of self-fertilization, it is customary to use only the last pulses of eggs. This system reduces rates of self-fertilization to 20 percent (A. Vergara, personal communication), but a residual proportion of eggs are still already fertilized with sperm from the same individual. As this process occurs within the reproductive tract, it is not possible to detect which individuals are selfed or outcrossed, although the rate of residual self-fertilization varies widely among families and produces biased estimates of heritability (Martinez and di Giovanni, 2006). Information from molecular markers can be of benefit under these circumstances (see below).

DNA MARKERS USED IN AQUACULTURE

Mutations in the genome create genetic variability (or polymorphism), which is reflected as allelic diversity of molecular markers. While genomic sequencing would greatly facilitate the development

of molecular markers, the many species in aquaculture would make this a costly task (Liu and Cordes, 2004). Hence, a variety of approaches have been taken to develop genetic markers for aquaculture species.

Dominantly-expressed markers have been used extensively in aquaculture studies. Amplified fragment length polymorphism (AFLP) markers (Vos *et al.,* 1995) provide a cost-effective alternative for species where DNA sequencing is not under way or when there are restricted resources for QTL mapping. Dominant AFLP markers are preferred over random amplified polymorphic DNA (RAPD) markers because they are more reproducible both in other lines or populations and in other laboratories (e.g. Nichols *et al.,* 2003), and they can generate hundreds of markers (a single polymerase chain reaction commonly generates over ten markers). Furthermore, heterozygotes can often be distinguished from homozygotes using the fluorescent band intensity (Piepho and Koch, 2000; Jansen *et al.,* 2001).

Microsatellite markers are simple sequence repeats (SSRs) arranged in tandem arrays scattered throughout the genome, both within known genes and in anonymous regions. Microsatellite markers are used increasingly in aquaculture species (reviewed by Liu and Cordes, 2004), due to their elevated polymorphic information content (PIC), co-dominant mode of expression, Mendelian inheritance, abundance and broad distribution throughout the genome (Wright and Bentzen, 1994). Microsatellites are generally Type II markers, which are associated with genomic regions that have not been annotated to known genes (O'Brien, 1991). Other molecular markers can be distinguished as Type I markers, which are linked to genes (of known function). Type I markers are more desirable because they are generally more conserved across evolutionarily distant organisms, enabling comparative genomics, assessment of genome evolution and candidate gene analysis.

Two procedures are used to generate microsatellite markers. The first uses a genomic library enriched with microsatellite-bearing sequences to generate clones that bear specific SSRs. These clones are then sequenced to identify microsatellite-bearing sequences and then to design primers to amplify the regions with specific SSR. Validation is required to study the level of polymorphism and the number of null alleles, and to identify any loci that are duplicates due to any recent evolutionary genome duplication event giving rise to multiple copies of loci in the haploid genome (Coulibaly *et al.,* 2005). This is done by screening a sample of individuals from the target population.

Many laboratories have been working on developing expressed sequence tags (ESTs) derived from complementary DNA (cDNA) libraries for a variety of fish and shellfish species (Panitz *et al.,* 2002; Rise *et al.,* 2004a; Hayes *et al.,* 2004; Rexroad *et al.,* 2005; A. Alcivar-Warren, personal communication). EST sequences can be used for marker development in species where the full genome is not currently being sequenced. The cDNA libraries are constructed using messenger RNA (mRNA) that was expressed in different tissues, such as kidney and gills. The expressed fragments of sequence data are not the full sequence of a known gene, but what was incorporated into a mature mRNA molecule.

In addition to the library-based method of marker development previously described, microsatellites can be developed from EST databases or from known gene sequences. As it is possible to connect the

TABLE 1
Recently published linkage maps for various fish and shellfish species used in aquaculture

Species	Number of markers	Marker type	Map length Female/Male	Male cM (Kosambi)	Female cM (Kosambi)	Reference
Atlantic salmon	473	AFLP	8.26:1	103	901	Moen *et al.*, 2004a
	54	Microsatellites				
	65	Microsatellites	3.92	np	np	Gilbey *et al.*, 2004
Rainbow trout	226	Microsatellites	-	4 590		Nichols *et al.*, 2003
	973	AFLP				
	4	Allozymes				
	72	VNTR				
	29	Known genes				
	12	Minisatellites				
	5	RAPDs				
	38	SINE*				
Oysters	115	Microsatellites	1.31:1	776	1 020	Houbert and Hedgecock, 2004
Sea bass	174	Microsatellites	1.6:1	567.4	905.9	Chistiakov *et al.*, 2005
Kuruma prawn	195	AFLP		1 780	1 026	Li *et al.*, 2003
Tilapia	525	Microsatellites	1:1	1 300		Lee *et al.*, 2005
	21	Genes				
Scallops	503	AFLP	1.27:1	2 468	3 130	Wang *et al.*, 2005
Common carp	110	Microsatellites	-	4 111		Sun and Liang, 2004
	105	Known genes				
	57	RAPDs				
Japanese flounder	111	Microsatellites	7.4:1	741.1	670.4	Coimbra *et al.*, 2003
	352	AFLP				
Channel catfish	313	Microsatellites	3.18:1	1 958[B]		Waldbieser *et al.*, 2001

[B] Sex-averaged
* Short interspersed elements
np = not published

function of the transcript of genes (from an EST sequence) with the presence of a microsatellite, these markers are Type I markers (O'Brien, 1991; Serapion *et al.*, 2004; Ng *et al.*, 2005). This strategy of developing microsatellite markers from known genes and ESTs has been used for common carp (Yue, Ho and Orban, 2004), rainbow trout (Rexroad *et al.*, 2005; Coulibaly *et al.*, 2005) and Atlantic salmon (Ng *et al.*, 2005; Vasemägi, Nilsson and Primmer, 2005).

In all these analyses, high levels of transferability between populations and species can be expected if the microsatellites are included in coding regions. Such transferability has been observed e.g. between Atlantic salmon and rainbow trout (Vasemägi *et al.*, 2005; Rexroad *et al.*, 2005), making these markers ideal for analyses of population genetics and comparative maps. For example, microsatellites derived from EST sequences have been used to study divergence of Atlantic salmon populations in salt, brackish and freshwater habitats (Vasemägi, Nilsson and Primmer, 2005).

Bioinformatic tools can be used for potential discovery of SNPs using DNA sequence alignment "*in silico*" (Marth *et al.*,

1999). Although it is possible to use base quality values to discern true allelic variations from sequencing errors, validation is a key step for true positive detection of SNPs (Marth *et al.*, 1999). This is generally carried out using a proportion of the SNPs detected in a sample of individuals from the target population. This strategy has been used recently for SNP detection using EST sequences from Atlantic salmon (Panitz *et al.*, 2002; Hayes *et al.*, 2004).

Linkage maps

A linkage map is an ordered collection of the genes and genetic markers occurring along the lengths of the chromosomes of a species, with distances between them estimated on the basis of the number of recombination events observed in the data. Genetic linkage maps have been published for rainbow trout (Young *et al.*, 1998; Sakamoto *et al.*, 2000; Nichols *et al.*, 2003), channel catfish (Waldbieser *et al.*, 2001), tilapias (Kocher *et al.*, 1998; Lee *et al.*, 2005) and Japanese flounder (Coimbra *et al.*, 2003). References to updated linkage maps of the major aquaculture species are given in Table 1. Dense linkage maps including a relatively large number of markers are under development.

Different patterns of recombination appear among regions of linkage groups in certain male maps, with markers clustered in centromeric regions, an extreme example being Atlantic salmon where recombination in males is greatly reduced (Moen *et al.*, 2004b). The molecular mechanisms responsible for the differences in recombination rates between sexes are not well understood, although studies on model organisms such as zebrafish, where genomic sequencing is currently under way, may help to clarify this (Singer *et al.*, 2002).

USING MARKERS TO AID CONVENTIONAL FISH AND SHELLFISH BREEDING PROGRAMMES

Molecular markers may be used in a number of ways to aid conventional breeding of fish and shellfish species, and some of these are described and exemplified below.

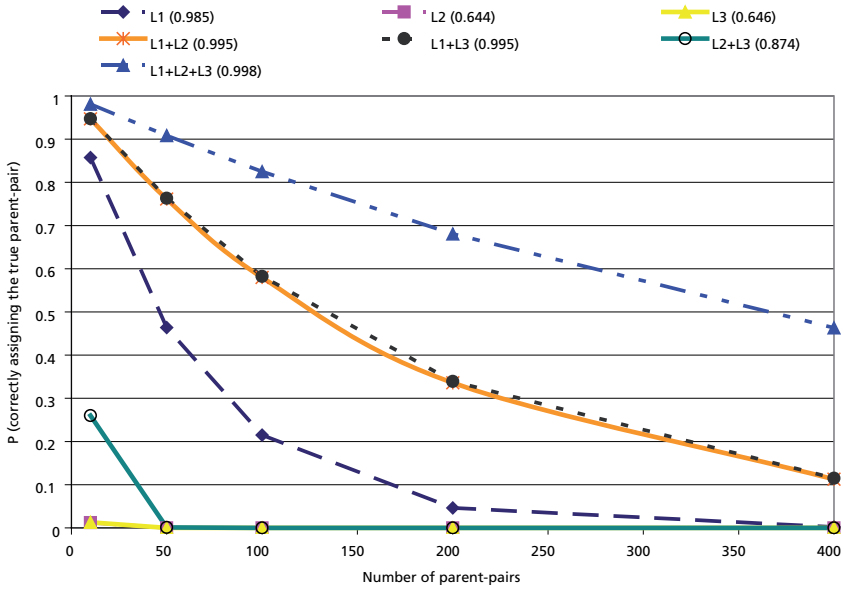
Parentage analysis

One of the main constraints facing effective breeding programmes for fish and shellfish is that newborn individuals are too small to be tagged individually. Application of the animal model approach (i.e. using a statistical genetic model to predict individual breeding values) requires tagging a constant number of individuals from each family with passive integrated transponders (PIT tags) when they become sufficiently large after a period of individual family rearing. However, this system of early management creates common environmental (i.e. tank) effects for full-sib families (Martinez, Neira and Gall, 1999). To address this issue, mixtures of equal-aged progeny from different families can be reared communally to preclude the development of such family-specific environmental effects, and genetic markers can be used subsequently to assign individuals to families after evaluation of individual performance (Doyle and Herbinger, 1994). Thus, the impact of early common environmental effects is considerably reduced if markers are used for parentage analysis when selecting individuals for early growth rate traits (Herbinger *et al.*, 1999; Norris, Bradley and Cunningham, 2000).

The amount of marker data needed to achieve acceptable levels of correct parentage assignment depends on the number of loci, the number of alleles and the number of parent-pairs (sires and dams) available for reconstructing the pedigree (Jamieson

FIGURE 1

Predicted probability of assigning the correct sire-dam (parent-pair) for a given number of parent-pairs (*x* axis) using different combinations of three microsatellites (L1, L2 and L3) amplified in *Oncorhynchus kisutch*



Probabilities of exclusion included between parentheses in the legend are obtained using the deterministic method of Villanueva, Verspoor and Visscher, 2002.
Data obtained from Diagnotec SA.

and Taylor, 1997; Villanueva, Verspoor and Visscher, 2002). The information from the marker data available for each species can be studied using exclusion probabilities, which are then used to calculate the probability (PC) of correctly assigning the true parent-pair (sire and dam) to offspring that are genotyped (Villanueva, Verspoor and Visscher, 2002).

Figure 1 presents the results for three microsatellites and combinations of microsatellites to predict the probabilities of exclusion and PC. The allelic frequencies of the three microsatellites were calculated with a sample (n=100) from a coho salmon (*O. kisutch*) farm in southern Chile managed under commercial conditions. The analysis showed that the probability of assigning the true parent-pair depended greatly on the number of parent pairs available for parentage. Only for an unrealistically small number of ten sires and dams is there a high probability of assigning the correct parent-pair to offspring. For a breeding programme of 200 or 300 parent-pairs, PC decreased considerably. Therefore, in this example, more markers are needed for accurate pedigree reconstruction. Successful parentage assignment experiments typically have used six to eight microsatellite markers (Herbinger *et al.*, 1995; Garcia de Leon *et al.*, 1998; Norris, Bradley and Cunningham, 2000; Castro *et al.*, 2004). In practice, the presence of genotyping errors, null alleles, realized mutations and non-Mendelian segregation can seriously affect the efficiency of parentage assignment (Castro *et al.*, 2004). Parentage assignment in the con-

text of fish breeding is also discussed by Sonesson (this volume).

For most breeding programmes, physical tagging will prove efficient both in economic and biological terms to achieve acceptable rates of genetic gain, while minimizing rates of inbreeding. Genetic marker technology can still be costly for routine assignment of parentage, although these costs can be reduced using multiplex polymerase chain reaction (PCR) technology (Paterson, Piertney and Knox, 2004; Taris, Baron and Sharbel, 2005) in which more than one marker can be genotyped simultaneously in a single gel lane or capillary. This is especially the case when only DNA markers are used without physical tagging, as individuals must be retyped when records for multiple traits are included in the selection criteria (Gjerde, Villaneuva and Bentsen, 2002).

It is expected that rates of genetic gain for economic traits will not be affected significantly when common environmental effects are present. This is because, in many species of cultured salmonids, the common environmental effect decreases considerably, from about 20 percent for alevin weight to 5 percent for body weight at harvest, which is the trait with most impact on profit (Herbinger *et al.,* 1999; Henryon *et al.,* 2002; Kause *et al.,* 2005). Hence, common environmental effects should not decrease the rates of genetic gain for traits measured at harvest when physical tagging is used. Furthermore, multistage selection offers the possibility of first selecting individuals on a within-family basis directly from tanks (for traits influenced by common environmental effects), and then selecting at a second stage for traits measured at harvest (Martinez, 2006a). This alternative would either maintain rates of gain while decreasing the costs associated with tagging, or even increase rates of gain, when recording from tanks (within families) can be carried out relatively inexpensively (Martinez, 2006a).

The sample size (i.e. the numbers of individuals and markers required for reconstructing the pedigree of a population accurately) is a practical issue, as not all individuals in a population can be genotyped for all markers available. Such issues arise in species where physical tagging is not possible or not economically sound, as in nucleus populations without electronic tagging (e.g. when recovering a back-up population for nucleus breeding programmes) or when disease challenges (e.g. for infectious pancreatic necrotic virus [IPNV]) are carried out early in the life cycle (Martinez *et al.,* in preparation). Small sample sizes, together with sperm competition (Withler and Beacham, 1994), mating preference (as in *Artemia*; G. Gajardo, personal communication) and other biological factors after fertilization can increase the variance of family size, thereby decreasing the effective population size to unsustainable levels (Brown, Woolliams and McAndrew, 2005).

Another problem arises in practice when selection is carried out before genotyping with markers. In this case, BLUP of breeding values is likely to be biased because not all phenotypic information is used when predicting breeding values. The magnitude of re-ranking is dependent on the amount of information from a family within the selected group. In these instances, the mixed model equations need to be modified to account for such selected data (Morton and Howarth, 2005).

Establishing breeding programmes using molecular information

The choices made at the founding of a breeding programme have a critical

bearing on its ultimate success. Criteria for choosing individuals that will be founders should be essentially the same as those used when the selection response is optimized under restricted co-ancestry when pedigree information is available (Meuwissen, 1997; Toro and Mäki-Tanila, 1999). Thus, it is necessary to avoid matings between close relatives for managing existing quantitative genetic variation at the start of the programme. Experiments with the planktonic microcrustacean *Daphnia* spp. have shown that neutral genetic variation gives little indication of the levels of quantitative genetic variation available for selection (Pfrender *et al.,* 2000). However, increasing the population size at the beginning of the breeding programme will diminish the subsequent effect of random genetic drift, and therefore larger founding populations will have an increased likelihood of showing response to selection. Lack of adequate base populations is the main reason for the lack of selection response observed in some species of fish (Gjedrem, 2000).

The effective population size (N_e) required for setting up a breeding programme depends on the policy regarding risk management (Brown, Woolliams and McAndrew, 2005), but to prevent decline in fitness, some authors have recommended N_e values ranging from 31 to 250, which in terms of rates of inbreeding should be less than 2 percent (Meuwissen and Woolliams, 1994). Due to the large family sizes possible for many fish and shellfish species, breeding programmes that fail to control the genetic contributions of parents in every generation are expected to incur relatively high rates of inbreeding (Meuwissen, 1997). The situation is even more extreme when selection is based on a complex breeding objective that includes information from relatives and many traits jointly (Martinez, 2006b).

Fish within commercial production populations generally are not tagged individually and pedigree information is therefore lacking. Genetic markers allow the estimation of pairwise relatedness between individuals or sib-ship reconstruction even with unknown ancestors (Toro and Mäki-Tanila, 1999; Thomas and Hill, 2000; Toro, Barragán and Óvilo, 2002; Wang, 2004; Fernandez and Toro, 2006). There is a plethora of estimators for calculating pairwise relatedness (Queller and Goodnight, 1989; Lynch and Ritland, 1999). The efficiency of inferring pairwise relatedness using markers without parental information is affected by assuming known allele frequencies in the base population and unlinked loci in Hardy-Weinberg equilibrium. Furthermore, pair-wise methods can lead to inconsistent assignations between triplets of individuals because they use information from only two individuals at a time (Fernandez and Toro, 2006). In addition, it is difficult to set thresholds for claiming different types of relatedness in the data (Thomas and Hill, 2000; Norris, Bradley and Cunningham, 2000). On the other hand, sib-ship reconstruction methods do not attempt to calculate co-ancestry; rather, they attempt to reconstruct full- or half-sib or other family groups (Thomas and Hill, 2000; Emery, Boyle and Noble, 2001; Smith, Herbinger and Merry, 2001). Such reconstructions of full- or half-sib families or even other groups of relatives appear robust to lack of knowledge of base population allele frequencies (Thomas and Hill, 2000; Fernandez and Toro, 2006).

Marker information can be used to infer relatedness between individuals available as candidate broodstock to generate the first generation of offspring in the breeding programme, and thereby avoid mating among close relatives. This approach uses molec-

ular information to infer the genealogical pedigree. A simulation was conducted to reconstruct the pedigree of 100 potential candidates from ten full-sib families (with a Poisson family size equal to ten using six equally-frequent microsatellites, without parental genotypes (Martinez, 2006c). The posterior probability of either full [*P*(FS)] or half-sib [*P*(HS)] groups was obtained using the Bayesian model of Emery, Boyle and Noble (2001). In the simulation results, there was a tendency to overestimate relationships, with posterior probabilities over 0.5 when individuals were in fact unrelated. On the other hand, not all true full-sibs were assigned to the correct full-sib family with the greatest probability, and some true full-sib family members were reconstructed as half-sibs. On average (among ten replicates), the probability of mating related individuals was significantly smaller when information from molecular markers was used, compared with what was expected by chance (4.7 percent versus 18.1 percent, $p = 0.002$). The practical implication is that inbreeding in the progeny generation would average 5 percent when random mating is used and 1 percent when optimization using molecular information is used.

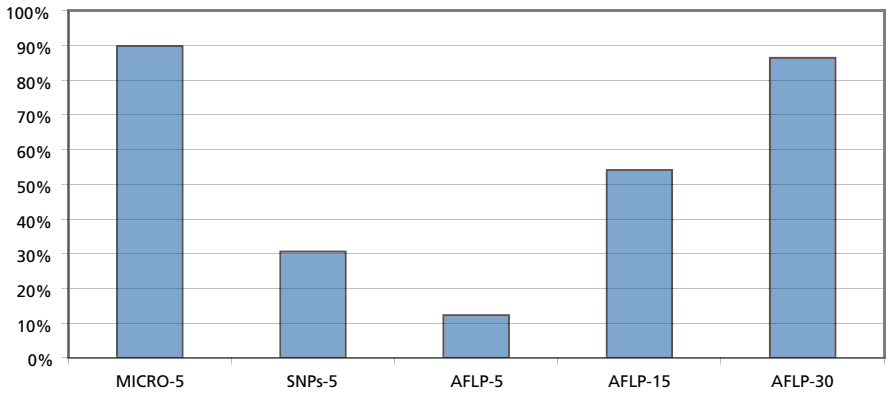
In practice, to perform mating in the base population, the relatedness inferred from molecular information does not need to be perfectly accurate, but it does require that relatedness is not underestimated greatly. Among the technical issues that arise when using marker data are that a pair of individuals could be misclassified as related when they are in fact unrelated (Type I error) or a pair may be wrongly classified as unrelated when the pair is in fact related (Type II error). Type II error is of greatest concern as this could result in related pairs being mated. This is because mating of individuals (males and females) as unrelated when in fact they are full sibs will increase true inbreeding in the population, while misclassification leading to unrelated individuals being assigned to a full-sib family would not increase the inbreeding in the progeny. The presence of mutations, null alleles or genotyping errors will underestimate the true relationships in the population and eventually increase the probability of mating true full-sibs (Butler *et al.,* 2004). Recently, Wang (2004) suggested a method for inferring relationships for marker data with a high error rate and mutation that can be used to address this issue. It should also be noted that studies dealing with estimation of heritability or prediction of breeding values with pedigrees reconstructed using molecular markers may be very inefficient when pedigrees are reconstructed with an increased rate of Type I errors (Mosseau, Ritland and Heath, 1998; Thomas, Pemberton and Hill, 2000).

Detecting self-fertilization in scallops

In scallops, a main drawback when implementing breeding programmes is the occurrence of self-fertilization, even when gametes from later spawning pulses are used for obtaining family material (Martinez and di Giovanni, 2006), i.e. a mixture of selfed and outcrossed individuals can be present even at later stages within a single family. Bias in estimating genetic parameters is expected due to this residual self-fertilization, which can occur with considerable frequency (average 20 percent) within particular families.

A simulation study was used to investigate to what extent markers with different information content can be used to discriminate between selfed and outcrossed individuals within a family (Figure 2). The results showed that microsatellites gave mean values of posterior probabilities greater

FIGURE 2
Identification of selfed individuals within families of scallops using different types of marker data[1]



Vertical bars represent the proportion of the 100 replicates in which the mean posterior probabilities of being selfed (for true selfed individuals) were greater than 0.95.

[1] MICRO-5: five microsatellites with six equally frequent alleles each. SNPs-5: five SNP markers with equal allele frequencies. AFLP-5, -15 or -30: 5, 15 or 30 AFLP markers. The design of the simulations of self-fertilization in scallops: the amount of self-fertilization was modelled using a truncated normal distribution which best fitted the empirical distribution of self-fertilization (Martinez and di Giovanni, 2006). A Bayesian model was used to infer mutually exclusive posterior probabilities of being either selfed or outbred (Anderson and Thompson, 2002). It was assumed that parental information was lacking, with unlinked markers and vague priors. Selfed individuals were regarded as having been detected correctly when the posterior probabilities of being selfed were greater than 0.95 (this criterion was determined empirically for operational reasons).

Source: V. Martinez, in preparation.

than 0.95 in about 90 percent of the families simulated (100 in total). Similar results were obtained with 30 AFLP markers, but these percentages were considerably reduced for smaller numbers of AFLPs or SNPs.

The information from these markers can be used to cull individuals, to construct a relationship matrix in which all unusual relationships are incorporated in analyses used for obtaining unbiased estimates of heritability and genetic correlations, and for estimating breeding values from real data sets (Martinez, 2006a).

IDENTIFYING QTL AND MAJOR GENES INFLUENCING COMPLEX QUANTITATIVE TRAITS

Molecular biology can greatly help the discovery of factors influencing the expression of quantitative traits. There are a number of ways in which this information can be used, the difference between them being the level of resolution with which these factors can be mapped. For example, loci with major effects on quantitative traits (QTL) are mapped by using markers to track inheritance of chromosomal regions in families or in inbred line crosses using the extent of linkage disequilibrium generated in the population. This approach gives a limited amount of mapping resolution. Fine mapping requires information from additional markers and individuals sampled across the outbred population and, while helping to narrow the confidence interval of the position of the QTL, this is only the starting point for identifying the polymorphisms in the performance-determining genes themselves. In practice, identification of genes influencing specific traits is achieved using

a combination of genetic mapping (linkage and fine mapping) to localize the QTL to a small region on the chromosome under analysis, and candidate gene or positional cloning approaches to identify the genes within the QTL region.

In some cases, sufficient biochemical or physiological information is available to investigate the association between the quantitative expression and the level of marker polymorphisms within specific genes. Nevertheless, this approach requires a great amount of detailed information in order to choose which gene explains the greatest effect and to have sufficient power to detect the association. This information is starting to appear in the aquaculture literature from multinational projects such as the Consortium of Genomic Resources for All Salmonids Project (cGRASP) (Ng *et al.,* 2005).

QTL mapping in fish using linkage disequilibrium: theoretical and practical considerations

Value of chromosomal manipulations

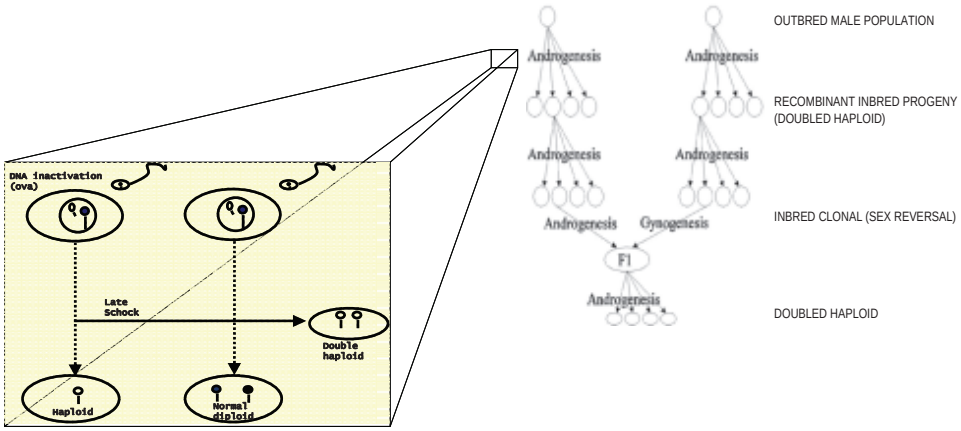
The great reproductive flexibility of fish enables different breeding designs to be implemented relatively easily. Completely homozygous fish can be produced in only one generation using chromosome set manipulations, without the many generations of inbreeding needed in other vertebrates. These manipulations enable doubling of the chromosomal complement of a haploid gamete (Young *et al.,* 1996; Corley-Smith, Lim and Bradhorst, 1996). Androgenetic double haploid individuals can be obtained by fertilizing eggs that were inactivated with gamma radiation, yielding haploid embryos containing only paternal chromosomes. Alternatively, gynogenetic double haploid individuals can be obtained by activating the development of eggs with ultraviolet-inactivated sperm, yielding haploid embryos containing only maternal chromosomes. In each case, diploidy is restored using methods that suppress the first mitotic division (Figure 3; Streisinger *et al.,* 1980; Corley-Smith, Lim and Bradhorst, 1996; Bijma, van Arendonk and Bovenhuis, 1997; Young *et al.*, 1998). The use of these reproductive manipulations to provide experimental populations for genetic analysis of complex quantitative traits has been well described (Bongers *et al.,* 1997; Robison, Wheeler and Thorgaard, 2001; Tanck *et al.,* 2001).

Double haploids from inbred line crosses

After a second round of uniparental reproduction (Figure 3), a collection of clonal lines can be obtained that collectively is likely to represent all the genetic variants from the base population (Bongers *et al.*, 1997). Crosses of sex-reversed double haploid individuals from lines that diverge for the traits of interest can produce F_1 lines in complete linkage disequilibrium. These F_1 populations can be used for further experimentation based on F_2 or backcross designs. Another round of androgenesis of F_1 individuals will produce a population of fully homozygous individuals. This design will have twice the power for detecting QTL as the standard F_2 design (Martinez, 2003). The standard deviation of QTL position estimates is halved for the double haploid design. This is due to an increase in the additive genetic variance, which is doubled for the double haploid design due to redistribution of the genotype frequencies in the progeny generation (Falconer and Mackay, 1996).

Informative double haploid populations of this sort have been utilized to perform QTL analysis for embryonic development rate in rainbow trout (Robison, Wheeler and Sundin, 2001; Martinez *et al.,* 2002;

FIGURE 3
Chromosomal manipulations in fish



ANDROGENESIS

In androgenesis, fertilization is carried out using radiation-inactivated ova, and a late shock is used to suppress first mitosis and thereby restore diploidy.
Adapted from Thorgaard and Allen, 1987.

Martinez *et al.*, 2005). At least four QTL of relatively large effect explain about 40 percent of the phenotypic variance of the mapping population and most of the 2.5 standard deviations of the difference between the original clonal lines used to generate the F_1 population (Robison, Wheeler and Thorgaard, 1999). Two linked QTL were in repulsion phase in the F_1 population, and were undetected in the analysis using composite interval mapping. This result was not surprising as evidence was accumulated among replicates of lines that were incubated at different temperatures (Robison, Wheeler and Sundin, 2001), and the Bayesian multiple QTL method incorporated all the available information of environmental co-variates in the analysis (Martinez *et al.*, 2005). Recently, these double haploid lines have been used for mapping QTL related to the number of pyloric caeca (Zimmerman *et al.*, 2005) and for confirming QTL influencing development rate (Sundin *et al.*, 2005).

When traits are associated and by taking into account the correlated structure of the data, multivariate estimation of QTL effects is expected to be more powerful than single trait analysis (Jiang and Zeng, 1995). Also, from a genetic standpoint, joint analysis provides the means for testing different hypotheses about the mode by which genes explained the genetic co-variation (Wu *et al.*, 1999). For example, after hypothesis testing (following Knott and Haley, 2000), a single pleiotropic QTL with opposite effects for development rate and length best explained the multivariate data (as detailed earlier by Martinez *et al.*, 2002b). This finding was also consistent with the negative correlation estimated with the data (Martinez *et al.*, 2002).

Double haploids in outbred populations

Martinez, Hill and Knott (2002) derived analytical formulae to predict the power of linkage analysis for interval mapping under three different mating designs in outbred

populations: full-sib mating, hierarchical mating, or double haploid designs. This analysis suggested that the use of double haploids appeared to be of benefit when detecting QTL, particularly when both the variance of the QTL and of the polygenic effects was small. Furthermore, given the relatively large size of full-sib families in fish, there appeared to be little advantage of hierarchical mating over full-sib mating designs for detecting QTL, the optimum family size depending on the size of the QTL and the population structure used for mapping (Martinez, Hill and Knott, 2002). The gain in power of the double haploid design comes from the increase in the variance of the Mendelian sampling term within families, which is effectively doubled for traits that are explained by additive effects (Falconer and Mackay, 1996).

As experimental settings constrain the total number of individuals genotyped, designs aimed at QTL mapping should include a small number of families of relatively large size in order to maximize the likelihood of detecting the QTL. This is because most of the information for mapping QTL uses linkage information that comes from within-family segregation (Muranty, 1996; Xu and Gessler, 1998). However, increasing power comes at the expense of reducing the accuracy of estimating the additive genetic variance for polygenic effects. A QTL mapping method has been developed for double haploids, which efficiently accommodates all the uncertainties that pertain to outbred populations, such as unknown linkage phases and differing levels of marker informativeness, using the identical-by-descent variance component method (see below; Martinez, 2003). Also, it is possible to combine double haploids and outbred relatives in the same family. Simulations of differing amounts of marker information and heritability for the QTL were used to compare the empirical power of the double haploid and full-sib designs. While the power of the full-sib design was lower than that for double haploids, QTL position estimates for double haploids had large confidence intervals (about 30 cM as compared with 40 cM for full-sibs; Martinez, 2003).

The double haploid design was used for mapping QTL for stress response in common carp using single marker analysis (Tanck *et al.*, 2001). The authors found only suggestive evidence for QTL, which is not surprising due to limited genome coverage for markers used in the analysis.

Published results have shown that double haploid lines are a useful resource for QTL detection studies. However, double haploid lines are difficult to develop due to the expression of deleterious recessive alleles (McCune *et al.*, 2002) and the low survival following shocks applied to restore diploidy to the haploid embryo. As the rate of male recombination is depressed, the precision of mapping QTL in androgenetic families is lower than that obtained using recombination events from females. Another practical matter is the labour needed for developing a clonal line, as at least two generations are required (Figure 3). This delay can be quite expensive and time-consuming for species with a long generation interval, such as salmon or trout (two to four years).

Aspects of QTL mapping in outbred populations of fish

Inbred line crosses are ideal for mapping QTL because they are expected to be completely informative for both markers and QTL, providing that the inbred lines are fixed for alternative alleles. Outbred populations are not completely informative for both QTL and markers; thus, experimental

power is expected to be lower than that for crosses between clonal lines. The power for detecting the QTL depends on allele frequencies, the probability of sampling an informative parent and family size.

Factors influencing the power of detecting QTL

Due to the large family sizes that can be obtained in many fish species, different mating designs using full-sib groups can be carried out for outbred populations. For example, full factorial designs may be used in which many males and females are mated to one another, and hierarchical designs may be applied in which each male is mated with multiple females, or each female with multiple males. For a given size of experiment, factorial and hierarchical designs have potentially a lower probability of sampling a heterozygous parent (because fewer sires and or dams are sampled overall), compared with the full-sib design in which each family has potentially two informative parents. For this reason, factorial and hierarchical designs can potentially give lower power compared with the simple full-sib design (Muranty, 1996; Martinez, Hill and Knott, 2002).

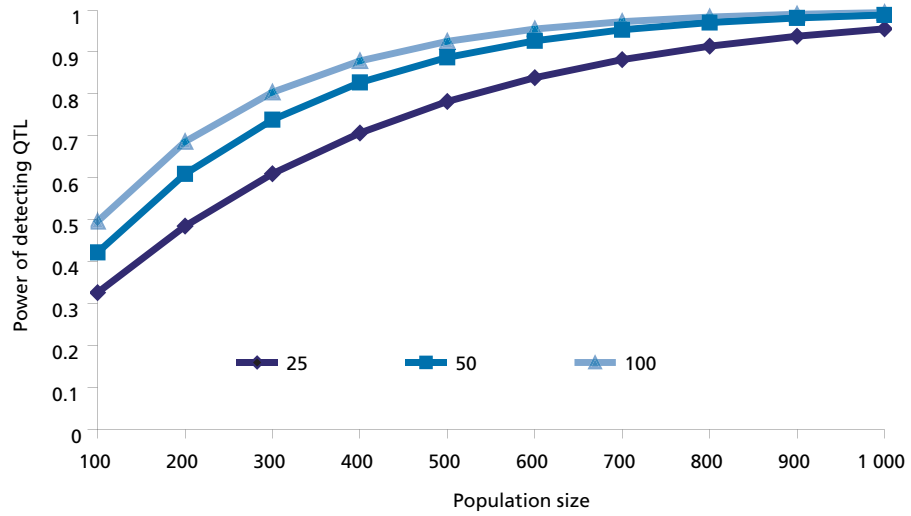
The optimum number of full-sib families sampled in the QTL mapping population depends on the intrinsic power of the experiment (i.e. size of the QTL effect and size of the population). As expected, large family sizes are needed for detecting QTL of small effects (Martinez, Hill and Knott, 2002). When the QTL explains 10 percent of the phenotypic variance, the optimum family size appears to be 50 individuals per family for a reasonably-sized QTL mapping experiment in outbred populations (Figure 4). Further increases in the number of individuals per family provide only a modest increase in power. Further, the same results used simulation models showing dominance and additive effects under the variance components method for mapping QTL (Martinez *et al.,* 2006a).

Methods of analysis

The method of choice when analysing data from outbred populations is the variance component method, in which QTL effects are included as random effects with a covariance proportional to the probability that relatives (e.g. full-sibs) share alleles identical by descent conditional on marker data (Xu and Atchley, 1995). This model is similar to the one used more generally for genetic evaluation of candidate fish for selection, but includes the random QTL effect.

A considerable proportion of the genetic variance for growth-related traits in fish populations has been explained by dominance (Rye and Mao, 1998; Pante, Gjerde and McMillan, 2001; Pante *et al.,* 2002). When mapping QTL using the random model, it is assumed that only additive effects are of importance and therefore only matrices of additive relationships conditional on marker data are fitted in the residual effect maximum likelihood procedure (George, Visscher and Haley, 2000; Pong-Wong *et al.,* 2001). However, the large family sizes in fish enable hypotheses for different modes of inheritance at the QTL to be tested using the within-family variance. While some authors have speculated that including dominance in the model will increase the power of detecting QTL (Liu, Jansen and Lin, 2002), others (Martinez, 2003; Martinez, 2006a) have shown that power to detect QTL was comparable between models including or not including dominance. This was particularly the case for the larger family sizes simulated and it was concluded that for most scenarios, the additive model was quite

FIGURE 4
Deterministic power calculation (following Martinez, Hill and Knott, 2002) for a QTL explaining 10 percent of the phenotypic variance for a variable total population size (x axis) and family size (25, 50 and 100)



robust for detecting QTL and there was little loss of information for detecting QTL when dominance is present but not used in the QTL mapping analysis.

QTL mapping in practice

To date, QTL mapping in fish using outbred populations has been carried out mostly with single marker analysis (microsatellites and AFLP markers), and using relatively sparse linkage maps when interval mapping is used. In tilapia, the F_2 design and a four-way cross between different species of *Oreochromis* have been used for detecting QTL affecting cold tolerance and body weight (Cnaani *et al.*, 2003; Moen *et al.*, 2004c). In outbred populations of salmonids, QTL that influence body weight have been mapped (Reid *et al.*, 2005 and references therein).

Studies seeking linkage of markers to traits amenable to MAS, such as disease resistance, have begun to appear in the literature over the past few years. For example, QTL for resistance have been mapped for infectious pancreatic necrosis virus (Ozaki *et al.*, 2001), infectious salmonid anemia (Moen *et al.*, 2004c), infectious haematopoietic necrosis (Rodriguez *et al.*, 2004; Khoo *et al.*, 2004), and stress and immune response (Cnaani *et al.*, 2004). Also, Somorjai, Danzmann and Ferguson (2003 and references therein) reported evidence of QTL for upper thermal tolerance in salmonids with differing effects in different species and genetic backgrounds.

From fine mapping to finding genes influencing complex traits

When the number of meioses in the genotyped pedigree is not sufficient for the linkage analysis to obtain a precise position for the QTL, there is a wide confidence interval around an estimated QTL position.

Fine mapping methods attempt to overcome this problem by quantifying the gametic phase or linkage disequilibrium (LD) present in an outbred population, i.e. across families. This method makes use of the number of generations as the appearance of a mutation and can produce extremely precise estimates of the QTL position (Pérez-Enciso *et al.,* 2003). The rationale behind using LD for mapping QTL is that when the population size is rather small, founders of the population would have only a limited number of haplotypes, and with very tightly linked loci there may not be sufficient time for recombination to break up the association between markers and the mutation affecting the quantitative trait.

LD mapping is carried out by calculating the probabilities that haplotypes shared by individuals are identical by descent from a common ancestor conditional on marker data (assuming t generations as the common ancestor and a certain N_e; Meuwissen and Goddard, 2001). The LD in the population depends on a number of population parameters such as the degree of admixture or stratification in the population and the actual level of association between the causal mutation and the polymorphisms. The correct determination of phases and of genotypes at the QTL is required for fine mapping purposes (Meuwissen and Goddard, 2001; Pérez-Enciso, 2003). For these reasons, a pure LD analysis is likely to result in a large number of false positives, i.e. falsely inferring association when there is no linkage.

Methods that incorporate the linkage information (within families) and LD jointly are preferred, because the likelihood of spurious association (i.e. LD without linkage) diminishes, making much better use of the whole data set (Meuwissen and Goddard, 2001, 2004; Pérez-Enciso, 2004). All of these methods, however, require a great deal of genotyping of tightly linked markers such as SNPs, which currently are not widely available for fine mapping in aquaculture species.

Using fine mapping techniques, the confidence interval for QTL position can be reduced considerably. However, to develop a direct test for a favourable polymorphism requires use of comparative mapping approaches with model species, such as zebrafish or *fugu*, to select the candidate genes that most likely affect the trait of interest. Otherwise, enrichment of markers in a specific region of the genome (to narrow further the most likely position of the polymorphism) following sequencing is needed to compare sequences between individuals that show different phenotypes or alternative QTL alleles.

Candidate gene analysis

It is tempting to invoke variation at genes with a known role in the physiology underlying a complex trait such as growth to explain phenotypic variability for the trait. These genes can be searched for polymorphisms (e.g. SNPs) and the variants then tested to determine whether they are correlated with the expression of the quantitative trait. This approach requires knowledge of the biology of the species, biochemical pathways and gene sequences in order to target variation at those specific genes. In aquaculture, most of this information is currently lacking, but it is expected that more genes will be incorporated in databases in the near future. The possibility exists to utilize data from highly studied model species, such as zebrafish or rainbow trout, in comparative bioinformatic approaches.

To date, this strategy has not proven particularly successful for explaining genetic

variation underlying complex (polygenic) traits. This is because although the biology of the trait and the genes most likely involved in the expression of the phenotype may be known, in complex traits many other genes may be involved in the metabolic pathway that are not obvious candidates. For example, in aquaculture species, candidate genes have been studied for growth-related traits using ten conserved gene sequences known to be related to the growth hormone axis (Tao and Boulding, 2003). In this study of Arctic charr, only a single SNP (of ten) from five of ten genes was found to be associated with growth rate.

Another example for disease resistance traits is the major histocompatibility complex (MHC). The genes of this complex encode highly polymorphic cell surface glycoproteins involved in specific immune responses and either specific alleles or heterozygotes at this complex were associated with resistance and susceptibility to *A. salmonicida* or infectious haematopoietic necrosis (IHN) virus (Langefors, Lohm and Grahn, 2001; Lohm *et al.,* 2002; Arkush *et al.,* 2002; Grimholt *et al.,* 2003; Bernatchez and Landry, 2003). Nevertheless, the background genome was quite important for explaining the difference in resistance between individuals within a family (Kjøglum, Grimholt and Larsen, 2005).

Microarrays, gene expression and identification of candidate genes for QTL analysis

Microarray technology (Knudsen, 2002) enables the expression of thousands of genes to be studied simultaneously. Until now, this information has been used primarily for following gene expression in treatment and control experiments in many fields such as disease exposure and stress response. This information can be used to discover new sets of candidate genes, possibly with or without functional assignment that may be related to the quantitative trait of interest (Walsh and Henderson, 2004). Genes whose expression differs between treatments are likely to be *trans*-acting genes, i.e. their expression is regulated by other genes. Therefore, it seems likely that seeking polymorphisms within these genes may not yield information about factors that explain the phenotype, and there might be problems assigning the correct significance threshold (Pérez-Enciso *et al.,* 2003). Further, because many genes are part of metabolic pathways and do not act individually, the expression of a single gene may be insufficient to explain phenotypic differences between individuals. Only those genes that directly affect phenotypic expression (i.e. *cis*-acting genes) can be treated as candidate genes for subsequent use in MAS after studying polymorphisms in their sequences. In salmonids, a microarray made available from the Consortium for Genomics Research on all Salmonids Project (cGRASP) has been used to study gene expression in fish exposed or not exposed to *Piscicricketsia salmonis* (Rise *et al.,* 2004b), and microarrays in other fish and shellfish species are currently under development.

A gene expression pattern can itself be regarded as a quantitative trait. Here, the interest is in finding associations between different patterns of gene expression and marker loci. This analysis was coined as "genetical genomics" by Jansen and Nap (2001). As is usual in QTL mapping, the analysis attempted to dissect the transcriptional regulation of the entire transcriptome and to identify the effects of individual QTL affecting gene expression (the so-called eQTL; e.g. Hubner *et al.,* 2005). To

date, this analysis relies upon the use of segregating populations (of known origin) such as recombinant inbred lines (Carlborg *et al.,* 2005), and the analysis of outbred populations poses greater challenges (Pérez-Enciso, 2004). Still, aquaculture species can provide sufficient information due to the large family sizes needed to unravel complex regulatory gene networks. How all this information can be included in MAS programmes is yet unclear.

INCORPORATING MOLECULAR MARKERS INTO BREEDING PROGRAMMES FOR FISH AND SHELLFISH

General aspects of incorporating molecular information in breeding programmes

The response to selection ΔG is estimated as:

$$\Delta G = i\sigma_H r$$

where i = the intensity of selection, r = the correlation between the breeding objective and the selection criteria (i.e. accuracy), and σ_H = the additive genetic standard deviation for the breeding objective. As the major impact of incorporating information from molecular markers will be on accuracy estimates, improvement of the response to selection will be higher for traits that have relatively small accuracy than for traits of relatively large accuracy. Thus, breeding programmes for traits with low heritability and relatively few records per trait measured such as carcass and disease resistance are those most benefiting from incorporating marker information (Meuwissen, 2003).

The relative increase in accuracy depends on the amount of variation explained by markers, which in turn depends on the number of QTL identified and used in MAS or GAS schemes (Lande and Thompson, 1990). QTL experiments in other species have shown that the effects of marked genes have a leptokurtic distribution, with a small number of genes having large effects and polygenes (Hayes and Goddard, 2001), which is likely to be the case in aquaculture species (Martinez *et al.,* 2005). Hence, it is expected that more than a single marked gene will be needed for MAS schemes to be efficient.

Due to the biology of many fish and shellfish species, multistage selection will likely prove useful in MAS or GAS schemes. Basically, a first stage of selection can be applied for traits expressed early in the life cycle (e.g. body weight), and a second stage of selection will incorporate information from relatives plus marked QTL. Optimization will be needed to determine the intensity of selection that should be applied at each stage to maximize profit while reducing the cost and labour of keeping individuals until later stages (Martinez *et al.,* 2006b).

Health and carcass traits are difficult to select for in fish and shellfish because phenotypic records are obtained from relatives and not from candidates for selection (Gjoen and Bentsen, 1997). Sib or pedigree evaluation has many disadvantages in relation to the amount of genetic progress that can be realized within a selection programme using only pedigree information to predict breeding values using an animal model. First, selection accuracy using sib information is lower than when predicting breeding values based on an individual's own information (Falconer and Mackay, 1996). Second, there is no variation of estimated breeding value for polygenic effects. Thus, variation of Mendelian sampling effects within a family cannot be used and consequently there may be a limited scope for constraining rates of inbreeding

to acceptable levels when the number of families is relatively low.

To date, little has been published regarding the economic profits arising from the extra genetic gain obtained by MAS or GAS schemes in aquaculture or terrestrial species. Information of this nature is essential because the additional gains are dependent on the magnitude of the allelic effects and thus the marginal increase should offset the costs of applying the technology. This trade-off may be more important when a single marked QTL, rather than multiple marked QTL (and multiple traits), is targeted by selection.

Pleiotropic effects can be important if the polymorphisms under MAS or GAS also have negative effects on fitness or other traits of economic importance. For example, negative genetic correlations have been found for resistance to viral and bacterial diseases (Gjøen *et al.,* 1997; Henryon *et al.,* 2002, 2005), which may be a problem in practical breeding when the goal is to select fish resistant to a range of pathogens. For example, in natural and selected populations, MHC polymorphism is likely to be maintained by frequency-dependent selection (Langefors, Lohm and Grahn, 2001; Lohm *et al.,* 2002; Bernatchez and Landry, 2003), suggesting that selection favours rare alleles, but works against the same alleles at high frequency. Therefore, it seems likely that a MAS scheme using MHC information or QTL in LD with disease resistance should focus on maintaining polymorphism rather than on selecting for a particular combination of alleles.

MAS in populations in linkage equilibrium

When populations are in LE between markers and QTL, the information used for selection purposes is given by the Mendelian co-segregation of markers and QTL within each of the full-sib families in the population under selection. In practical terms, this means that co-ancestry conditional on marker information needs to be computed within a family for a given segment in the genome. In effect, the segregation of regions that individuals share as identical-by-descent ("more" or "less" than average) is being traced and, under such circumstances, the accuracy of predicting breeding values using marker information is mainly dependent on the proportion of the within-family variance due to the QTL (Ollivier, 1998).

The effect of family size on the relative accuracy of predicting breeding values (comparing MAS and BLUP) using marker information was studied in detail using simulations (Table 2; V. Martinez, unpublished data). Compared with the GAS schemes presented below, for LE-MAS to be efficient, large full-sib families are required for predicting breeding values for the QTL accurately. This is because breeding value prediction is carried out on a within-family basis; thus, large families are required to obtain breeding values for predicting QTL effects with reasonable accuracy. When individuals do not have records for the quantitative trait, the extra accuracy of MAS was highest for the largest family size simulated (50 individuals, 25 with records and 25 without records; the difference is equal to 7 percent). The accuracy of predicting breeding values was very similar in BLUP or MAS for individuals that have records for the trait in most of the scenarios simulated, suggesting that MAS is expected to be of little use under these circumstances (Villanueva, Pong-Wong and Woolliams, 2002).

The advantage of MAS will come both from increased accuracy and from

TABLE 2
Empirical correlation between predicted breeding values using molecular* and pedigree information (M+BLUP) or pedigree information (BLUP) and true breeding values

Scenario	Individuals with records	Family size (number of families)					
		10 (100)		20 (50)		50 (20)	
		M+ BLUP	BLUP	M+ BLUP	BLUP	M+ BLUP	BLUP
I	NO	0.47	0.45	0.55	0.52	0.64	0.57
	YES	0.60	0.60	0.65	0.64	0.70	0.65
II	NO	0.41	0.41	0.49	0.47	0.56	0.52
	YES	0.58	0.58	0.62	0.61	0.64	0.63

* Molecular information comprises a completely informative marker bracket of 10 cM around a QTL and all individuals genotyped for the markers. The matrix of identity-by-descent values was calculated using the deterministic method of Martinez (2003). The estimated values of h^2 using residual effect maximum likelihood for the polygenic and QTL effects were, on average, 0.13 and 0.09, respectively. The results are presented for different nuclear family sizes (number of families, between parentheses) and for candidates with or without phenotypic records. The population size was equal to 1 000, where 50 percent (Scenario I) or 25 percent (Scenario II) of the individuals within each full-sib family had records for the trait.
Source: Martinez, unpublished data.

increasing the realized selection intensity in sustainable breeding schemes with restricted rates of inbreeding. In sib-testing schemes, candidates without records can only be selected randomly within families because an estimate of the Mendelian sampling terms cannot be obtained. Markers provide an estimate of the QTL effects that segregate within a family, and therefore the realized selection differential (at the same rates of inbreeding) is expected to be greater than that obtained using standard sib/family testing.

All the benefits outlined above come at an expense. MAS using LD within families requires a great deal of genotyping and recording of phenotypes on relatives, due to the fact that the linkage phase between markers and QTL needs to be re-estimated in each generation. This is because LD between markers and the QTL is established only within families in each generation and not across the population. For this reason, it is not possible to predict breeding values for the QTL using molecular marker data without records when exploiting information from a single generation. Therefore, pre-selection using this approach is more difficult to apply in practice. This means that for disease resistance or carcass quality traits, challenge (measurement) will have to be carried out at every generation, in all the families available within the programme, as is always the case for conventional breeding programmes.

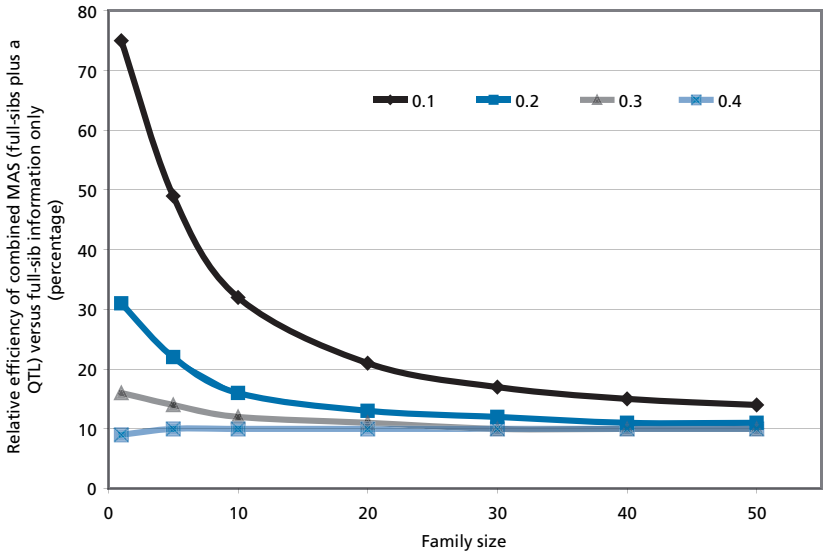
Due to the low resolution when mapping the QTL, it is likely that inaccurate estimates of position will lead to over-optimistic estimates of rates of genetic gain. In the simulations, it was assumed that the QTL position was known within the interval and the markers surrounding the QTL were completely informative. Thus, the increase in accuracy presented in Table 2 represents the upper bounds of accuracy estimates.

Utilizing direct test of genes in GAS schemes

The mean phenotype of the population for a quantitative trait can be modified by increasing the frequency of favourable alleles of genes influencing the trait. In the literature, greater genetic gain has been predicted for GAS schemes than for MAS schemes (using LE populations) at

FIGURE 5

Relative efficiency of combined GAS (for different family sizes [full-sibs]) and a known QTL, explaining 10 percent of the genetic variance) versus an index using information from full-sibs only for different values of the overall heritability (h^2)



The data were obtained from the ratio of square root of the variance of the indices. Selection index formulae were derived from Lande and Thompson (1990).

the same rate of inbreeding (Pong-Wong *et al.*, 2002). This is because the accuracy of predicting QTL effects using markers is always smaller than when the QTL effects are known, as in GAS schemes. In reality, it is likely that MAS will be carried out using information from many markers to predict the allelic effects of more than one QTL simultaneously whereas, in GAS schemes, only a limited number of polymorphisms are likely to be available. Therefore, on the whole, MAS schemes may yield greater genetic response because a greater proportion of the genetic variation is marked and used. Still, more marker genotyping is required for MAS schemes, which means that the additional proportion of the variance typed should pay for the increase in the cost of many markers typed simultaneously.

Due to the biology of many species in aquaculture, large family sizes can be used in a breeding programme. Following the deterministic model of Lande and Thompson (1990), Figure 5 describes the effect of family size and amount of polygenic variation on the relative efficiency of accuracy estimates for an index using different numbers of full-sibs measured for the trait, versus an index also including information on candidates for selection genotyped at loci targeted for GAS schemes (V. Martinez, unpublished results). For a single QTL explaining 10 percent of the genetic variance, when the heritability is relatively large, family size has a small impact on the accuracy. On the other hand, when the heritability of the trait is small, selection for a known QTL has a major impact on relative efficiency, particularly when the family size

is relatively small. Hence, this approach can be important for traits that are expensive or difficult to measure such as carcass quality, disease resistance or antibody response.

Given the research efforts carried out at diverse laboratories worldwide, it is likely that direct tests will be available in the near future for GAS schemes for different traits. With an increasing amount of data on ESTs, together with a greater understanding of the function of known genes in aquaculture species and new gene discovery, there is a possibility of more rapidly identifying and subsequently using polymorphisms that are within coding regions. However, the research effort required to develop tests for polymorphisms explaining allelic effects cannot be underestimated, and the factors influencing the profitability of GAS will include:

- the amount of variation explained by the test and the number of tests (genes) available for explaining the phenotype;
- the frequency of the favourable allele in, and the presence of the direct test (e.g. SNPs), for the target population;
- the interaction between the polymorphism and the background genome and possible pleiotropic effects on fitness;
- the trade-off between the marginal return given by the additional genetic gain obtained through the non-linear changes in the allele frequency of the favourable allele until fixation;
- fixed costs of implementing genotyping and patenting.

MAS in populations in LD

Using information from dense marker maps, it is possible to make use of LD between the markers and the beneficial mutations influencing the quantitative traits across the population. Under this scenario, there are two possible ways to use the LD in MAS programmes i.e. using information on a single haplotype effect in LD with the beneficial polymorphism across the population, or predicting the total genetic value using genome-wide, dense marker maps (genome-wide marker-assisted selection, or G-MAS) (Lande and Thompson, 1990; Meuwissen, Hayes and Goddard, 2001).

The effectiveness of each scenario is largely dependent on the actual magnitude of the effects associated with the polymorphism, either across the whole genome or at specific genes. It is likely that, in the near future, high-throughput SNP technology will make dense marker maps cost effective for selective breeding purposes in aquaculture. Thus, it can be expected that LD-MAS will be implemented over the whole genome, basically using markers to unravel the genetic architecture of quantitative traits. Information from multiple traits jointly and for multiple genes (and their interactions within and between loci) will be used, rather than first relying on mapping QTL in experimental populations and then implementing this information in MAS programmes. A profit analysis including multiple traits (e.g. to study undesirable pleiotropic effects on the breeding goal) will be needed on a case-by-case basis to determine whether the use of a single or multiple haplotypes simultaneously is most profitable and which method of LD-MAS better suits the population under selection.

Specific genes are not being evaluated when LD is used across the population; rather, haplotype effects on the phenotype are being estimated. As this is done on a single generation across the whole genome, it would be possible to use these haplotype effects for selecting candidates some generations after the initial estimation without relying on phenotypes (Meuwissen, Hayes and Goddard, 2001). Recombination will

erode the initial LD and therefore it is expected that accuracy of estimating the breeding value of many haplotypes will decay (Zhang and Smith, 1992), the extent of the erosion being dependent on several population parameters (Meuwissen, Hayes and Goddard, 2001). In practice, the response to selection obtained needs to be verified in each generation; thus, re-estimation can be used based on a random sample of individuals from the population.

One possible caveat is that by assuming a certain mode of gene action (i.e. only additive effects), there may in fact be a more complicated genetic architecture influencing quantitative traits. For example, when estimating dominance and epistasis with the same data, more haplotype effects need to be estimated. Therefore, it is likely that the accuracy of individual effects will decrease. Another potential complication that arises when the true model involves non-additive effects is that assignment of potential mates needs to be optimized to increase the mean phenotype of the population simultaneously through heterosis arising from combination of different QTL alleles. In the long term, the frequency of homozygotes that are identical-by-descent will increase within the population as a whole; consequently, methods are required to constrain the rates of inbreeding to obtain similar changes of the population mean across generations. Furthermore, expression of different combinations of alleles after selection will require re-estimation of between-haplotype effects in each generation.

CONCLUSION

QTL mapping and MAS are not as well advanced in aquaculture species as in terrestrial plants and animals. However, the merger between genetics and genomics is expected to be a fertile area of research in the coming years due to the plethora of information that is currently being gathered by many laboratories around the world. It is through these research efforts that variations affecting complex traits in fish and shellfish species may be detected and used for increasing the usefulness of MAS schemes. In the final analysis, however, all these techniques must be cost-effective if they are to be profitable in actual breeding programmes.

ACKNOWLEDGEMENTS

I am indebted for comments from Mr Eric Hallerman who greatly helped clarify the manuscript. This chapter has been funded partially by: INNOVA-Corporación de Fomento de la Producción (CORFO) (05CT6 PP-10) on "Using functional genomics for understanding disease resistance in salmon" from the Government of Chile; Vicerrectoria de Investigacion y Creación, Universidad de Chile hatch project 03/03312; and Fondo de Cienciias y Tecnologia (FONDECYT) 1061190.

REFERENCES

Anderson, E.C. & Thompson, E.A. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160: 1217–1229.

Arkush, K.D., Giese, A.R., Mendonca, H.L., McBride, A.M., Marty, G.D. & Hedrick, P.W. 2002. Resistance to three pathogens in the endangered winter-run Chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Can. J. Fisheries and Aquatic Sci.* 59: 966–975.

Bernatchez, L. & Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evolut. Biol*. 16: 363–377.

Bijma, P., van Arendonk, J.A.M. & Bovenhuis, J.H. 1997. Breeding value and variance component estimation from data containing inbred individuals: application to gynogenetic families in common carp.(*Cyprinus carpio* L.). *Genet.*145: 1243–1249.

Bongers, A.B., Bovenhuis, J.H., van Stokkom, A.C., Wiegertjes, G.F., Zandieh-Doulabi, B., Komen, J. & Ritcher, C.J.J. 1997. Distribution of genetic variance in gynogenetic or androgenetic families. *Aquaculture* 153: 225–238.

Brown, C.R., Woolliams, J.A. & McAndrew, B.J. 2005. Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture* 247: 219–225.

Butler, K., Field, C., Herbinger, C.M. & Smith, B.R. 2004. Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. *Mol. Ecol*. 13: 1589–1600.

Carlborg, Ö., De Koning, D.J., Manly, K.F., Chesler, E., Williams, R.W. & Haley, C.S. 2005. Methodological aspects of the genetic dissection of gene expression. *Bioinformatics* 21: 2383–2393.

Castro, J., Bouza, C., Presa, P., Pino-Querido, A., Riaza, A., Ferreiro, I., Sanchez, L. & Martinez, P. 2004. Potential sources of error in parentage assessment of turbot (*Scophthalmus maximus*) using microsatellite loci. *Aquaculture* 242: 119–135.

Christiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A. & Volckaert, F.A.M. 2005. A microsatellite linkage map of the European seabass, *Dicentrarchus labrax* L. *Genetics* 170: 1821–1826.

Cnaani, A., Hallerman, E.M., Ron, M., Weller, J.I., Indelman, M., Kashi, Y., Gall, G.A.E. & Hulata, G. 2003. Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F_2 tilapia hybrid. *Aquaculture* 223: 117–128.

Cnaani, A., Zilberman, N., Tinman, S., Hulata, G. & Ron, M. 2004 Genome-scan analysis for quantitative trait loci in an F_2 tilapia hybrid. *Mol. Genet. Genom*. 272: 62–172.

Coimbra, M.R., Kobayashi, M.K., Koretsugu, S., Hasegawa, O., Ohara, E., Ozaki, A., Sakamoto, T., Naruse, K. & Okamoto. N. 2003. A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* 220: 203–218.

Corley-Smith, G.E., Lim, C.H.J. & Brandhorst, B.P. 1996. Production of androgenetic zebrafish (*Danio rerio*). *Genetics* 142: 1265–1276.

Coulibaly, I., Gharbi, K., Danzmann, R., Yao, J. & Rexroad, C.E. 2005. Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. *Anim. Genet*. 36: 309–315.

Doyle, R.W. & Herbinger, C.M. 1994. The use of DNA fingerprinting for high intensity, within-family selection in fish breeding. *Proc. 5th Wrld. Congr. Genetics Appl. Livestock Prodn.*, Vol. 19, pp. 364–371. Guelph, Ontario.

Emery, W., Boyle, C. & Noble, F. 2001. Assignment of paternity groups without access to parental genotypes: multiple mating and development plasticity in squid. *Molec. Ecol*. 10: 1265–1278.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to quantitative genetics*, *4th edition.* Harrow, Essex, U.K, Longman.

Fernandez, J. & Toro. M. 2006. A new method to estimate relatedness using molecular markers. *Molec. Ecol*. (in press).

Garcia de Leon, F.J., Canonne, M., Quillet, E., Bonhomme, F. & Chatain, B. 1998. The application of microsatellite markers to breeding programs in the sea bass, *Dicentrarchus labrax*. *Aquaculture* 159: 303–316.

George, A., Visscher, P. & Haley, C.S. 2000. Mapping quantitative trait loci in complex pedigrees: a two step variance component approach. *Genetics* 156: 2081–2092.

Gilbey, J., Verspoor, E., McLay, A. & Houlihan, D. 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Anim. Genet*. 5: 98–105.

Gjedrem, T. 2000. Genetic improvement of cold-water fish species. *Aquaculture Res*. 31: 25–33.

Gjerde, B. 1986. Growth and reproduction in fish and shellfish. *Aquaculture* 57: 37–55.

Gjerde, B. & Schaeffer, L.R. 1989. Body traits in rainbow trout: II. Estimates of heritabilities and of phenotypic and genetic correlations. *Aquaculture* 80: 25–44.

Gjerde, B., Villanueva, B. & Bentsen, H.B. 2002. Opportunities and challenges in designing sustainable fish breeding programs. *Proc. 7th Wrld. Congr. Genetics Appl. Livestock Prodn.,* CD-ROM 06-01. Montpellier, France.

Gjøen, H. & Bentsen, H. 1997. Past present and future of genetic improvement in salmon aquaculture. *ICES J. Marine Sci*. 54: 1009–1014.

Gjøen, H.M., Refstie, T., Ulla, O. & Gjerde, B. 1997. Genetic correlations between survival of Atlantic salmon in challenge and field tests. *Aquaculture* 158: 277–288.

Griffing, B. 1967. Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Austr. J. Biol. Sci*. 20: 127–139.

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S. & Stet, R.J. 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogen*. 55: 210–219.

Hayes, B. & Goddard, M.E. 2001. The distribution of the effects of genes affecting quantitative traita in livestock. *Genet. Sel. Evol*. 33: 209–229.

Hayes, B., Lærdahl, J., Lien, S., Adzhubei, A. & Høyheim, B. 2004. Large scale discovery of single nucleotide polymorphism (SNP) markers in Atlantic salmon (*Salmo salar*). AKVAFORSK, Institute of Aquaculture Research (available at www.akvaforsk.no/norsk/HentFil.asp?fid=41).

Henryon, M., Jokumsen, A., Berg, P., Lund, I., Pedersen, P.B., Olesen, N.J. & Slierendrecht, W.J. 2002. Genetic variation for growth rate, feed conversion efficiency and disease resistance exists within a farmed population of rainbow trout. *Aquaculture* 209: 59–76.

Henryon, M., Berg, P., Olesen, N.J., Kjaer, T.E., Slierendrecht, W.J., Jokumsen, A. & Lund. I. 2005. Selective breeding provides an approach to increase resistance of rainbow trout (*Onchorhynchus mykiss*) to the diseases, enteric redmouth disease, rainbow trout fry syndrome and viral haemorrhagic septicaemia. *Aquaculture* 250: 621–636.

Hershberger, W.K., Myers, J.M., Iwamoto, R.N., Mcauley, W.C. & Saxton, A.M. 1990. Genetic changes in the growth of coho salmon (*Oncorhynchus kisutch*) in marine net-pens, produced by ten years of selection. *Aquaculture* 85: 187–197.

Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M. & Cook, D. 1995. DNA fingerprint-based analysis of paternal and maternal effects of offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137: 245–256.

Herbinger, C.M., O'Reilly, P.T., Doyle, R.W., Wright, J.M. & O'Flynn, F. 1999. Early growth performance of Atlantic salmon full-sib families reared in single family tanks versus in mixed family tanks. *Aquaculture* 173: 105–116

Houbert, S. & Hedgecock, D. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster (*Crassostrea gigas*) *Genetics* 168: 351–362.

Hubner, N., Wallace, C.A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., Mueller, M., Hummel, O., Monti, J., Zidek, V., Musilova, A., Kren, V., Causton, H., Game, L., Born, G., Schmidt. S., Muller, A., Cook, S.A., Kurtz, T.W., Whittaker, J., Pravenec, M. & Aitman, T.J. 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genet*. 37: 243–253.

Hulata, G. 2001. Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica* 111: 155–173.

Jamieson, A. & Taylor, S.T. 1997. Comparison of three probability formulae for parentage exclusion. *Anim. Genet*. 28: 397–400.

Jansen, R.C., Geerlings, H., Van Oeveren, A.J. & Van Schaik R.C. 2001. A comment on codominant scoring of AFLP markers. *Genetics* 158: 925–926.

Jansen, R.C. & Nap, J.P. 2001. Genetical genomics: the added value from segregation. *Trends Genet*. 17: 388–391.

Jiang, C. & Zeng, Z.B. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140: 1111–1127.

Kause, A., Ritola, O., Paananen, T., Ritola, O., Paananen, T., Mäntysaari, E. & Eskelinen, U. 2002. Coupling body weight and its composition: a quantitative genetic analysis in rainbow trout. *Aquaculture* 211: 65–79.

Kause, A., Ritola, O., Paananen, T., Wahlroos, H. & Mantysaari, E.A. 2005. Genetic trends in growth, sexual maturity and skeletal deformations and rate of inbreeding in a breeding programme for rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 247: 177–187.

Khoo, S.K., Ozaki, A., Nakamura, F., Arakawa, T., Ishimoto, S., Nickolov, R., Sakamoto, T., Akutsu, T., Mochizuki, M., Denda, I. & Okamoto, N. 2004. Identification of a novel chromosomal region associated with infectious hematopoietic necrosis (IHN) resistance in rainbow trout. *Fish Path*. 39: 95–102.

Kjøglum, S., Grimholt, U. & Larsen S. 2005. Non-MHC genetic and tank effects influence disease challenge tests in Atlantic salmon (*Salmo salar*). *Aquaculture* 250: 102–10.

Knott, S.A. & Haley, C.S. 2000. Multitrait least squares for quantitative trait loci detection. *Genetics* 156: 899–911.

Knudsen, S. 2002. *A biologist's guide to analysis of DNA microarray data*. New York, NY, USA, John Wiley and Sons.

Kocher, T. D., Lee, W. J., Sobolewska, H., Penman, D. & McAndrew, B. 1998. A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* 148: 1225–1232.

Lande, R. & Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743–756.

Langefors, J., Lohm, M. & Grahn, O. 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. R. Soc. Lond. B* 268: 479–485.

Lee, B.Y., Lee, W.J., Streelman, J.T., Carleton, K.L., Howe, A.E., Hulata, G., Slettan, A., Stern, J.E., Terai, Y. & Kocher, T.D. 2005. A second-generation genetic linkage map of tilapia (*Oreochromis* spp). *Genetics* 170: 237–244.

Li, Y., Byrne, K., Miggiano E., Whan, V., Moore, S., Keys, S., Crocos, P. & Lehnert, S. 2003. Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. *Aquaculture* 219: 143–156.

Liu, Y., Jansen, G.B. & Lin, C.Y. 2002. The covariance between relatives conditional on genetic markers. *Genet. Select. Evol*. 34: 657–667.

Liu, Z.J. & Cordes, J.F. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238: 1–37.

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A. & von Schantz, T. 2002. Experimental evidence for major histocompatibility complex allele-specific resistance to a bacterial infection. *Proc.R. Soc. Lond. B* 269: 2029–2033.

Lynch, M. & Ritland, K. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* 152: 1753–1766.

Marth, G.T., Korf, I., Yandell, M.D., Yeh, R.T., Gu, Z., Zakeri, H., Stitziel, N.O., Hillier, L., Kwok, P.Y. & Gish, W.R. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genet*. 23: 452–456.

Martinez, V. 2003. *Quantitative genetic analysis of fish populations using molecular markers*. University of Edinburgh, UK (Ph.D. thesis).

Martinez, V. 2006a. Using marker data in conventional breeding programmes: a case of study in selective breeding programs of scallops. In *Proc. Internat. Assocn. Genet. Aquaculture*, p 66. Montpellier, France.

Martinez, V. 2006b. Further insights of the variance component method to detect QTL in fish populations: relaxing the assumption of additive genetic effects. In *Proc. Internat. Assocn. Genet. Aquaculture,* p. 66. Montpellier, France.

Martinez, V. 2006c. Importance and implementation of molecular technologies in selective breeding programs for aquaculture species. In *Proc. 8th Wrld. Congr. Genet. Appl. Livest. Prodn*., Communication 09-01. Minas Gerais, Brazil.

Martinez, V. & di Giovanni. S. 2006. Breeding programmes of scallops: effect of self-fertilization when estimating genetic parameters. In *Proc. Internat. Assocn. Genet. Aquaculture*, p.65. Montpellier, France.

Martinez, V., Hill, W.G. & Knott, S.A. 2002. On the use of double haploids for detecting QTL in outbred populations. *Hered.* 88: 423–431.

Martinez, V., Neira, R. & Gall, G.A.E. 1999. Estimation of genetic parameters from pedigreed populations: lessons from analysis of alevin weight in coho salmon (*O. kisutch*). *Aquaculture* 330: 22–30.

Martinez, V., Sillanpaa, M., Thorgaard, G., Robison, B., Woolliams, J. & Knott, S. 2002. Evidence of a pleiotropic QTL influencing components of early development in double haploid lines of rainbow trout. *Proc. 7th Wrld. Congr. Genet. Appl. Livestock Prodn*., CD ROM Communication 06-08. Montpellier, France.

Martinez, V., Thorgaard, G., Robison, B. & Sillanpää, M. 2005. Simultaneous consideration of environmental and genetic effects in QTL mapping: an application to early development in double haploid lines of rainbow trout. *Genet. Res*. 86: 209–221.

Martinez, V., Kause, A., Mäntysaari, E. & Mäki-Tanila, A. 2006a. The use of alternative breeding schemes to enhance genetic improvement in rainbow trout: I. One-stage selection. *Aquaculture* 254: 182–194.

Martinez, V., Kause, A., Mäntysaari, E. & Mäki-Tanila, A. 2006b. The use of alternative breeding schemes to enhance genetic improvement in rainbow trout: II. Two-stage selection. *Aquaculture* 254: 195–202.

McCune, A.R., Fuller, R.C., Aquilina, A.A., Dawley, R.M., Fadool, J.M., Houle, D., Travis, J. & Kondrashov, A.S. 2002. A low genomic number of recessive lethals in natural populations of bluefin killifish and zebrafish. *Science* 296: 2398–2401.

Meuwissen, T.H.E. 1997. Maximizing the response of selection with predefined rate of inbreeding. *J. Anim. Sci*. 75: 934–940.

Meuwissen, T.H.E. 2003. Genomic selection the future of marker-assisted selection and animal breeding. In *Marker-assisted selection: a fast track to increase genetic gain in plant and animal breeding?* (available at www.fao.org/biotech/torino.htm).

Meuwissen, T.H.E. & Goddard, M.E. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Select. Evol*. 33: 605–634.

Meuwissen, T.H.E. & Goddard, M.E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Select. Evol.* 36: 261–279.

Meuwissen, T.H.E. & Woolliams, J.A. 1994. Effective sizes of livestock populations to prevent a decline in fitness. *Theor. Appl. Genet*. 89: 1019–1026.

Meuwissen, T.H.E., Hayes, B. & Goddard, M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genet*. 157: 1819–1829.

Moav, R. & Wohlfarth, G. 1976. Two-way selection for growth rate in the common carp (*Cyprinus carpio L.*). *Genet.* 82: 83–101.

Moen, T., Hoyheim, B., Munck, H. & Gomez-Raya, L. 2004a. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Anim. Genet*. 35: 81–92.

Moen, T., Kjersti, K.T., Fjalestad, H. & Gomez-Raya, L. 2004b. A multistage testing strategy for detection of quantitative trait loci affecting disease resistance in Atlantic salmon. *Genet.* 167: 851–858.

Moen, T., Agresti, J.J., Cnaani, A., Moses, H., Famula, T.R., Hulata, G., Gall, G.A.E. & May, B. 2004c. A genome scan of a four-way tilapia cross supports the existence of a quantitative trait locus for cold tolerance on linkage group 23. *Aquacult. Res*. 35: 893–904.

Morton, R. & Howarth, J.M. 2005. Estimation of breeding values using selected pedigree records. *Biomet.* 61: 422–432.

Mousseau, T.A., Ritland, K. & Heath, D.D. 1998. A novel method for estimating heritability using molecular markers. *Hered*. 80: 218–224.

Muir, W.M. 2005. Incorporation of competitive effects in forest tree or animal breeding programs. *Genet*. 170: 1247–1259.

Muranty, H. 1996. Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Hered*. 76: 156–165.

Ng, S.-H., Chang, S., Brown, A., Koop, B. & Davidson, W.S. 2005. Type I microsatellite markers from Atlantic salmon (*Salmo salar*) expressed sequence tags. *Mol. Ecol. Notes* 5: 762–766.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies I., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A. & Thorgaard, G.H. 2003. A consolidated genetic linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim. Genet*. 34: 102–115.

Norris, A.T., Bradley, D.G. & Cunningham, E.P. 2000. Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. *Aquaculture* 182: 73–83.

O'Brien, S.J. 1991. Molecular genome mapping: lessons and prospects. *Curr. Opin. Genet. Devel*. 1: 105–111.

Ollivier, L. 1998. The accuracy of marker-assisted selection for quantitative traits within populations in linkage equilibrium. *Genet.* 148: 1367–1372.

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M.R., Akutsu, T. & Okamoto, N. 2001. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genom*. 265: 23–31.

Panitz, F., Holm, L.-E., Høyheim, B., Hagen-Larsen, H., Taggart, J., Cairney, M., Davey, G.C., Powell, R. & Bendicen, C. 2002. EST sequencing in Atlantic salmon. *Proc. 7th Wrld. Congr. Genet. Appl. Livestock Prodn*. CD-ROM Communication No 06-14. Montpellier, France.

Pante, M.J.R, Gjerde, B. & McMillan, I. 2001. Effect of inbreeding on body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 192: 201–211.

Pante, M.J.R., Gjerde, B., McMillan, I. & Mistzal, I. 2002. Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, *Oncorhychus mykiss*. *Aquaculture* 204: 383–392.

Paterson, S., Piertney, S.B. & Knox, D. 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar L.*) microsatellites. *Mol. Ecol. Notes* 4: 160–162.

Pérez-Enciso, M. 2003. Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: a Bayesian unified framework. *Genet*. 163: 1497–1510.

Pérez-Enciso, M. 2004. In silico study of transcriptome genetic variation in outbred populations. *Genet.* 166: 547–554.

Pérez-Enciso, M., Toro, M.A., Tenenhaus, M. & Gianola. D. 2003. Combining gene expression and molecular marker information for mapping complex trait genes: a simulation study. *Genet*. 164: 1597–1606.

Pfrender, M.E., Spitze, K., Hicks, J., Morgan, K.L. & Lynch, L.M. 2000. Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. *Conserv. Genet*. 1: 263–269.

Piepho, H-P. & Koch, G. 2000. Codominant analysis of banding data from a dominant marker system by normal mixtures. *Genet*. 155: 1459–1468.

Pong-Wong, R., Villanueva, B. & Woolliams, J.A. 2002. Comparison of direct and marker-assisted selection with optimised contributions. In *Proc. 7th Wrld. Congr. Genet. Appl. Livestock Prodn*. CD-ROM Communication No. 33-02. Montpellier, France.

Pong-Wong, R., George, A.W., Woolliams, J.A. & Haley C.S. 2001 A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genet. Select. Evol.* 33: 453–471.

Ponzoni, R.W., Hamzah, A., Tan, S. & Kamaruzzaman, N. 2005. Genetic parameters and response to selection for live weight in the GIFT strain of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 247: 203–210.

Queller, D.C. & Goodnight, K.F. 1989. Estimating relatedness using genetic markers. *Evol.* 43: 258–275.

Quinton, C., McMillan, I. & Glebe, B.D. 2005. Development of an Atlantic salmon (*Salmo salar*) genetic improvement program: Genetic parameters of harvest body weight and carcass quality traits estimated with animal models. *Aquaculture* 247: 211–217.

Reid, D.P., Szanto, A., Glebe, B., Danzmann, R.G. & Ferguson, M.M. 2005. QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Hered.* 94: 166–172.

Rexroad, C.E., Rodriguez, M.F., Gharbi, K., Danzmann, R.G., DeKoning, J., Phillips, R. & Palti, Y. 2005. Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). *BMC Genom*. 6: 54.

Rise, M.L., Von Schalburg, K.R., Brown, G.D., Mawer, M.A., Devlin, R.H., Kuipers, N., Busby, M., Beetz- Sargent, M., Alberto, R., Gibbs, A.R., Hunt, P., Shukin, R., Zeznik, J.A., Nelson, C., Jones, S.R., Smailus, D.E., Jones, S.J., Schein, J.E., Marra, M.A., Butterfield, Y.S., Stott, J.M., Ng, S.H., Davidson, W.S. & Koop, B.F. 2004a. Development and application of a salmonid EST database and cDNA microarray:data mining and interspecific hybridization characteristics. *Genome Res*. 14: 478–490.

Rise, M.L., Jones, S.R., Brown G.D., von Schalburg, K.R., Davidson, W.S. & Koop, B.F. 2004b. Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiol. Genom*. 20: 21–35.

Robison, B., Wheeler, P. & Sundin, K. 2001. Composite interval mapping reveals a QTL of major effect on embryonic development rate in rainbow trout (*Onchorhynchus mykiss*). *J. Hered*. 92: 16–22.

Robison, B., Wheeler, P. & Thorgaard, G. 1999. Variation in development rate among clonal lines of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173: 131–141.

Rodriguez, F.M., LaPatra, S., Williams, S., Famula, T. & May, B. 2004. Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* 241: 93–115.

Rye, M. & Mao, I.L. 1998. Nonadditive genetic effects and inbreeding depression for body weight in Atlantic salmon (*Salmo salar* L.). *Livest. Prodn. Sci*. 57: 23–32.

Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., Okamoto, N., Ferguson, M.M., Holm, L-E., Guyomard, R. & Hoyheim, B. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genet*. 155: 1331–1345.

Serapion, J., Waldbieser, G.C., Wolters, W. & Wolters, Z.J. 2004. Development of Type I markers in channel catfish through intron sequencing. *Anim. Genet.* 35: 463–466.

Singer, A., Perlman, H., Yan, Y., Walker, C., Corley-Smith, G., Brandhorst, B. & Postlethwait, J. 2002. Sex-specific recombination rates in zebrafish (*Danio rerio*). *Genet*. 160: 649–657.

Smith, B.R., Herbinger, C.M. & Merry, H.R. 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. *Genet*. 158: 1329–1338.

Somorjai, I.M.L., Danzmann, R.G.& Ferguson, M.M. 2003. Distribution of temperature tolerance quantitative trait loci in Arctic charr (*Salvelinus alpinus*) and inferred homologies in rainbow trout (*Oncorhynchus mykiss*). *Genetics* 165: 1443–1456.

Streisinger, G.,Walker, C.H., Dower, N., Knauber, D. & Singer, F. 1980. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* 291: 293–296.

Sun, X. & Liang, L. 2004. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* 238: 165–172.

Sundin, K., Brown, K.H., Drew, R.E., Nichols, K.M., Wheeler, P.A. & Thorgaard, G.H. 2005. Genetic analysis of a development rate QTL in backcrosses of clonal rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 247: 75–83.

Tanck, M.W.T., Palstra, A.P., van de Weerd, M., Leffering, C.P., van der Poel, J.J., Bovenhuis, H. & Komen, J. 2001. Segregation of microsatellite alleles and residual heterozygosity at single loci in homozygous androgenetic common carp (*Cyprinus carpio* L.). *Genome* 44: 743–751.

Tao, W.J. & Boulding, E.G. 2003. Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.). *Hered.* 91: 60–69.

Taris, N., Baron, S. & Sharbel, T.F. 2005. A combined microsatellite multiplexing and boiling DNA extraction method for high-throughput parentage analyses in the Pacific oyster (*Crassostrea gigas*). *Aquaculture Res*. 36: 516–518.

Thomas, S. C., Pemberton, J.M. & Hill, W.G. 2000. Estimating variance components in natural populations using inferred relationships. *Hered*. 84: 427–436.

Thomas, S.C. & Hill, W.G. 2000. Estimating quantitative genetic parameters using sib-ships reconstructed from marker data. *Genet*. 155: 1961–1972.

Thorgaard, G.H. & Allen S.K. 1987. Chromosome manipulation and markers in fishery management. *In* N. Ryman & F.M. Utter, eds. *Population genetics and fishery management*, pp. 319–331. Seattle, WA, USA, University of Washington Press.

Toro, M. & Mäki-Tanila, A. 1999. Establishing a conservation scheme. *In* J.K. Oldenbroek, ed. *Gene banks and the conservation of farm animal genetic resources,* pp.75–90. Lelystad, Netherlands, DLO Institute of Animal Health.

Toro, M., Barragán, C. & Óvilo, C. 2002. Estimation of coancestry in Iberian pigs using molecular markers. *Conserv. Genet*. 3: 309–320.

Toro, J.E., Sanhueza, M.A., Winter, J.E., Aguila, P. & Vergara, A.M. 1995 Selection response and heritability estimates for growth in the Chilean oyster *Ostrea chilensis* (Philippi, 1845). *J. Shellfish Res*. 14: 87–92.

Vasemägi A., Nilsson, J. & Primmer, C.R. 2005. Expressed sequence tag (EST)-linked microsatellites as a source of gene associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol. Biol. & Evol*. 22: 1067–1076.

Villanueva, B.,Verspoor, E. & Visscher, P.M. 2002. Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. *Anim. Genet*. 33: 33–41.

Villanueva, B., Pong-Wong, R. & Woolliams, J.A. 2002. Marker assisted selection with optimised contributions of the candidates to selection. *Genet. Select. & Evol*. 34: 679–703.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van De Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. & Zabeau, M. 1995. AFLP: A new technique for DNA fingerprinting. *Nucl. Acids Res*. 23: 4407–4414

Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J. & Wolters, W.R. 2001. A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. *Genet.* 158: 727–734.

Walsh, B. & Henderson, D.J. 2004. Microarrays and beyond: What potential do current and future genomics tools have for breeders? *Anim. Sci*. 82: E292–E299.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. *Genet.* 166: 1963–1979.

Wang, L., Song, L., Chang, Y. & Xu, W. 2005. A preliminary genetic map of Zhikong scallop (*Chlamys farreri,* Jones et Preston 1904). *Aquaculture Res*. 36: 643–653.

Ward, R D., English, L. & McGoldrick, J. 2000. Genetic improvement of the Pacific oyster, *Crassostrea gigas* (Thunberg) in Australia. *Aquaculture Res.* 31: 35–44.

Withler, R. & Beacham, T. 1984. Genetic consequences of the simultaneous or sequential addition of semen from multiple males during hatchery spawning of chinook salmon (*Oncorhynchus tshawytscha*). *Aquaculture* 126: 11–20.

Wright, J.M. & Bentzen, P. 1994. Microsatellites – genetic markers for the future. *Rev. Fish Biol. & Fisheries* 4: 384–388.

Wu, W.-R., Li, W.-M., Tang, D.-Z., Lu, H.-R. & Worland, A.J. 1999. Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genet.* 151: 297–303.

Xu, S. & Atchley, W. 1995. A random model approach to interval mapping of quantitative trait loci. *Genet.* 141: 1189–1197.

Xu, S. & Gessler, D. 1998. Multipoint genetic mapping of quantitative trait loci using a variable number of sibs per family. *Genet. Res*. 71: 73–83.

Young, W.P., Wheeler, P.A., Fields, R.D. & Thorgaard, G.H. 1996. DNA fingerprinting confirms isogenicity of androgenetically-derived rainbow trout lines. *J. Hered.* 87: 77–81.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P. & Thorgaard, G.H. 1998 A detailed linkage map of rainbow trout produced using doubled haploids. *Genet.* 148: 839–850.

Yue, G.H., Ho., M.Y. & Orban, L. 2004. Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* 234: 85–98.

Zhang, W. & Smith, C. 1992. Computer simulation of marker assisted selection utilizing linkage disequilibrium. *Theor. Appl. Genet*. 83: 813–820.

Zimmerman, A.M., Wheeler, P.A., Ristow, S.S. & Thorgaard, G.H. 2005. Composite interval mapping reveals three QTL associated with pyloric caeca number in rainbow trout. *Aquaculture* 247: 85–95.

SECTION VI

Selected issues relevant to applications of marker-assisted selection in developing countries

CHAPTER 18

Marker-assisted selection in crop and livestock improvement: how to strengthen national research capacity and international partnerships

Maurício Antônio Lopes

SUMMARY

It is generally recognized that marker-assisted selection (MAS) is a tool that breeders can use to accelerate the speed and precision of crop and livestock breeding in developing countries. However, its practical application has been more difficult than previously expected. Although advances in molecular marker technology have uncovered many possibilities for transferring genes into desired crops and livestock through MAS, more methodological development and better planning and implementation strategies will be needed for its successful and expeditious application to breeding programmes. Also, this technology should not be regarded as an end in itself, but as an interacting part of complex strategies and decision-making processes. An appropriate mix of technologies and capabilities together with effective approaches to networking must be viewed as key ingredients for its correct development and application to breeding programmes. This chapter describes some strategies to guide decisions about structures, methods and capacities that may contribute to enhancing the access and successful use of MAS in developing countries.

INTRODUCTION

The tremendous advances made in molecular marker techniques in the past two decades have led to increased understanding of the genetic basis of many agricultural traits in a variety of plant and animal species. The use of these techniques has also made it possible to accelerate the transfer of desirable traits among varieties and to introgress novel genes from related wild species.

DNA markers have many advantages over conventional approaches available to breeders. They are especially advantageous for traits that are otherwise difficult to tag, such as resistance to pathogens, insects and nematodes, tolerance to abiotic stresses and quality parameters. They offer great scope for improving the efficiency of conventional breeding by carrying out selection not directly on the trait of interest but on linked genomic regions. Additionally these markers are unaffected by environmental conditions and are detectable during all stages of growth (Mohan *et al*., 1997).

Molecular marker techniques have therefore moved beyond their early projected role as tools for identifying chromosomal segments and genes to uncovering many possibilities for easing the transfer of genes into desired cultivars and lines. MAS generated great enthusiasm as it was seen as a major breakthrough, promising to overcome many limitations of conventional breeding processes (FAO, 2003). However, despite advances in the theory of MAS, direct utilization of the information it provides for selecting superior individuals with complex traits is still very limited (Young, 1999; Ferreira, 2003). Nevertheless, there is still optimism about the contributions of MAS, which is now balanced by the realization that genetic improvement of quantitative traits using this tool may be more difficult than previously considered (FAO, 2003). In 1999, Young reviewed the development of MAS, analysing in detail its main drawbacks, many of which remain today. He concluded that because MAS technology was so challenging it should not be a reason for discouragement but, instead, reason for more ingenuity and better planning and execution.

Recent developments in high-throughput genotyping, single nucleotide polymorphism (SNP) and the integration of genomic technologies are advances that will play an important role in the development of MAS as an effective tool for sustainable conservation and increased use of crop genetic resources (Ferreira, 2006). However, research teams, funding agencies, commodity groups and the private sector will need to work together to develop MAS technology further and ensure that breeders have the best available tools. Also, the tools and strategies will need to go beyond markers themselves to include genome-based knowledge derived from model systems, high-throughput cost effective technology, as well as better technologies and strategies for handling large volumes of information.

The purpose of this chapter is to discuss the access to and utilization of MAS technology by breeding programmes, with special emphasis on strategies to help strengthen research capacity and partnerships in developing countries. Whenever possible, recommendations are presented to help guide decisions that may contribute to enhancing the access and successful use of MAS by national programmes.

PERCEPTIONS ABOUT THE USE OF MAS IN CROP AND LIVESTOCK IMPROVEMENT

As MAS is still an evolving technology, there are not many detailed studies available describing the state-of-the-art of its

application to breeding programmes. Also, there are very few prospective studies indicating future trends in the application of this technology. The FAO Biotechnology Forum hosted an e-mail conference on "Molecular marker-assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries". This provided a comprehensive overview of the perceptions of scientists from different parts of the world about key aspects of the application of MAS to genetic improvement in developing countries (www.fao.org/biotech/logs/c10logs.htm).

As described in Chapter 21, this FAO conference was very inclusive, with a total of 627 people subscribing. Eight percent of these (52 people) submitted 85 messages, which were received from all major regions of the world, including Asia (33 percent), Europe (26 percent), Latin America and the Caribbean (14 percent), Africa (9 percent) Oceania (9 percent) and North America (8 percent). People from 26 different countries participated, with a total of 50 messages (59 percent) from developing countries and 35 messages (41 percent) from developed countries. Institutional representation was also ample, including national research institutes, centres belonging to the Consultative Group on International Agricultural Research (CGIAR), universities, consultants, farmer organizations, government agencies, non-governmental organizations (NGOs), etc. Although only 52 people out of 627 subscribers participated directly in the conference, the number is significant considering the broad representation, the high level of the (moderated) discussions and the number of relevant issues discussed (www.fao.org/biotech/logs/c10logs.htm).

To prepare this chapter, a detailed review was carried out of the conference results in an attempt to capture the main perceptions and concerns related to access to and utilization of MAS in developing countries. This analysis revealed a variety of ideas and creative suggestions to overcome the problems of MAS utilization. Although there is a risk of narrowing views on important issues discussed during the conference, four major perceptions were clear from the rich content of the discussions:

Perception 1. There is a need for development of priority-setting mechanisms and cost benefit analysis to guide informed decisions on how best to apply MAS and other technological innovations to crop and livestock breeding in developing countries.

Perception 2. MAS has to be understood as part of a complex process. Complementarities, mix of technologies, integration of capabilities and networking must always be viewed as key ingredients for its correct application in breeding programmes.

Perception 3. There is a need for an objective definition of public-private functions and responsibilities in relation to funding and development of technological innovation in developing countries. Public-private and north-south partnerships are essential to accelerate progress and effective application of MAS and other innovations to breeding programmes in developing countries.

Perception 4. Developing countries must focus on capacity building and human resource development oriented to shape effective strategies of technological innovation.

In the following sections, possible strategies and alternatives to deal with the challenges and opportunities indicated above are outlined, including the need

for objective priority-setting, development of partnerships, complementarities and capacity building for compatible human resource formation.

MAS AS PART OF A COMPLEX PROCESS – SETTING PRIORITIES AND TAKING ACTION

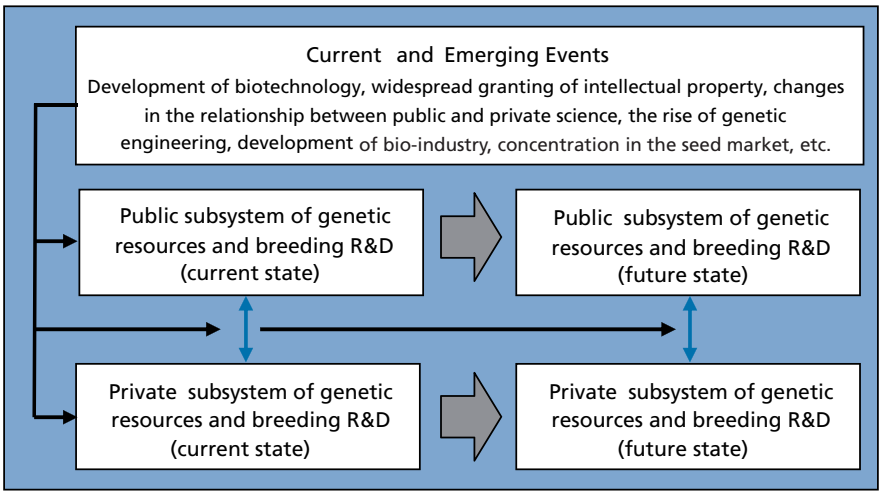
Before discussion of MAS as a technological alternative to increase the capacity of breeding programmes it is important to discuss and consider the future of the breeding process itself. Until recently, selection was based on observable phenotypes, without knowledge of the genetic architecture of the selected characteristics (Dekkers and Hospital, 2002). However, advances in molecular marker techniques and rapid advances in large-scale sequencing are creating new perspectives for exploiting the immense reservoir of polymorphism in genomes. Molecular genetic analysis of traits in plant and animal populations is leading to a better understanding of quantitative trait genetics. More recently, the discovery and scoring of single nucleotide polymorphisms (SNPs) using automated and high-throughput instrumentation are already providing the increased resolution needed to analyse sets of genes involved in complex quantitative traits (Altshuler *et al*., 2000; De La Vega *et al.*, 2002; Rafalsky, 2002, Lörz and Wenzel, 2005; Ferreira, 2006).

What impacts will all these developments have on breeding programmes? As anticipated by Stuber, Polacco and Senior, in 1999, "when genomics is added to future strategies for plant and animal breeders, the projected outcomes are mind-boggling. There is every reason to believe that the synergy of empirical breeding, MAS and genomics will truly produce a greater effect than the sum of the various individual actions." Despite the positive view of many who find technological development an open venue for enhancement or complete redesign of traditional breeding, there are many uncertainties about its future. The rise of genetic engineering and the bio-industry, and the widespread granting of intellectual property rights, followed by profound changes in the relationship between public and private science make it very difficult to anticipate future developments in both publicly funded breeding research and the commercial biotechnology industry.

Unfortunately, very little effort has been directed to thinking about the future of breeding, especially in developing countries (Castro *et al*., 2002, 2006). Many past and current events are changing the performance, the relationships and the space that public and private research organizations have in the market, raising the need for a deeper understanding of their unfolding impacts on the public activity of research (Price, 1999; Graff *et al*., 2003). The current scenario of changes and uncertainties has generated the necessity for strategic re-alignment of public research in many parts of the world. Therefore, research organizations need information that is not currently available about such changes and influences and their impact on the future of key activities, such as crop and livestock breeding. To obtain and to organize this information, prospective studies need to be developed on the present and future performance of breeding programmes and their related production systems.

The future configuration of breeding programmes depends on knowledge to guide strategic decisions about structures, methods and capacities in order to take advantage of new opportunities and technological niches. Foresight methodologies have been applied to this end, using systemic analysis of the past and

FIGURE 1
Conceptual framework of a prospective study on genetic resources and breeding R&D in Brazil



Dark arrows indicate the impact of current and emerging events on both public and private subsystems of R&D in genetic resources and breeding, at present and in the future, considering several alternative scenarios. Vertical arrows indicate the state of the relationship between the public and the private R&D subsystems as it is affected by current and emerging events.
Source: Castro *et al.,* 2002, 2006.

present performance of a research field, determining critical factors of performance (Linstone and Turoff, 1975; Castro, de Cobbe and Goedert, 1995; Castro, de Lima and Freitas Filho, 1998; Castro *et al*. 2002, 2006; Lima *et al*., 2000).

An innovative model of a prospective study was proposed and tested by Castro *et al*. (2002, 2006), based on the Brazilian national system of genetic resources and breeding. The effort started with the distinction between two component subsystems – public and private. The authors considered that the two subsystems admit two possible states or situations, current and future, after the effect of current and emerging events (Figure 1). Prospective efforts based on this framework can be very useful to guide diagnosis of national programmes, identifying the main determinants of current and past system performance that can be used to guide decisions about the configuration of genetic resources, breeding programmes and the associated seed industry.

This type of study can help identify changes in the system and in the corresponding technology market, analysing their current and future impacts, determining future opportunities and threats to the strategic positioning of research organizations in the technology market. There is also the perspective of developing possible alternative scenarios for the relationships between public and private research, and of these with the market, to guide the strategic positioning of public research. Results of this effort could indicate new opportunities and niches for public breeding programmes, as well as areas of extreme value where the public sector would have to acquire capacity in the future. Key decisions on investments in new technologies and processes applied

to genetic resources and breeding research, such as MAS, genomic tools, transgenic technology and others, are better taken if these results are available.

The results of this forward-looking approach developed in Brazil allowed the identification of some important trends that must be considered by managers in the process of adapting breeding efforts for the future (Castro *et al.*, 2002, 2005, 2006). Current and emerging events identified in the process will certainly affect the performance, methods, technological processes, portfolio of products and institutional relations in the public and private R&D sectors dedicated to plant breeding in Brazil. This complexity indicates that it is quite dangerous for developing countries, pressured by market evolution and rapid expansion of methods and technologies, to face the challenge of identifying priority areas for investment without a minimum prospective effort.

In summary, the ability to predict changes that might affect the performance of public and private R&D organizations is essential for decision-makers and managers to guide adjustments in the focus of these sectors in a timely manner, avoiding threats and promoting access to new tools and opportunities. Although the same prospective methodology may be applied to a wide range of countries, it is important to point out that situations differ drastically from country to country, thereby requiring examination of future configuration of a sector on a case-by-case basis.

MAS AS PART OF A COMPLEX PROCESS – BUILDING CAPACITIES, COMPLEMENTARITIES AND ENHANCING NETWORKING

MAS cannot be considered an end in itself or a tool detached from the complexities of breeding strategies. It has to be understood and analysed in the context of an interacting mix of tools and strategies that have to be targeted towards crop and livestock improvement in a coordinated manner. Independently of the outcome of any priority-setting effort, the need for an expanded networking approach to breeding and biotechnological research will always be an objective to be pursued. This need arises because networking and partnerships are essential to enable organizations to attain otherwise unattainable goals, add value to their products and processes and reduce costs. Also, the continuous demand for efficiency and relevance presses R&D programmes to move in the direction of cooperation and alignment of efforts.

One of the key problems limiting the use of MAS and other advanced technologies in developing countries is exactly the difficulty of building effective teams and networks. Unfortunately, very few developing countries have trained scientists and advanced facilities concentrated in one place or institution. Usually, these scarce resources are scattered over different places and institutions, and many times away or disconnected from the relevant breeding programmes. This is a serious drawback as the increasing interdependence of traditional and upstream disciplines makes it necessary to build and manage multidisciplinary teams consisting of breeders, agronomists, molecular biologists, biochemists, pathologists, entomologists, physiologists, soil scientists, statisticians, etc. – a goal always difficult to achieve. In addition to the challenge of working within team alignments and cooperation, there is the pressing need to develop ways to share capacities, infrastructure, materials and information among research teams located across a country, a region, or even continents.

The main problem in fostering collaboration and effective cooperation to achieve common goals seems to be the difficulty of recognizing that different teams and organizations have different general interests and norms. For this reason, competition usually prevails. While it has been well accepted that competition is one of the key forces that keep industry competitive and dynamic, this view is being challenged by the concept that many activities can benefit from a rational mix of competition and cooperation that leads to complementary products and expansion of possibilities through the formation of new relationships or even new modes of operation and management. Increasingly, the same is also true for R&D organizations, which can benefit from working with partners (competitors) whose abilities make their own more attractive in the eyes of clients (Brandenburger and Nalebuff, 1997). Also, faced with growing competition from industry and increasing pressures and demands, public R&D institutions must look at ways to do more with fewer resources. Collaboration through team nets and other networking strategies have the potential to reduce costs, add value and promote capacity to respond quickly to changes. Besides, with the new tools of information technology, collaboration with any part of the world is possible as this promotes information and other resource sharing without the need for geographical proximity (Lipnack and Stamps, 1993).

How should a R&D organization behave in a multifaceted relationship, when partners can be also competitors? Organizations that enter competitive collaboration must be aware that their partners may be out to disable them. This dilemma has been faced by a growing number of organizations, which rapidly understand that effectiveness will be more and more a product of recognizing and using interdependence. With networks and interdependent teams, cooperation must be designed in the name of mutual needs and with a clear sense of sharing risks to reach objectives that are common to all partners (Lopes, 2000).

In many parts of the world, including in developing countries like Brazil, competitive funding systems for agricultural R&D are assuming growing importance as new sources of funding and as drivers for cooperation among universities, R&D institutes and the private sector, in many cases allowing collaboration even among institutions that are traditional competitors (Lopes, 2000). Although the rules and procedures governing the competitive granting system indicate the need for partnership and the general mode of interaction, experience has shown that industry/university/R&D institutes cooperations succeed only if they are founded on trust and understanding and promise mutual benefits. Also, successful experiences have come from the clear recognition of objectives and well structured management with intense communication.

Two experiences are described below that rely heavily on cooperation and networking directed to effective application of advanced technologies, including MAS, to genetics and breeding. Both are excellent examples of strategies that promote effective partnerships and collaboration by researchers from different institutions, disciplines or countries working on specific high-priority projects.

The case of the CGIAR Generation Challenge Programme: an international R&D network in genetic resources, genomics and breeding

As the number of stakeholders in the agricultural decision-making process increases and

the agricultural research agenda expands, organizations must be able to respond to an increasingly diverse and complex portfolio of priorities by strengthening interactions within the system and developing links and partnerships with groups traditionally outside the system. Towards this end, the CGIAR has designed a strategy to nurture the definition of objective R&D agendas in key themes and to guide scientists and teams worldwide towards integrated, synergistic involvement and operation. This strategy became known as the "Global Challenge Programmes" (www.cgiar.org/impact/challenge/index.html).

The strategy of the Global Challenge Programmes recognizes both that the cost of conducting research is escalating and that the complexity of the science needed for agricultural research is increasing. Research in most fields requires not only specialized equipment and facilities but also highly trained technical support in diverse disciplines. Increasingly, multidisciplinary teams of scientists will be required to address the complex issues facing agriculture and, in many cases, the professional expertise needed may have to be accessed in different parts of the world.

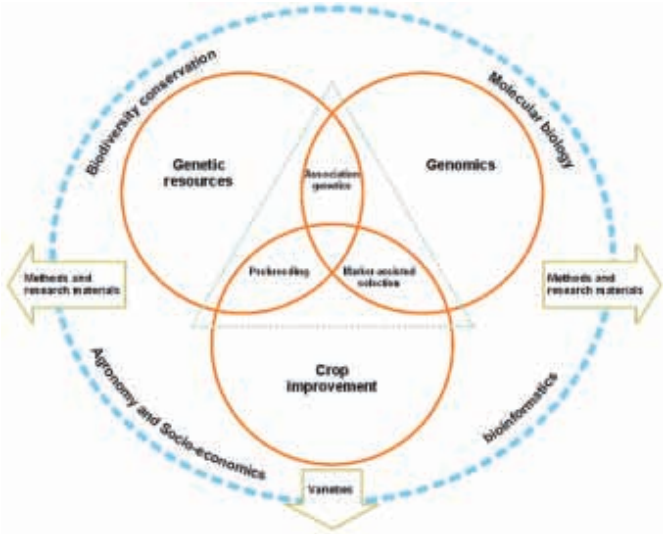
One such Challenge Programme, entitled "Unlocking Genetic Diversity in Crops for the Resource-Poor", also known as the "Generation Challenge Programme (GCP)" (CGIAR, 2003) is an international, multi-institutional, cross-disciplinary public platform for accessing and developing new genetic resources using advanced molecular technologies associated with conventional methods. Founded in July 2003 by the Executive Council of the CGIAR with start-up funding from the World Bank and the European Commission, the GCP has a membership of twenty-two public research institutions around the world, including nine CGIAR centres, four advanced research institutes and nine national agricultural research system institutions. Its budget in 2005 totalled at US$14 million (GCP, 2005).

This platform was designed to ensure that the advances of crop science and technology are applied to the specific problems and needs of resource-poor people who rely on agriculture for subsistence and their livelihoods. The GCP aims to "bridge that gap by using advances in molecular biology and harnessing the rich global stocks of crop genetic resources to create and provide a new generation of plants that meet these farmers' needs".

The concrete objective of the GCP is to access and develop genomic and genetic resources as enabling technologies and intermediate products for crop improvement programmes. It will not produce and release finished crop varieties for farmers, but develop new genetic resources and make the initial gene transfers to locally adapted germplasm, and then transfer the derived materials to crop improvement programmes, particularly those conducted in national agricultural research systems of developing countries, and to any other entities that have crop improvement goals, especially those dedicated to resource-poor farmers.

The GCP is, to date, the most comprehensive effort to cover, in a well structured and feasible manner, the complex interactions between genetic resources, genomics and breeding (Figure 2) in order to capture the benefits of the revolutions in biology and direct them to help solve some of the agricultural problems in the world's most difficult and marginal environments. It addresses its three key component parts in a separate but interconnected manner: (1) genetic resource collections provide the

FIGURE 2
Conceptual basis for the Generation Challenge Programme – Unlocking Genetic Diversity in Crops for the Resource-Poor



Reprinted by permission from the proposal for a CGIAR Challenge Programme (CGIAR, 2003).

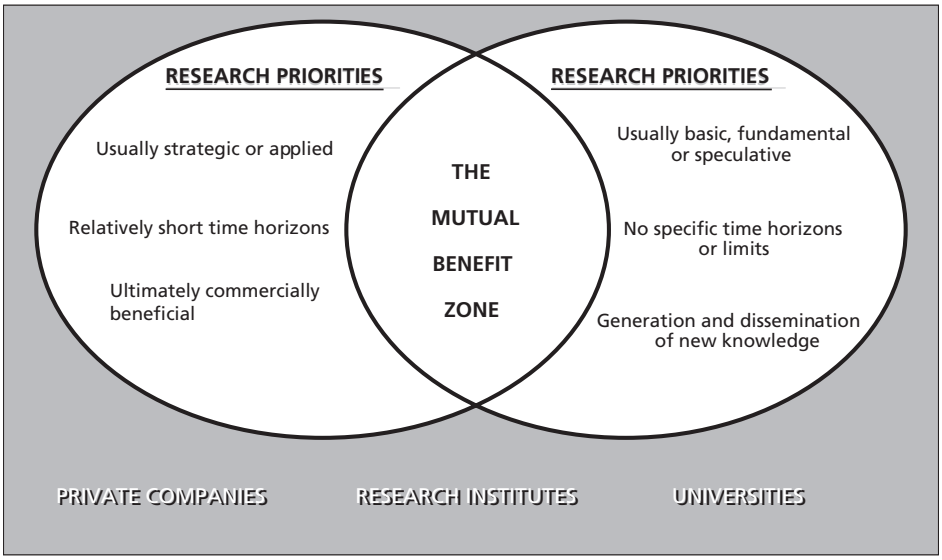
raw materials; (2) genomic science provides the means to exploit genetic resources (i.e. identify new alleles); and (3) crop improvement applies traditional and modern methods of gene/allele transfer into functional crop varieties (CGIAR, 2003).

The GCP is therefore an ambitious initiative to put into action a complex mix of tools, capacities, concepts and strategies. It is organized and managed to direct these resources towards the pursuit of goals that are not attainable through the disciplinary and isolated modes of operation that unfortunately prevail in the international agricultural R&D arena. As such, it is possibly the best structured international effort for development, adaptation and promotion of effective (and inclusive) access and use of tools such as MAS.

As part of its complex strategy, the GCP will define protocols for more efficient gene transfer including molecular markers that are closely linked to the genes for the desired trait, rapid tests for phenotype recognition, and genetic transformation of new genes into locally adapted genetic materials, such as improved varieties and landraces. All of these strategies depend on the adaptation and development of marker technology and marker-assisted procedures, hopefully helping to consolidate a networking approach to breeding and biotechnological research with effective impact, especially on resource-poor countries.

Research activities commenced in January 2005 with the first round of competitive research grants awarded for 17 three-year projects of approximately US$1 million each. In early 2005 a new round of commissioned grants was started, which served as the basis of the GCP platform of tools and technologies for genetic studies and applications. In total, the GCP initiated

FIGURE 3
Research priorities and partnerships among private companies, research institutes and universities



Source: adapted from Laider (1998).

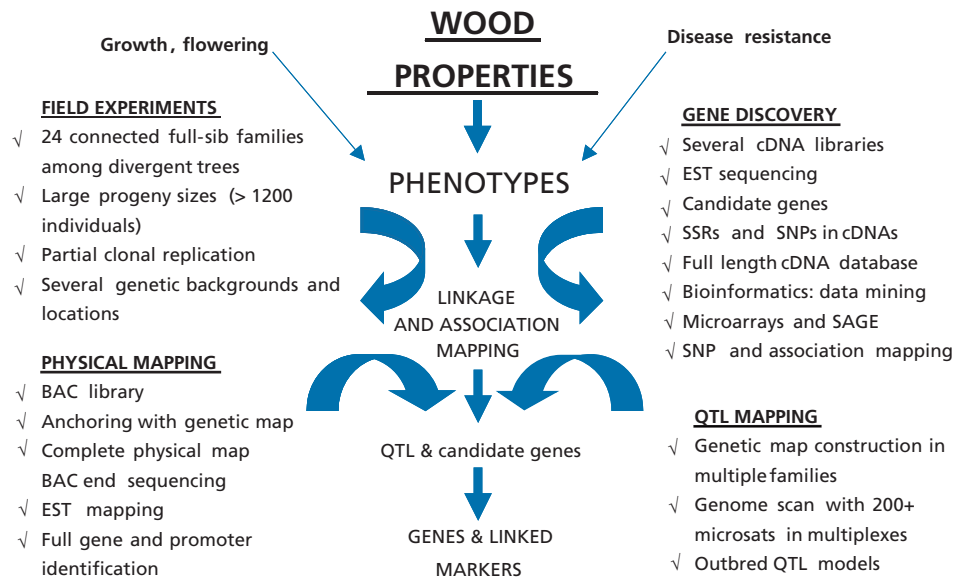
67 competitive and commissioned research projects and capacity-building activities in 2005. The 2005 Annual Report and 2006 Work Plan summarizes research progress and capacity building achievements in 2005 and presents an overview of the competitive and commissioned research portfolio and the capacity-building and delivery activities for 2006 (GCP, 2005, 2006).

The Case of the Genolyptus Programme in Brazil: a public/private network in genomics and molecular breeding of *Eucalyptus*

"The challenge and the opportunity for publicly supported agricultural research are not in duplicating the private sector's research agenda, but in building unique public/private partnerships or perhaps even jointly supported consortia for agricultural research" (CAST, 1994). Increasingly, agricultural research will be conducted through partnerships among private companies, public research institutes and universities (Figure 3). In forming such alliances, these organizations must recognize that developing productive relationships involves non-competitive dialogue and understanding of each others' abilities and limitations. Partnerships will flourish only if founded on trust and understanding and if differences in drivers and objectives are recognized and accommodated in initiatives with a real perspective of mutual benefits (Lopes, 2000).

An example of a successful public/private partnership with clear understanding of partners' abilities and limitations and clear definition of responsibilities and benefits to be pursued is the Genolyptus Network in Brazil (Grattapaglia, 2003). This R&D network was created to establish a foundation for a genome wide understanding of the molecular basis of wood formation in

FIGURE 4
The Genolyptus project – from phenotypes to genotypes in an integrated way



Source: Grattapaglia, 2003.

Eucalyptus, coupled with the translation of knowledge into improved tree breeding technologies.

This programme relies heavily on the development of aligned R&D efforts in genetic resources, genomics and molecular breeding (Figure 4). It mobilizes capacities and infrastructure in constructing physical maps, developing expressed sequence tag (EST) databases, generating a database of expression profiling of genes that control key traits and developing methods for MAS for traits of high heritability in wood formation. Also, the network develops a capacity-building and training programme for professionals in universities and forestry companies, targeting the integration of genetics, genomics and breeding efforts of *Eucalyptus*.

The rationale of the network is based on the understanding that, even with the more powerful tools allowing a much more global and integrated view of genetic processes, genomics will only succeed in contributing to the development of improved eucalypt if it is deeply interconnected with intensive fieldwork and creative breeding. The Genolyptus project therefore differs from other plant genome initiatives in the intensity, refinements and scope of the effort devoted to field experiments to generate the diversity of phenotypes necessary to study gene function. Quantitative trait loci (QTL) detection, the development of SNP haplotypes for association mapping and physical mapping will link the phenotypes to genes that control processes of wood formation that define industrial level traits (Grattapaglia, 2003).

A key feature of the Genolyptus network is its pre-competitive nature. The research programme was designed collaboratively with no immediate intention of marketing its results, even although its

planned outputs will eventually lead to the creation of many new products and processes of commercial value. Thus, the activities during its first phase are designed to resolve basic, common technological problems – a sufficient reason to mobilize several private companies (that are competitors in the market) and public organizations. After the first phase of six years, the network will have generated a consolidated base of knowledge and tools that will promote the development of specific interest projects, either in partnership or individually, according to specific business strategies and market targets.

Also, team organization and management are based on modern tools and concepts, involving a competent, highly respected scientist with talent to lead network operations, a steering committee and a technical committee for adequate planning, decision-making and follow-up, as well as contract models and negotiation strategies appropriate to the complexity of the network. Intellectual property rights provisions are based on access limited to participants, with all genetic materials and patents produced being co-owned by the 20 participating institutions. Scientific publications are highly encouraged.

As in the Generation Challenge Programme, the Genolyptus network is an original initiative to integrate and align a complex mix of tools, capacities, concepts and strategies. The ability to mobilize such a wide range of organizations, including 12 private companies operating in a highly competitive market space, indicates that the network design was successful, while its pre-competitive nature, organization and management strategy allowed the definition of a "zone of mutual benefits" (Figure 3), facilitating the pursuit of goals that are not attainable through isolated research. The Genolyptus network is therefore an excellent example of the feasibility of developing a structured public/private effort for integrated and effective use of advanced tools such as MAS.

CONCLUSIONS

- Although advances in molecular marker technology have uncovered many possibilities for easing the transfer of genes into desired crops and livestock through MAS, there is still limited recorded impact of these technologies in breeding programmes.
- It is generally recognized that genetic improvement of complex traits using MAS is more difficult than previously considered. Therefore, more methodology development, better planning and implementation strategies will be needed for its successful and rapid application to breeding programmes.
- The future configuration of breeding programmes is dependent on knowledge to guide strategic decisions about structures, methods and capacities that take advantage of new opportunities and technological niches. Unfortunately, there are very few efforts directed at thinking about the future of breeding programmes, especially in less developed countries. Research organizations need information, which is not currently available, about changes and influences and their impact in the future on key activities such as crop and livestock breeding. To acquire and organize this information, prospective studies on the present and future performance of breeding programmes and their related activities will have to be developed.
- Priority-setting strategies, together with cost–benefit analysis are necessary to guide informed decisions on how best

to apply MAS and other advanced technologies to crop and livestock breeding in developing countries.

- MAS has to be understood and undertaken as part of a complex process. Complementarities and a mix of technologies and capabilities, together with effective approaches to networking, must be viewed as key ingredients for its appropriate development and application to breeding programmes.
- One of the key problems limiting the use of MAS and other advanced technologies in developing countries is the difficulty of building effective teams and networks. Approaches to networking and partnerships are key to enabling organizations to attain new goals with less cost and to adding more value to their products and processes. Also, the demand for efficiency and relevance presses R&D programmes to move in the direction of cooperation and alignment of efforts.
- The present and future challenges and opportunities for agricultural research organizations are to build public/private partnerships or new types of consortia dedicated to innovation. In forming such alliances, these organizations must recognize that developing productive relationships involves non-competitive dialogue and understanding of each others' abilities and limitations. In order to survive and flourish, partnerships have to be sustained on trust and understanding.
- Developing countries must focus on training to build and shape capacities and effective strategies to support research in advanced biology applied to breeding. Also, new management strategies are needed to deal with the complex nature of modern breeding programmes.

REFERENCES

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L. & Lander, E.S. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513–516.

Brandenburger, A.J. & Nalebuff, B.J. 1997. Co-opetition, New York, USA, Doubleday.

CAST (Council for Agricultural Science and Technology). 1994. Challenges confronting agricultural research at land grant universities. Issue Paper 5. Ames, IA, USA.

Castro, A.M.G., de Cobbe, R.V. & Goedert, W.J. 1995. Manual de prospecção de demandas para o SNPA. Brasília, Brazil, Embrapa.

Castro, A.M.G., Lima, S.M.V. & Freitas Filho, A. 1998. Manual de Capacitación en Análisis de Cadenas Productivas. Brasília, Brazil, Embrapa.

Castro, A.M.G., Lima, S.M.V., Lopes, M.A. & Martins, M.A.A.G. 2002. Estratégia de P&D para o Melhoramento Genético em uma Época de Turbulência. *Anais do XXII simpósio de gestão da inovação tecnológica,* p.17. Salvador, BA, Brasil.

Castro, A.M.G., Lopes, M.A., Lima, S.M.V. & Bresciani, J.C. 2005. Trends and technological strategy in the Brazilian seed sector (Cenários do Setor de Sementes e Estratégia Tecnológica*). Revista de Política Agrícola Brasília* 3: 58–72.

Castro, A.M.G., Lima, S.M.V., Lopes, M.A., Machado, M.S. & Martins, M.A.G. 2006. O Futuro do Melhoramento Genético Vegetal no Brasil – Impactos da Biotecnologia e das Leis de Proteção de Conhecimento (The future of plant breeding in Brazil – impacts of biotechnology and intellectual property legislation). Brasília, DF, Embrapa Informação Tecnológica.

CGIAR (Consultative Group on International Agricultural Research). 2003. *Unlocking Genetic Diversity in Crops for the Resource-poor: A Proposal for a CGIAR Challenge Program* (available at www.generationcp.org/sccv10/sccv10_upload/CP_proposal.pdf).

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Revs. Genet*. 3: 22–32.

De La Vega F.M., Dailey D., Ziegle J., Williams J., Madden D. & Gilbert D.A. 2002. New generation pharmacogenomic tools: a SNP linkage disequilibrium map, validated SNP assay resource, and high-throughput instrumentation system for large-scale genetic studies. *Biotechniques* Suppl 32: S48–S54.

FAO. 2003. Molecular marker-assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries. *Background document to Conference 9 of the FAO Biotechnology Forum* (available at www.fao.org/biotech/C10doc.htm).

Ferreira, M.E. 2003. Melhoramento Genético do Arroz: Impactos da Genômica. *In* A. Borém, M. Giúdice & T. Sediyama, eds. *Melhoramento genômico.* Universidade Federal de Viçosa,Viçosa, Brazil.

Ferreira, M.E. 2006. Molecular analysis of gene banks for sustainable conservation and increased use of crop genetic resources. *In* J. Ruane & A. Sonnino, eds. *The role of biotechnology in exploring and protecting agricultural genetic resources*. FAO, Rome (available at www.fao.org/docrep/009/a0399e/a0399e00.htm).

GCP (Generation Challenge Programme). 2005. *Generation Challenge Programme research highlights 2005.* Mexico, DF. CIMMYT (available at http://www.generationcp.org/sccv10/sccv10_upload/Highlightslittle.pdf).

GCP. 2006. *Generation Challenge Program Annual Report and Year Three Workplan.* Mexico, DF, (available at www.generationcp.org/comm/manual/doc_100.pdf).

Graff, G., Cullen, S., Bradford, K., Zilberman, D. & Bennet, A. 2003. The public-private structure of intellectual property ownership in agricultural biotechnology. *Nature Biotech*. 21: 989–995.

Grattapaglia, D. 2003. Genolyptus. *In*: A. Borém, M. Giúdice, & T. Sediyama, eds. *Melhoramento genômico,* pp. 51–71. Universidade Federal de Viçosa,Viçosa, Brazil.

Laider, D.A. 1998. *Research partnerships between industry & universities. A guide to better practice*. Runcorn, UK, ICI Technology.

Lima, S.M.V., Castro, A.M.G., de Mengo, O. & Medina, M. 2000. *Entorno y Enfoque Estratégico en I & D: Proyecto Nuevo Paradigma.* Artigo submetido ao XXI Simpósio Internacional de Gestão Tecnológica, FIA/ USP/PACTO.

Linstone, H.A. & Turoff, M. 1975. *The Delphi method: techniques and applications.* Readings, MA, USA, Addison Wesley.

Lipnack, J. & Stamps, J. 1993. *The age of the network.* New York, USA, Oliver Wright Publications.

Lopes, M.A. 2000. Cooperation and competition in competitive grants: is "coopetition" being achieved? *In Proc. of the Competitive Grants in the New Millenium: A Global Workshop for Designers and Practitioners,* pp. 219–230. Brasília, Brazil, Embrapa/IDB/The World Bank.

Lörz, H. & Wenzel, G. 2005. Molecular marker systems in plant breeding and crop improvement. *Biotech. in Agric. & Forestry* 55: XXII. Berlin, Germany, Springer.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G. & Yano, M. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breeding* 3: 87–103.

Price, S.C. 1999. Public and private plant breeding. *Nature Biotech*. 17: 938.

Rafalski, J.A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci.* 162: 329–333,

Stuber, C.W., Polacco, M. & Senior, M.L. 1999. Synergy of empirical breeding, marker-assisted selection, and genomics to Increase crop yield potential. *Crop Sci*. 39: 1571–1583.

Young, N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Mol. Breeding* 5: 505–510.

CHAPTER 19

Technical, economic and policy considerations on marker-assisted selection in crops: lessons from the experience at an international agricultural research centre

H. Manilal William, Michael Morris,
Marilyn Warburton and David A. Hoisington

SUMMARY

Molecular markers and related technologies have been used extensively in genetic characterization and identification of loci controlling traits of economic importance in many crop species. However, the application of such tools for crop improvement has not been extensive, at least in the public sector. Although there are clear advantages in using molecular markers as tools for indirect selection of traits of importance, available examples indicate that their use is restricted to traits with monogenic inheritance or when the inheritance is conditioned by a few genes with large effects. Another important limitation of large-scale marker applications is the cost involved in marker assays, which may be beyond the capacities of many public plant breeding enterprises. For an effective marker-assisted selection (MAS) activity to facilitate ongoing crop improvement programmes, especially in the context of the developing countries, laboratories with adequate capacity and adequately trained scientific personnel as well as operational resources are required. Although recent technological advances such as single nucleotide polymorphisms (SNPs) and associated assay protocols are likely to reduce assay costs significantly, for many of these operations, assay platforms with significant capital investments including computational capacity are required. Coupled with these limitations, private sector domination of biotechnology research with proprietary rights to important products and processes with immediate benefits to developing countries may further constrain the benefits these technologies may offer to resource-poor farmers. Policy-makers in different national programmes and international development and research agencies have a responsibility to sustain and augment the capacity of national public agricultural research organizations to ensure that biotechnology tools and processes are infused appropriately into national research efforts. They must also ensure that any biotechnology efforts undertaken are well integrated with national crop improvement activities.

INTRODUCTION

Due to their usefulness in characterizing and manipulating genetic factors responsible for qualitative as well as quantitative traits, molecular markers are considered to be valuable tools for crop improvement. These uses of molecular markers have been invaluable in helping researchers understand complex traits, dissect them into single Mendelian genetic factors, and establish their chromosomal locations via the use of linkage maps and/or cytogenetic stocks. Availability of well characterized genetic linkage maps is a prerequisite for tagging important agronomic or other traits with molecular markers, enabling their use in MAS related activities. To date, however, few practical applications have been published from these studies. This paucity of published studies may indicate the long-term nature of this research, or it might simply reflect the fact that marker technology has been applied to plant breeding efforts mostly by scientists working in the private sector (Hoisington and Melchinger, 2004).

Maize was one of the first crop species for which molecular linkage maps were developed, and Gardiner *et al.* (1993) consolidated several individual maps into a consensus map. Rice is another species for which high-density linkage maps have been developed (reviewed in Gowda *et al.*, 2003) while, due to its high ploidy level and large genome (21 linkage groups, as opposed to 10 in maize and 12 in rice), efforts to develop well characterized, saturated linkage maps with wheat have lagged behind. Other important cereals and legumes are at various stages of linkage map development. The availability of well-defined linkage maps and the extent of genetic studies conducted on them therefore vary among different crops, and this influences the feasibility of any MAS-related activity. Thus, while it is possible to carry out MAS to some degree in cereals such as rice, maize and wheat, and in legumes such as soybean, for species such as cassava and sweet potato, the so-called "orphan crops", genetic improvement with MAS may not yet be feasible. These crop species may benefit more readily from genetic modification arising from direct introduction of genes isolated from other species or organisms, which is not the focus of this chapter.

Citing practical lessons learned at the International Maize and Wheat Improvement Center (CIMMYT) as well as findings of studies conducted elsewhere, this chapter describes some actual and potential applications as well as the advantages and disadvantages of MAS, and outlines possible applications of MAS in developing country plant breeding programmes.

LESSONS LEARNED FROM CROPS

Numerous scientific reports describe molecular mapping and analysis of quantitative trait loci (QTL) for nearly every agronomic trait in a diverse array of crop species. The traits covered include many parameters associated with tolerance to drought and other abiotic stresses, maturity, plant height, quality parameters, qualitative and quantitative factors of disease and pest resistance, and numerous seed traits and yield. Although these efforts have resulted in a vast amount of knowledge and better understanding of the underlying genetic factors that control these traits, application of this knowledge to manipulate genes in an effective or simple manner for improving crop species has had limited success. The scientific community is faced with the challenges of accurate and precise QTL identification and application of the information derived to successful MAS efforts.

Scientific advances have been instrumental in increasing the power and accuracy of computational parameters as well as designing ways of combining the information generated from molecular genetics with traditional crop improvement efforts. Numerous simulation studies have been undertaken to evaluate the effectiveness of MAS, taking into account the influence of heritability, population size, linkage distance, etc. (Xie and Xu, 1998; Moreau *et al*., 1998; Ribaut, Jiang and Hoisington, 2002), and MAS procedures have been used to incorporate traits of interest from exotic species including wild relatives into elite cultivars through advanced backcross QTL analysis (Tanksley and Nelson, 1996; Fulton *et al*., 2000).

Manipulation of qualitative traits

Molecular markers that are tightly linked to genes having a strong effect on the expression of a trait can be used to introgress the genes (and thus the trait) into different backgrounds through backcross breeding schemes that rapidly and efficiently improve the recurrent parent for the target trait. In conventional backcross breeding schemes and line conversion activities, the donor parent containing the trait of interest is crossed with the recurrent parent, normally a well-adapted variety lacking the trait of interest. The resulting progeny are screened to identify the trait of interest, and individuals exhibiting the trait are crossed to the recurrent parent. The entire process is repeated several times. For traits that are conditioned by recessive gene action, a cycle of selfing is also required after each crossing cycle. After several cycles of backcrossing and a final self-pollination, plant breeders are often able to recover lines that are nearly identical to the recipient parent but also contain the trait of interest. Compared with traditional backcrossing, the use of DNA markers enables faster recovery of the recurrent parent genotype along with the introgressed target trait in line conversion activities. Ribaut and Hoisington (1998) reported that MAS should enable the recovery of the target genotype after three cycles of backcrossing, compared with a minimum of six cycles with traditional approaches (Tanksley *et al*., 1989).

CIMMYT has a long history of using molecular markers for certain traits in maize improvement. Although maize is widely used for both food and feed, maize kernels do not provide sufficient quantities of two essential amino acids, lysine and tryptophan. The *opaque2* mutation, identified at Purdue University (United States of America) in the mid-1950s, confers elevated levels of these two amino acids. Although initial efforts to introduce the *opaque2* mutation into breeding materials were not successful (Villegas, 1994), researchers eventually succeeded in producing nutritionally enhanced maize lines. These came to be known as quality protein maize (QPM). CIMMYT breeders have used traditional backcrossing to transfer the *opaque2* mutation and associated modifiers into elite lines. To perform phenotypic selection in segregating progenies for lines carrying the *opaque2* mutation, it is necessary either to wait until the plants produce mature ears, or to do random pollination on a large number of plants. Although reliable laboratory screening techniques are available, co-dominant microsatellite markers present within the *opaque2* mutation can be used earlier in the growing season. Using these markers in backcross progenies, plants heterozygous for the *opaque2* mutation can be selectively identified as a qualitative trait for use in the next crossing cycle. Markers

are not used to select for the background recurrent parent genotypes, but only to select lines carrying the *opaque2* mutation allele. Although CIMMYT uses markers for detecting the presence of the *opaque2* mutation, markers are not available to select for the modifiers, which are important in determining seed texture and quality and for which other traditional screening techniques are being used.

A well known example of marker-assisted backcrossing of a qualitative trait involves the introgression of the *Bt* transgene into different maize lines (Ragot *et al.,* 1994). Whenever plant transformation techniques are used to produce genetically modified organisms (GMOs), usually there are some cultivars that are more receptive to transformation procedures than others. When the cultivar with the best agronomic type is not the most receptive to transformation, it is often possible to transform another cultivar that is receptive and then use the diagnostic marker that detects the transgene to introgress it into more desirable backgrounds. This type of MAS-aided line conversion can be accomplished for any crop species. The presence of markers to detect the transgene enables the detection of converted progeny with a high degree of accuracy.

Another MAS-related CIMMYT experience involves the case of maize streak virus (MSV) resistance, for which a major QTL was identified on maize chromosome 1 that explains 50–70 percent of total phenotypic variation (Pernet *et al*., 1999a, b). As maize has a well-saturated molecular linkage map, several microsatellite markers associated with this QTL were identified in the specific chromosomal region. These markers were tested in three populations generated using three different MSV tolerant lines crossed with one susceptible line. After screening the F_2 progeny and F_3 families, lines identified by markers were sent to Africa, where MSV is prevalent. By phenotypic screening of the lines selected by MAS, it was established that MAS-selected lines were significantly more resistant to MSV (J-M. Ribaut, personal communication).

In legumes, resistance to soybean cyst nematode (SCN) is one example of an effective MAS approach. Routinely used phenotypic assays for SCN screening take approximately five weeks and extensive greenhouse space and labour. Successful identification of closely linked microsatellite markers has enabled transfer of the resistance gene *rhg1* with about 99 percent accuracy (Cregan *et al.,* 1999; Young 1999). Many public and commercial soybean cultivar improvement efforts use these markers to screen for SCN resistance (Young, 1999). Another example of successful MAS in common beans was reported by Yu, Park and Poysa (2000) who used markers associated with common bacterial blight. These markers identified a locus that explained about 62 percent of the phenotypic variation and have been used in MAS experiments.

As described earlier, linkage map construction in wheat is more challenging than in species such as rice or maize. The allohexaploid nature allows wheat to withstand chromosomal imbalances as the loss of one chromosome can be compensated by the presence of a homologous chromosome. As a result, wheat can be crossed with a range of wild relatives (both intergeneric and interspecific), enabling introgression of genetic material possessing resistances to different biotic and abiotic stresses. When translocations (especially intergeneric translocations) are present in wheat, markers can be readily developed

for the translocated chromosome segments. If a translocated segment carries a trait of importance, markers can then be used to transfer it into different wheats. Diagnostic or perfect markers (i.e. markers with complete linkage to the genes of interest with no possibility of recombination) have been developed for genes conferring resistance to different biotic stresses in wheat. CIMMYT's wheat improvement efforts use a set of markers routinely on a seasonal basis for introgression of a set of genes into high-yielding backgrounds. Examples of the perfect markers that are currently in use are:

- Cereal cyst nematode (CCN) resistance gene *Cre*1 (2BL), identified in wheat landrace AUS10894 and *Cre*3 (2DL), derived from *Triticum tauschii* (Lagudah, Moullet and Appels, 1997). These markers are used routinely in segregating populations to enable selective advancement of lines containing the *Cre* genes targeted to all environments, but particularly to marginal ones, where healthy root architecture is essential to allow plants to take advantage of minimal soil moisture. Phenotypic evaluation for CCN resistance is labour intensive as well as expensive. Given that it is impossible to screen for CCN resistance in Mexico (where CIMMYT headquarters are located) due to the lack of required screening facilities, the use of markers is essential for improving this trait.
- Barley yellow dwarf virus (BYDV) resistance, derived from a chromosome segment introgressed from *Thinopyrum intermedium*, on chromosome 7DL (Ayala *et al.*, 2001). BYDV is an important viral disease in certain wheat growing regions of the world. Environmental influence makes field screening less reliable. The diagnostic marker for the translocated chromosome segment allows the alien-derived resistance to be combined with the BYDV tolerance available in wheat.
- Marker for *Aegilops ventricosa*-derived resistance to stripe rust (*Yr17*), leaf rust (*Lr37*) and stem rust (*Sr38*) (O. Robert, personal communication). The translocation from *Ae. ventricosa* is present on chromosome 2AS. The diagnostic marker for the translocation is used mainly in bread wheat x durum wheat crosses, to identify the durum derivatives carrying the translocation.

In addition, CIMMYT uses a set of linked markers for transferring a locus with major effects for boron tolerance (*Bo-1*), crown rot resistance, scab resistance and stem rust resistance in its MAS efforts. These efforts with linked genes are conducted with the objective of increasing the allele frequency for desirable alleles in segregating populations (William, Trethowan and Crosby-Galvan, 2007).

Gene pyramiding/stacking

MAS lends itself well to gene pyramiding efforts for disease resistance. When a cultivar is protected by one gene with major effects against a specific disease, it is often not possible to introgress additional genes conferring resistance to the same disease because of the difficulty of phenotypic screening for the presence of additional genes (as the plant already shows resistance to the disease). However, if several genes can be tagged with closely linked molecular markers, MAS strategies can be used to develop lines with stacked genes, giving the cultivar more durable protection than that afforded by a single resistance gene.

Resistance to bacterial blight provides an excellent example of using MAS for gene pyramiding. Bacterial blight is caused by

Xanthomonas oryzae and is one of the most important diseases of rice. Several genes that confer resistance to bacterial blight have been tagged with molecular markers. Huang *et al.* (1997) and Hittalmani *et al.* (2000) developed strategies for combining four resistance genes, namely *Xa-4, Xa-5, Xa-13* and *Xa-21,* in a single cultivar using pairwise combinations of the genes. Due to the co-dominant nature of the markers used, the authors were able to select from F_2 generations without having to perform progeny testing. The derived lines containing pyramided genes showed higher level of resistance and/or a wide spectrum of resistance compared with the parental material. Another gene pyramiding example using MAS involves stacking of the resistance genes *rym4, rym5, rym9* and *rym11* for the barley yellow mosaic virus complex using molecular markers and doubled haploids (Werner, Friedt and Ordon, 2005). Other examples include pyramiding for barley stripe rust resistance (Castro *et al.*, 2003), and powdery mildew resistance in wheat (Liu *et al.,* 2000) and, in MAS applications at CIMMYT, crosses have been made to combine two genes for cereal cyst nematode resistance and three different genes for stem rust resistance (*Sr24, Sr26* and *Sr25*) in targeted wheat germplasm.

Manipulation of quantitative traits

Quantitatively inherited traits are genetically complex, are conditioned by a number of genes each having relatively small effects, and their expression often depends on interactions among different genetic components (epistasis). The environment also has a high degree of influence on the expression of the trait, which confounds the interpretation of QTL identification and often renders the results obtained from QTL studies cross-specific. When it is necessary to manipulate several genomic regions simultaneously, each having different effects on the same trait of interest, MAS-based approaches become more complicated and present formidable challenges. Mapping studies conducted at CIMMYT identified five genomic regions associated with the anthesis-silking interval which is a parameter associated with drought tolerance in maize (Ribaut *et al.*, 1996, 1997). The drought tolerant parent was used in MAS experiments as the donor parent to transfer the five QTL to CML 247, an elite inbred line with good combining ability that was drought-susceptible but high-yielding under favourable conditions. Markers were used to generate 70 BC_2F_3 lines containing the favourable alleles from the drought-resistant parent after two backcrosses and two self pollinations. These lines were crossed with two testers for field evaluation. Field tests indicated that under severe drought stress conditions, the 70 MAS-derived lines were significantly better yielding than the controls. The differences were less prominent under reduced drought stress (Ribaut and Ragot, 2007).

Other CIMMYT experiments aimed at comparing MAS with phenotypic selection have been conducted for stem borers in tropical maize (Willcox *et al.,* 2002). In the case of maize stem borer resistance, three QTL identified through mapping experiments were used in MAS. Three BC_2S_2 families that carried all three target QTL from the donor parent in homozygous state were developed. Comparative studies with MAS and traditional phenotypic selection did not establish a clear advantage for MAS, but both approaches yielded significant genetic gains in reducing leaf damage. MAS is not being used currently on a routine basis at CIMMYT for drought and stem borer resistance.

Other reports describing the manipulation of quantitatively inherited traits include those of Bouchez *et al.* (2002) for introgressing favourable alleles at three QTL for earliness and grain yield in maize, and by Yousef and Juvik (2001) who reported on MAS for seedling emergence and eating quality characters in sweet corn. Also, Han *et al.* (1997) attempted to select for barley malting traits using MAS. Additional scientific reports are available that describe MAS-related efforts for quantitatively inherited traits.

In general, manipulating several QTL associated with multiple genomic regions in segregating progenies is considerably more challenging. Often the success in genetic gains depends on the stability of these QTL as well as the cost efficiency of large-scale MAS applications.

Genetic diversity studies

In addition to being used in MAS activities, molecular markers have been used extensively for genetic diversity studies. Numerous scientific publications are available that describe the use of molecular markers in estimating the degree of relatedness of a set of cultivars in many cultivated crop species. In common with their use in trait manipulations, the practical outcomes of the numerous genetic diversity studies using molecular markers are not clear. Evaluation of genetic relatedness using molecular markers will have implications on understanding the genetic structure of existing populations, enabling the design of strategies for proper acquisition of germplasm for conservation purposes. The genetic uniqueness of accessions or populations in germplasm collections can be accurately estimated by the use of DNA profiling (Brown and Kresovich, 1996; Smith and Helentjaris, 1996). Molecular markers have also been used for identifying redundancies in existing germplasm collections in rice (Xu, Beachell and McCouch, 2004) and sorghum (Dean *et al.,* 1999). In cassava, Chavarriaga-Aquirre *et al.* (1999) used morphological traits, isozyme profiles and agronomic criteria to identify a core set of 630 accessions from a base collection of approximately 5 500 accessions.

Modern farming in advanced countries is based on high performing, genetically uniform new cultivars, which are generally derived from well adapted, genetically related parental material. Tanksley and McCouch (1997) have concluded that most modern soybean cultivars grown in the United States can be traced back to a very limited number of strains from a small area of northeastern China, while a majority of hard red winter wheats is derived from a few lines originated in Poland and the Russian Federation. The genetic basis of modern rice varieties grown in the United States is also considered narrow (Dilday, 1990).

Another application in the area of genetic diversity is the use of markers in identifying heterotic groups. Molecular markers have been used extensively in the construction of heterotic groups since the 1990s in many different crop species of economic importance. Heterotic groups are clusters of germplasm usually with similar characteristics and a high degree of relatedness that, when crossed with materials from another heterotic group, tend to give rise to progeny with high levels of heterosis. Although markers randomly distributed in the genome can be used to develop heterotic groups, their usefulness in determining hybrid performance is not clear. While it is reasonable to assume that heterosis depends on the interactions among favourable alleles belonging to the two parents, unless molecular markers that are known to be linked to

these favourable alleles are used in heterotic studies, the predictive power of markers in estimating heterosis for practical applications may not be very high.

At CIMMYT, large-scale, rapid characterization methods for inbred lines and populations have been optimized using up to 120 microsatellite markers spread throughout the maize genome. In the past, characterizing maize populations was costly and time-consuming, given that as many as 22 individuals had to be analysed individually to calculate allele frequencies for each marker. Currently, a bulking method in which 15 individuals from a population are amplified in the same polymerase chain reaction (PCR) and run on an automatic DNA sequencer, provides a reliable estimate of the allele frequencies within that particular population. Between one and two bulks can now be used to fingerprint populations with considerable savings in time and resources. Other studies of maize genetic diversity have been conducted for CIMMYT maize breeders as well as outside collaborators with objectives that include: determining how maize inbred lines from different national breeding programmes are related to each other (and to determine the possibility of sharing among regions or using lines from one region to expand diversity in another); establishing heterotic groups; determining levels of genetic diversity present in synthetic varieties; determining how landraces and farmers' varieties from different regions are related to each other; monitoring homozygosity levels in inbred lines; and tracking changes in lines that have been intensively selected for a given trait.

A core set of 100 microsatellite markers has been selected for wheat genetic diversity studies. Recent fingerprinting studies by CIMMYT and national programme scientists have been conducted to assist in regenerating gene bank accessions without losing genetic diversity, measuring the contribution of wild ancestors and exotic species in advanced backcross progenies of synthetic bread wheat, and to track the changes over time in diversity levels of CIMMYT wheat cultivars from the original Green Revolution varieties to modern breeding lines.

Marker implementation

To facilitate the use of MAS activities in wheat and maize improvement efforts, CIMMYT has recently established a marker implementation laboratory. This provides the facilities and technical expertise to provide CIMMYT wheat and maize breeders with access to biotechnology tools, including MAS. The laboratory carries out two main MAS-related activities, marker adoption and research support. The first includes constantly reviewing the literature to identify markers developed by third parties and verifying that these can be used to detect traits or genes of interest in CIMMYT germplasm improvement efforts, and developing efficient protocols for their in-house use. The second consists of a range of routine tasks that include growth and/or sampling of plant tissue, DNA extraction, marker detection, data analysis and dissemination of results to breeders.

Close cooperation between field and laboratory staff is important to be able to apply molecular markers in crop improvement efforts. Ideally, laboratory staff should have an understanding of field activities and field workers should have basic knowledge of different aspects of MAS-associated laboratory procedures. MAS is used when there is a high probability that markers will help plant breeders achieve genetic gains faster and more economically than field

or laboratory-based phenotypic selection methods. When perfect markers are available to screen for a particular trait, such markers are preferred. However, for traits that cannot be screened conveniently using traditional approaches and even when perfect markers are not available, if markers are available with close linkages to the trait(s) of interest, these can be used to increase the desirable allele frequency for the target gene. MAS-related activities in both wheat and maize at CIMMYT are conducted as collaborative projects involving both breeders and biotechnologists. The breeders use information coming from wheat MAS activities to define better their parental crossing block materials and to make selective crosses using parents identified by markers. Moreover, segregating early generation progenies in certain crosses are selected in the field based on whole plant phenotype, which are then further refined by sampling leaf tissue from field-tagged plants and processing for MAS assays in the laboratory. Only those entries that contain the target genes identified with MAS are advanced to the next generation. This enables breeders to reduce population sizes for the traits under evaluation and accumulate certain gene combinations in elite backgrounds. The material thus generated is advanced through several cycles of selfing and eventually used in field screening to identify the best performing lines.

ECONOMICS OF MAS

Establishing the capacity to conduct MAS

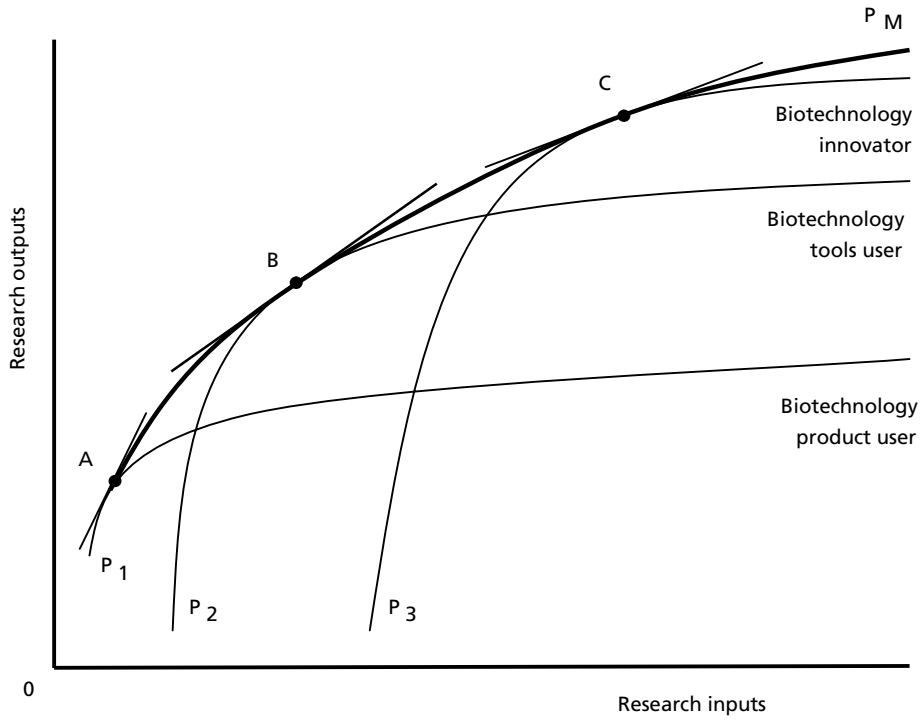
For MAS to be a viable option for a plant breeding programme, adequately equipped laboratory facilities must be in place and appropriately trained scientists must be available. Therefore, one of the first decisions facing research managers considering MAS is whether to invest in biotechnology research capacity.

Economic theory suggests that the most efficient level of research investment can be determined with the help of a research production function that relates research inputs to research outputs. At the national level, the research production function can be thought of as a meta-function encompassing the frontiers of many smaller functions, each representing a different level of research capacity distinguished by complexity and scope (Figure 1) (Brennan 1989; Byerlee and Traxler, 2001; Maredia, Byerlee and Maredia, 1999; Morris *et al.*, 2001). Movement outwards along the meta-function, accomplished by adding subprogrammes and thereby increasing the number of researchers and the extent of available research infrastructure, is associated with changes in focus and increases in the capacity of the national research programme.

For a plant breeding programme, adding new biotechnology-based subprogrammes is equivalent to taking a series of discrete steps involving increased complexity and cost. These steps have the effect of moving the programme from one level of research capacity to the next. These levels of research capacity can be broadly characterized as follows:

- *Biotechnology product user.* Here, the research programme imports germplasm products developed using biotechnology and incorporates them into its conventional crop improvement schemes, either by backcrossing them into local germplasm or by testing them for potential immediate release.
- *Biotechnology tools user* where the research programme imports biotechnology tools and uses them, if necessary, after adapting them to local

FIGURE 1
Biotechnology research production meta-function



Source: Morris *et al.* (2001).

circumstances, to improve current crop improvement practices.

- *Biotechnology methods innovator,* in which the research programme establishes the full capacity needed to develop innovative biotechnology tools and products.

Moving from one level of biotechnology research capacity to the next usually requires significant investments in laboratory facilities and staff training. The practical decision facing research managers is not to determine the optimal level of research investment, but rather to select from among the different levels of biotechnology research capacity characterized by increasing complexity and cost (A or B or C in Figure 1). The choice should be based on whether a given level of research capacity can be expected to generate enough additional benefits to justify the additional expenditure. For most plant breeding programmes, benefits consist of value added to crop production enterprises. Therefore, the incentive to invest in additional research capacity will tend to increase with the size of the area planted and/or the value of the crops expected to benefit from the research.

There are few published estimates of the cost of moving from one level of biotechnology research capacity to the next, and new estimates are not provided here. Empirical estimates would quickly be outdated, as the cost of biotechnology laboratory equipment and materials continues to change very rapidly. However,

for the purposes of this chapter it is important to point out that although establishing capacity to develop new molecular markers requires substantial investment, establishing the capacity to use freely available existing molecular markers requires only a modest investment.

Variable cost of MAS

At CIMMYT the capacity to carry out MAS on a reasonable scale has been developed, but the need now is to make the technology work on a high-throughput scale to reduce the cost per data point, while being able to handle large quantities of assays per growing season. In this regard, there are several challenges to consider as markers are not always cost-effective even when they improve the precision of selection. Depending on the nature of the target trait (quantitative or qualitative), the type of gene (major or minor), the form of gene action that controls expression of the trait (dominant or recessive effect), and the ease with which the trait can be measured (visually detected or more expensive field or laboratory analysis required), conventional selection may be cheaper than MAS. The desirability of using genetic markers therefore depends in part on the costs of genotypic versus phenotypic screening, which vary among applications.

Information about the cost of using MAS at CIMMYT for specific breeding projects is available from case studies. For example, Dreher *et al.* (2002, 2003) examined the costs and benefits of using MAS for a common application in maize breeding. This study generated three noteworthy conclusions.

First, for any given breeding project, detailed budget analysis is needed to determine the cost-effectiveness of MAS relative to conventional selection methods. Although the costs of field operations and laboratory procedures required for molecular marker analysis may remain relatively constant across applications, every breeding project is likely to involve unique phenotypic evaluation procedures whose costs will frequently differ.

Second, direct comparisons of unit costs for phenotypic and genotypic analysis provide useful information to research managers, but in many cases technology decisions are not made solely on the basis of cost. Factors other than cost often influence the choice of screening methods. Time considerations are often critical, as genotypic and phenotypic screening methods may differ in their time requirements. Even when labour requirements are similar, for applications in which phenotypic screening requires samples of mature grain, genotypic screening can often be completed much earlier in the plant growth cycle.

Third, conventional and MAS methods are not always direct substitutes. Using molecular markers, breeders may be able to obtain more information about what is going on at the genotypic level than they can obtain using phenotypic screening methods. For example, in conventional backcross breeding or line conversion projects (see section *Manipulation of qualitative traits*), background molecular markers can be used to identify those plants among a set of progeny that not only possess a desirable allele but also closely resemble the recurrent parent at the genetic level. Based on this additional information, breeders are often able to modify their entire breeding strategy, with potentially significant implications in terms of cost and/or time requirements (this issue is discussed in the next section).

The CIMMYT case study thus confirmed what many practising plant breeders

intuitively know: namely, the costs and benefits of MAS projects are likely to vary depending on the crop being improved, the breeding objective being pursued, the skill of the breeder, the capacity of the research organization, the location of the work being carried out, the cost of key inputs, and many other factors.

Economic trade-offs

While caution is required when extrapolating from the results of a case study, general conclusions regarding the cost-effectiveness of molecular markers in crop genetic improvement work can be drawn based on the findings of the CIMMYT study and a number of other studies carried out elsewhere. Broadly speaking, two types of benefits associated with MAS can be distinguished: cost savings and time savings.

Cost savings

For certain applications, MAS methods can substitute directly for conventional selection methods, and for these applications the relative cost-effectiveness of the two methods can easily be determined by comparing the screening cost per sample. Generally, as the cost of phenotypic screening rises, markers are more likely to represent a cost-effective alternative. For applications in which phenotypic screening is easy and cheap (e.g. visual scoring of plant colour), MAS will not offer any obvious advantages in terms of cost. However, for applications in which phenotypic screening is difficult or expensive (e.g. assessing root damage caused by nematodes or for a disease that is not present in the field site), MAS will often be preferable.

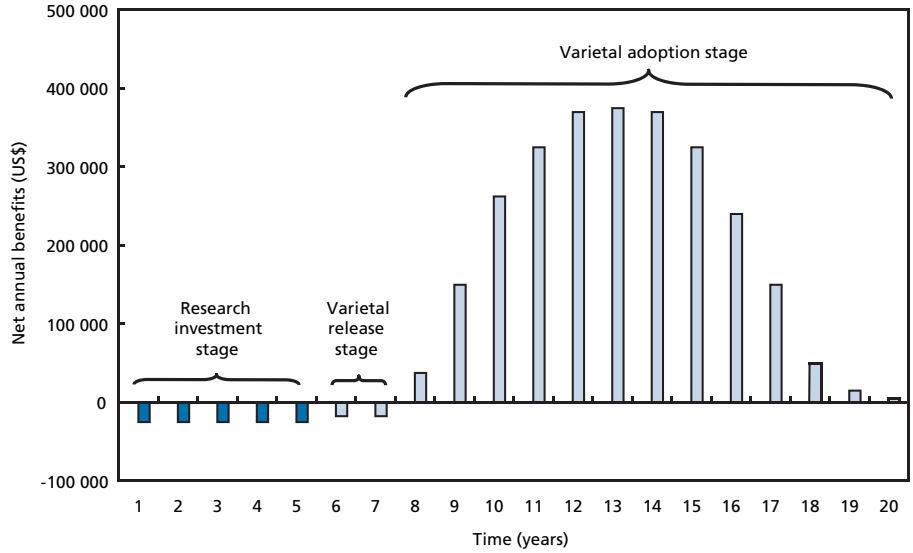
Time savings

Cost is an important factor affecting the choice of breeding technology, but it is not the only one. Plant breeders worry about controlling costs, but they also worry about getting products out quickly. Therefore, it is not sufficient to consider potential cost savings alone. The time requirements of alternative breeding strategies must also be taken into account, because even when MAS costs more than conventional selection (as it does in some, although not all, cases), breeders who use it may be able to generate a desired output quicker. Accelerated release of improved varieties can translate into large benefits, especially for the private seed industry, so time is an important consideration in addition to cost.

For breeding applications in which MAS offers cost and time savings, the advantages of MAS compared with conventional breeding are clear. More problematic, however, are the many applications in which MAS methods cost more to implement than conventional selection methods but also reduce the time needed to accomplish a breeding objective. This commonly happens, for example, with inbred line conversion schemes based on backcrossing procedures. In such schemes, MAS methods can often be used to derive converted inbred lines containing one or more incorporated genes in much less time than would be possible using conventional selection methods alone.

In applications that involve a trade-off between time and money, under what circumstances is the higher cost of MAS relative to conventional breeding justified? The choice of the plant breeding method can be viewed as an investment decision and evaluated using conventional investment criteria (Sanders and Lynam, 1982). Using data from the CIMMYT case study, Morris *et al.* (2003) explored the relationship between time and money as it relates to crop improvement research

FIGURE 2
Stylized economic model of a plant breeding programme



by developing a simple model of a plant breeding programme and using it to compare the returns with alternative inbred line conversion schemes based on conventional selection and MAS. Two measures of project worth were used: the net present value (NPV) of the discounted streams of costs and benefits, and the internal rate of return (IRR) to the investment.

Figure 2 depicts the stylized "variety life cycle" assumed by the model. The stream of costs and benefits associated with the development, release and adoption by farmers of an improved variety can be divided into three stages: a research stage during which the variety is developed; a release stage during which the variety is evaluated and registered for release, and commercial seed is produced; and an adoption stage during which the variety is taken up and grown by farmers. During the first two stages, net benefits are negative, because costs are incurred without any benefits being realized. During the third stage, net benefits turn positive as the variety is taken up and grown by farmers; they continue to increase until the peak adoption level is achieved and then decline when the variety is replaced by newer varieties.

The model was used to estimate the NPV and IRR of conventional and marker-assisted inbred line conversion schemes. Research cost data were taken from the CIMMYT case study. Plausible values were used for key parameters relating to the varietal release and adoption stages (for details, see Morris *et al.*, 2003). Figure 3 shows the streams of annual net benefits generated by each of the two breeding schemes. Annual net benefits are calculated as follows:

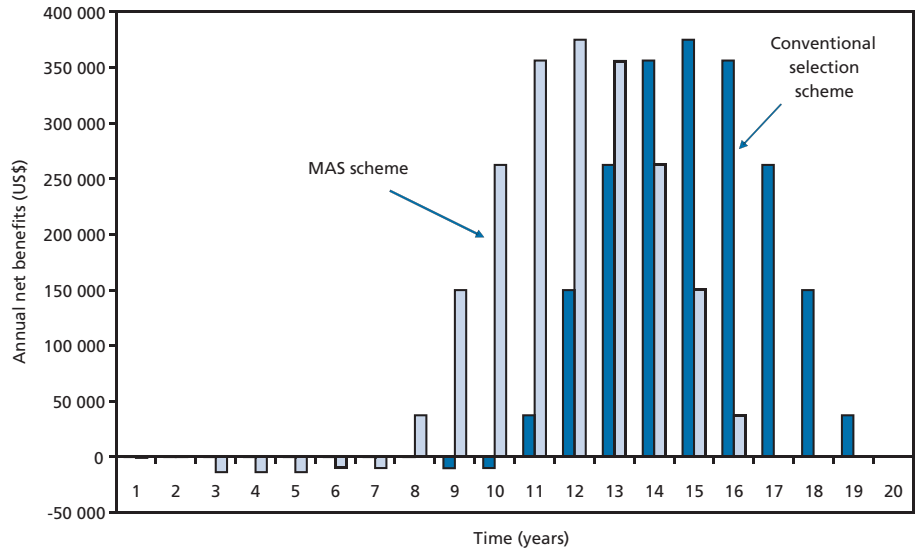
$$NB_t = (GB_t - VR_t - RC_t)$$

where:

NB = net benefits

GB = gross benefits (calculated as area planted to the variety x incremental benefits associated with adoption)

FIGURE 3
Annual net benefits, conventional vs. marker-assisted line conversion scheme



VR = varietal release expenses (cost of evaluation trials, registration procedures, seed multiplication, advertising and promotion, etc.)
RC = research investment costs
t = year (1…n)

NPVs were calculated by adding the discounted stream of net benefits associated with each breeding scheme over the life of the variety (n years):
where:

$$NPV = \sum_{t=1}^{n} (GB_t - VR_t - RC_t) / (1 + r)^t$$

NPV = net present value
r = discount rate

IRRs were calculated conventionally by solving the discount rate that drives the NPV to 0.

The profitability rankings of the two breeding schemes, MAS and conventional, were found to differ depending on the measure of project worth that was used. The MAS scheme generated the highest NPV, whereas the conventional breeding scheme generated the highest IRR on investment. These results, generated using a stylized model of a plant breeding programme and plausible values for varietal release and adoption parameters, provide an important insight into the relative cost-effectiveness of conventional selection methods and MAS in applications involving trade-offs between time and money. From an economic perspective, the relative attractiveness of conventional versus MAS methods will depend on the availability of research investment capital. If investment capital is abundant (meaning that the breeding programme can afford to absorb the higher up-front costs associated with MAS without curtailing other ongoing breeding projects), MAS may become a desirable option, because it generates the largest NPV. On the other hand, if investment capital is constrained (i.e. the breeding

programme cannot absorb the higher up-front costs associated with MAS, or that it can absorb them only by forgoing other potentially profitable breeding projects), it makes sense to choose conventional selection, because it generates the largest IRR.

IMPLICATIONS FOR DEVELOPING COUNTRIES

When discussing policy implications of MAS efforts in developing country scenarios, it is appropriate to consider the experience gained over the past several decades, mainly in industrialized countries. In advanced countries, the private sector has made significant investments in MAS efforts while there are a few publicly-funded research groups using MAS in breeding routinely and these are restricted to a few target crops (Eagles *et al.*, 2001; Dubcovsky, 2004; William, Trethowan and Crosby-Galvan, 2007). Information about the traits and the breeding strategies used in MAS applications in large agribusiness enterprises are not publicly available freely. To date, significant investments have been made in biotechnology applications only for widely grown crop species such as rice, maize, wheat, soybean, cotton and canola. While GM crops and their implications are not the focus of this chapter, it is reasonable to assume that technologies associated with GM crops offer significant potential for addressing biotic and abiotic stress tolerance in widely grown cereals and legumes as well as species that are important but thus far neglected such as tef, millets, yams and other tuber crops in the developing countries. For example, GM technologies that can make one crop species perform better are likely to be valuable with slight modifications to enhance the performance of a neglected crop species. When useful GM varieties of a particular crop are made available, they also become prime candidates to apply MAS-based introgression of the introduced gene construct/s to other well adapted cultivars in different agro-ecological regions.

Reports indicate that two rice varieties with improved bacterial blight resistance have been developed with MAS approaches and deployed in Indonesia (Toenniessen, O'Toole and DeVries, 2003). Moreover, rice varieties carrying multiple disease resistance genes are being developed by several national programmes with technical backstopping by the International Rice Research Institute (IRRI) (Hittalmani *et al.*, 2000). There are also reports describing the use of MAS in China for improving certain quality traits in rice (Zhou, P.H. *et al.*, 2003) and wheat (Zhou, W-C. *et al.*, 2003) and fibre related traits in cotton (Zhang *et al.*, 2003), but it is not clear whether these are one-time research efforts or there is continued activity using MAS.

Although it is not possible to obtain entirely reliable estimates of the costs, benefits and cost-effectiveness of MAS applications, the costs associated with MAS are frequently considered as the main constraint to their effective use by many plant breeders, especially in small- to medium-scale breeding enterprises. However, new marker technologies, especially those based on single nucleotide polymorphisms (SNPs) and associated ongoing large-scale genome sequencing projects, should enable the development and deployment of gene-based markers in the near future (Rafalski, 2002). SNPs are defined as single base differences within a defined segment of DNA at corresponding positions. These SNP-based polymorphisms are known to be abundantly present in human as well as in plant genomes. Consequently, the potential exists to develop SNP markers associated with many important traits in a diverse array of

economically important crop species. For species such as maize, rice and soybeans, robust SNP-based assay platforms already exist in the private sector as well as in some public sector enterprises. The added advantage of SNP-based marker systems is that they avoid gel-based allele separations for visualization and have the potential for automation in high-throughput assay platforms. These ongoing research efforts will inevitably lead to the development of more robust, high-throughput assays that are both simple and cost effective (Jenkins and Gibson, 2002).

When is it advantageous to use MAS?

In addition to the cost and time savings described above, for a number of breeding scenarios, MAS methods are likely to offer significant advantages compared with conventional selection methods. These scenarios assume the availability of markers for multiple traits and take into consideration the advantages of MAS under optimum situations (Dreher *et al.,* 2002; Dudley, 1993).

- *Gene stacking for a single trait.* MAS offers potential savings compared with conventional selection when it allows breeders to identify the presence of multiple genes/alleles related to a single trait, and the alleles do not exert individually detectable effects on the expression of the trait. For example, when one gene confers resistance to a specific disease or pest, breeders would be unable to use traditional phenotypic screening to add another gene to the same cultivar in order to increase the durability of resistance. In such cases, MAS would be the only feasible option, provided markers are available for such genes.
- *Early detection.* MAS offers potential savings compared with conventional selection when it allows alleles for desirable traits to be detected early, well before the trait is expressed and can be detected phenotypically. This benefit can be particularly important in species that grow slowly, for example, tree crops.
- *Recessive genes.* MAS offers potential savings compared with conventional selection when it allows breeders to identify heterozygous plants that carry a recessive allele of interest whose presence cannot be detected phenotypically. In traditional breeding approaches, an extra step of selfing is required to detect phenotypes associated with recessive genes.
- *Heritability of traits.* Up to a point, gains from MAS increase with decreasing heritability. However, due to the difficulties encountered in QTL detection, the gains are likely to decline beyond a certain threshold heritability estimate.
- *Seasonal considerations.* MAS offers potential savings compared with conventional selection when it is necessary to screen for traits whose expression depends on seasonal parameters. Using molecular markers, at any time of the year, breeders can screen for the presence of an allele (or alleles) associated with traits that are expressed only during certain growing seasons. For example, CIMMYT's wheat breeding station in northern Mexico is usually used for screening segregating germplasm for leaf rust resistance. However, expression of leaf rust is not uniform in all growing seasons. The same concept is true for field screening for drought tolerance. When there are seasons with low expression of leaf rust or less intense drought due to unexpected rainfall, markers, if available, can be a valuable alternative as a tool for screening.
- *Geographical considerations.* MAS offers potential savings when it is necessary

to screen for traits whose expression depends on geographical considerations. Using molecular markers, breeders in one location can screen for the presence of an allele (or alleles) associated with traits expressed only in other locations.

- *Multiple genes, multiple traits.* MAS offers potential savings when there is a need to select for multiple traits simultaneously. With conventional methods, it is often necessary to conduct separate trials to screen for individual traits.
- *Biological security considerations.* MAS offers potential advantages over selection based on the use of potentially harmful biological agents (e.g. artificial viral infections or artificial infestations with insect pests), which may require specific security measures.

In view of the above-mentioned factors, it is desirable to consider MAS approaches on a case-by-case basis, taking into account factors such as the importance of a trait in the overall breeding scheme, the amount of available resources in terms of both staff and operational expenditures, and the nature of the breeding materials. There are no "one size fits all" recommendations that can be made for MAS approaches. Usually, no breeding scheme focuses on improving just one trait. At current levels of capacity, MAS is likely to be used to achieve genetic gains for single traits such as host plant resistance to pests and/or diseases. Therefore, MAS activities should be integrated into an overall breeding programme.

Challenges for developing countries

The rapid expansion of agricultural biotechnology is generating a wide array of methodologies with potential applications, and therefore national programmes in developing countries face the difficult challenge of identifying priority areas for investment. To complicate matters further, the private sector dominates many fields of biotechnology research and therefore has proprietary rights to many technologies and products that have immediate applications in developing countries (e.g. transgenic technology). This is quite different from conventional plant breeding technologies, most of which were developed by publicly-funded research programmes and thus have remained more accessible.

There is no single answer to meeting these challenges, especially as developing countries are not uniform in their public agricultural research capacities. Broadly speaking, developing countries fall into the following categories:

- countries (a few) with strong public sector research infrastructure enabling biotechnology applications, as well as upstream research capability to develop tools for their own specific needs;
- countries with intermediate capacity in applied plant breeding, as well as in using biotechnology tools that are publicly available or can be acquired through bilateral partnerships with the private sector;
- countries (a considerable number) with moderate plant breeding capacity and practically no, or very little, capacity for biotechnology applications.

More advanced developing countries with major commercial farming sectors are more likely to succeed in adopting agricultural biotechnology. In addition, the presence of commercial opportunities will attract investment by private industry and thus allow the country to benefit from future advances in biotechnology. This is not always a positive outcome for the public sector because, as competition increases, it may be more difficult to justify large public investments in biotechnology. This

has occurred to some degree in maize biotechnology, even in the United States.

Developing countries, in which agriculture is still dominated by subsistence farming and where there is limited or no capacity for biotechnology research, are at an added disadvantage. Resource-poor farmers in such countries rarely offer adequate market incentives for the private industry that dominates biotechnology research. For example, the involvement of the private sector in research and development activities for root crops or grain legumes is doubtful as these crops are grown mainly by small-scale farmers in poorer regions of the world and there would be potentially low returns on investment. Therefore, it is important that international development agencies ensure that neither the "orphan commodities" yielding broad socio-economic benefits, nor the less advantaged and least developed countries, are left out from the prospect of harnessing potential benefits associated with biotechnology. In doing so, they must evaluate what biotechnology tools can be of immediate benefit to such crops and countries and then develop strategies leading to successful adoption by the target groups. This can only be accomplished if the efforts made are serious, long-term and sustainable. Many examples can be cited where international aid agencies have invested in purchasing equipment designed for biotechnology research in developing countries but, when the aid programmes terminate their short-term involvement, the capital investments either have not been optimally utilized or have remained idle.

Policy-makers in different national programmes must also bear in mind that sustained capacity in public agricultural research is a pre-requisite for successful application of biotechnology tools including MAS for crop improvement. Biotechnology tools can be used to enhance genetic gains for a few traits in a few crops, but their ultimate impact depends on how well they are adopted and integrated into existing plant breeding activities. This is a sobering thought, because in many developing countries public sector research capacity is being eroded and public sector extension services are being severely curtailed.

Other factors essential for the successful application of biotechnology tools are training and capacity building. Many biotechnology applications require learning new skills, some research infrastructure and effective operational capacity. It is especially important to train and nurture national scientists capable of using emerging technologies. In general, it may not be possible for older plant scientists to acquire the capacity for biotechnology applications. Therefore, policy-makers in developing countries have to consider long-term investments in training and nurturing a new generation of scientific talent. They also need to consider how to utilize this talent effectively by providing adequate resources and optimum work environments. Specialized technical training must in turn be underpinned by complementary government investments in basic education, e.g. by including biotechnology-related subjects in national university curricula.

Although it is widely assumed that enormous investments are needed to establish a capacity to carry out MAS, this is not always true. Certainly, a minimum level of investment is needed for laboratory facilities, equipment and trained staff. However, considering that most MAS work in developing countries is likely to be geared towards the use of existing markers rather than the development of new ones, investments in facilities and capital need not be

large. Developing countries are likely to have difficulty obtaining the required laboratory materials including consumables that are manufactured mostly in the industrialized world. Other factors such as local support for servicing and maintaining laboratory equipment and reliable basic services such as an uninterrupted power supply can also be challenging. In the less advanced developing countries, international research organizations and development assistance agencies will have a more significant role to play in ensuring the availability of the technology as well as the capacity to use it effectively, though on a limited scale.

Many developing countries are likely to use genetically modified cultivars with value added traits in the near future. Associated with transgenic technology are the complex, yet important, issues of biosafety and management of intellectual property. Policy-makers should therefore also consider ways of increasing the efficiency of publicly-funded research efforts, as well as finding opportunities and providing incentives for formulating productive public–private sector partnerships. As most tools of biotechnology that have potential practical applications are developed and patented by private industry, policy-makers have the challenges of addressing the need to forge research partnerships that allow the competitive private sector to maintain its interest in financial rewards while permitting technologies to be used by public sector researchers in relevant areas to serve farmers in species of importance that have so far been neglected. Coupled with these partnerships is the requirement to manage intellectual property issues.

In many situations, international development agencies are able to play a role in areas such as biotechnology priority-setting, raising funds for establishing the required biotechnology infrastructure and maintenance capacity, supporting public–private sector partnerships, and assisting in technology transfer and capacity building. International agricultural research institutes, which have had long-term involvement with national programmes in a large number of developing countries, should play a role in identifying key areas for contributing further in helping relevant national programmes identify, optimize and adopt MAS tools when it is feasible. International research centres can also play an active role in capacity building by identifying areas where it is needed and by providing necessary backstopping.

Novel marker systems based on SNP platforms are likely to bring the costs associated with MAS applications to an affordable level by many breeding programmes and it will be challenging to establish these technologies based on robotics and other automated, large-scale, screening platforms in many developing countries as the technology development and associated intellectual property rights remain in large private sector enterprises. This is an area where developing country policy-makers, together with international aid bodies and research organizations, should ideally work together to find partnerships with the private sector to devise ways of infusing these technological breakthroughs and associated benefits to the developing countries, at least on a limited scale.

In conclusion, MAS technologies have matured to the extent that they can be used for making genetic gains in certain traits and in some important crop species. National programmes in developing countries should evaluate the feasibility of applying MAS approaches for crop improvement as, despite the considerable limitations that exist in many developing

countries, the technology can be used at a relatively low operational cost. At least for major crops such as rice, maize, wheat and soybean, significant numbers of linked markers have been identified for genes of interest, and ongoing selection programmes have found them to be useful for making rapid genetic gains. Incorporating these tools into active breeding strategies will allow more rapid and efficient improvement of varieties for target traits.

As national programmes in developing countries vary in their capacities to absorb biotechnology tools, priority-setting and identification of MAS strategies should be done on a case-by-case basis, ideally supported by strong breeding programmes. Individual national programmes will have to be selective in their choice of technologies and markers to ensure that the level of investment is appropriate to justify the costs and produce the most rapid returns. This means that, while fully functioning biotechnology laboratories may not be feasible in all countries, initiating MAS is an important first step towards using modern biotechnology approaches in plant improvement. As the success of biotechnology applications depends on the existence of strong crop improvement programmes, policy-makers and international development agencies must ensure that the limited funds allocated to traditional agricultural research are not curtailed to support biotechnology activities. International aid agencies and agricultural research institutes should play a role in building research capacity within national programmes, encouraging public–private sector partnerships, and promoting technology transfer.

REFERENCES

Ayala, L., Henry, M., Gonzalez-de-Leon, D., van Ginkel, M., Mujeeb-Kazi, A., Keller, B. & Khairallah, M. 2001. A diagnostic molecular marker allowing study of *Th. Intermedium* derived resistance to BYDV in bread wheat segregating populations. *Theor. Appl. Genet*. 102: 942–949.

Bouchez, A., Hospital, F., Causse, M., Gallais, A. & Charcosset, A. 2002. Marker assisted introgression of a favorable alleles at quantitative trait loci between maize elite lines. *Genetics* 162: 1945–1959.

Brennan, J.P. 1989. An analysis of economic potential of some innovations in a wheat breeding programme. *Austr. J. Agric. Econ.* 33 (1): 48–55.

Byerlee, D. & Traxler, G. 2001. The role of technology spillovers and economies of size in the efficient design of agricultural research systems. *In* P. Pardey & M. Taylor, eds. *Agricultural science policy: changing global agendas,* pp. 161–186. Baltimore, MD, USA, The Johns Hopkins University Press.

Brown, S.M. & Kresovich, S. 1996. Molecular characterization for plant genetic resource conservation. *In* A.H. Paterson, ed. *Genome mapping in plants,* pp. 85–93. Austin, TX, USA, R.G. Landers Co.

Castro, A.J., Chen, X., Corey, A., Filichkina, T., Hayes, P., Mundt, C., Richardson, K., Sandoval-Islas, S. & Vivar, H. 2003. Pyramiding and validation of quantitative trait loci (QTL) alleles determining resistance to barley stripe rust: effects on adult plant resistance. *Crop Sci*. 43: 2234–2239.

Chavarriaga-Aquirre, P., Maya, M.M., Tohme, J., Duque, M.C., Iglesias, C., Bonierbale, M.W., Kresovich, S. & Kochert, G. 1999. Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundency in the cassava core collection and to assess the usefulness of DNA based markers to maintain germplasm collections. *Mol. Breeding* 5: 263–273.

Cregan, P.B., Mudge, J., Fickus, E.W., Danesh, D., Denny, R. & Young, N.D. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. *Theor. Appl. Genet*. 99: 811–818.

Dean, R.E., Dahlberg, J.A., Hopkins, M.S. & Kresovich, S. 1999. Genetic redundancy and diversity among 'Orange' accessions in the US sorghum collection as assessed with simple sequence repeat (SSR) markers. *Crop Sci*. 39: 22–32.

Dilday, R.H. 1990. Contribution of ancestral lines in the development of new cultivars of rice. *Crop Sci*. 30: 905–911.

Dreher, K., Morris, M.L., Khairallah, M., Ribaut, J-M., Pandey, S. & Srinivasan, G. 2002. Is marker-assisted selection cost-effective compared with conventional plant breeding methods? The case of quality protein maize. *In* R. Evenson, V. Santaniello & D. Zilberman, eds. *Economic and social issues in agricultural biotechnology*, pp. 203–236. Wallingford, UK, CABI.

Dreher, K., Khairallah, M., Ribaut, J-M. & Morris, M.L. 2003. Money matters (I): Costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. *Mol. Breeding* 11: 221–234.

Dubcovsky, J. 2004. Marker assisted selection in public breeding programs: the wheat experience. *Crop Sci*. 44: 1895–1898.

Dudley, J.W. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Sci.* 33: 660–668.

Eagles, H.A., Bariana, H.S., Ogbonnaya, F.C., Rebetzke, G.J., Hollamby, G.H., Henry, R.J., Henschke, P.H. & Carter, M. 2001. Implementation of markers in Australian wheat breeding. *Austr. J. Agric. Res*. 52: 1349–1356.

Fulton, T.M., Grandillo, S., Beck-Bunn, T., Fridman, E., Frampton, A., Lopez, J., Petiard, V., Uhlig, J., Zamir, D. & Tanksley, S.D. 2000. Advanced backcross QTL analysis of a *Lycoperssicon esculentum* x *Lycopersicon parvifolium* cross. *Theor. Appl. Genet*. 100: 1025–1042.

Gardiner, J., Melia-Hancock, S., Hoisington, D.A., Chao, S. & Coe, E.H. 1993. Development of a core RFLP map in maize using an immortalized F_2 population. *Genetics* 134: 917–930.

Gowda, M., Venu, R.C., Roopalakshmi, K., Sreerekha, M.V. & Kulkarni, R.S. 2003. Advances in rice breeding. *Mol. Breeding* 11: 337–352.

Han, F., Romagosa, I., Ullrich, S.E., Jones, B.L., Hayes, P.M. & Wesenberg, D.M. 1997. Molecular marker-assisted selection for malting quality traits in barley. *Mol. Breeding* 3: 427–437.

Hittalmani, S., Parco, A., Mew, T.V., Zeigler, R.S. & Huang, N. 2000. Fine mapping and DNA marker assisted pyramiding of the three major genes for blast resistance in rice. *Theor. Appl. Genet*. 100: 1121–1128.

Hoisington, D.A. & Melchinger, A.E. 2004. From theory to practice - marker assisted selection in maize. *In*: H. Lorz & G. Wenzel, eds. *Biotechnology in agriculture and forestry: molecular marker systems,* pp. 335–352. Berlin, Heidelberg, Germany, Springer-Verlag.

Huang, N., Angelels, E.R., Domingo, J., Mgapantay, G., Singh, S., Zhang, G., Kumaravadivel, N., Bennett, J. & Kush, G.S. 1997. Pyramiding of bacterial blight resistance genes in rice: marker assisted selection using RFLP and PCR. *Theor. Appl. Genet*. 95: 313–320.

Jenkins, S. & Gibson, N. 2002. High-throughput SNP genotyping. *Comp. Funct. Genom*. 3: 57–66.

Lagudah, E.S., Moullet, O. & Appels, R. 1997. Map-based cloning of a gene sequence encoding a nucleotide binding domain and a leucine-rich region at the *Cre3* nematode resistance locus of wheat. *Genome* 40: 659–665.

Liu, J., Liu, D., Tao, W., Li, W., Wang, C.P., Cheng, S. & Gao, D. 2000. Molecular marker-facilitated pyramiding of different genes for powdery mildew resistance in wheat. *Plant Breeding* 119: 21–24.

Maredia, M., Byerlee, D. & Maredia, K. 1999. Investment strategies for biotechnology in emerging research systems. Paper presented at 2nd *Ann. Conf. Internat. Consort. Agricult. Biotech. Res. (ICABR)*, Rome, Italy.

Moreau, L., Charcosset, A., Hospital, F. & Gallais, A. 1998. Marker-assisted selection efficiency in populations of finite size. *Genetics* 148: 1353–1365.

Morris, M.L., Ribaut, J-M., Khairallah, M. & Dreher, K. 2001. Potential impacts of biotechnology-assisted selection methods on plant breeding programs in developing countries. *In* P.G. Pardey, ed. *The future of food: biotechnology markets and policies in an international setting*, pp.197–218. Washington, DC, Johns Hopkins and IFPRI.

Morris, M.L., Ribaut, J-M., Dreher, K. & Khairallah, M. 2003. Money matters (II): Costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Mol. Breeding* 11: 235–247.

Pernet, A., Hoisington, D., Franco, J., Isnard, M., Jewell, D., Jiang, C., Marchand, J-L., Reynaud, B., Glaszmann, J-C. & Gonzalez-de-Leon, D. 1999a. Genetic mapping of maize streak virus resistance from the Mascarene source. I. Resistance in line D211 and stability against different virus clones. *Theor. Appl. Genet.* 99: 524–539.

Pernet, A., Hoisington, D., Dintinger, J., Jewell, D., Jiang, C., Khairallah, M., Letourmy, P., Marchand, J-L., Galszmann, J-C. & Gonzalez-de-Leon, D. 1999b. Genetic mapping of maize streak virus resistance from the Mascarene source. II. Resistance in line CIRAD390 and stability across germplasm. *Theor. Appl. Genet.* 99: 540–553.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5: 94–100.

Ragot, M., Biasiolli, M., Delbut, M.F., Dell'orco, A., Malgarina, L., Thevenin, P., Vernoy, J., Vivant, J., Zimmermann, R. & Gay, G. 1994. Marker assisted backcrossing: a practical example. *Techniques et utilizations des marqueurs moleculaires,* Les Coloques 72, pp. 45–56. Paris, INRA.

Ribaut, J-M. & Hoisington, D.A. 1998. Marker-assisted selection: new tools and strategies. *Trends in Plant Sci.* 3 (6): 236–239.

Ribaut, J-M. & Ragot, M. 2007. Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations and alternatives. *J. Exp. Bot.* 58: 351–360.

Ribaut, J-M., Jiang, C. & Hoisington, D.A. 2002. Simulation experiments on efficiencies of gene introgression by backcrossing. *Crop Sci.* 42: 557–565.

Ribaut, J-M., Hoisington, D.A., Deutsch, J.A., Jiang, C. & Gonzalez-de-Leon, D. 1996. Identification of quantitative trait loci under drought conditions in tropical maize: 1. Flowering parameters and the anthesis-silking interval. *Theor. Appl. Genet*. 92: 905–914.

Ribaut, J-M., Jiang, C., Gonzalez-de-Leon, D., Edmeades, G.O. & Hoisington, D.A. 1997. Identification of quantitative trait loci under drought conditions in tropical maize: II. Yield components and marker assisted selection strategies. *Theor. Appl. Genet*. 94: 887–896.

Sanders, J.H. & Lynam, J.K. 1982. Definition of the relevant constraints for research resource allocation in crop breeding programs. *Agric. Admin.* 9 (4): 273–284.

Smith, S. & Helentjaris, T. 1996. DNA fingerprinting and plant variety protection. *In* A.H. Peterson, ed. *Genome mapping in plants,* pp 95–110. Austin, TX, USA, R.G. Landers Co.

Tanksley, S.D. & McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063–1066.

Tanksely, S.D. &. Nelson, J.C. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet*. 92: 191–203.

Tanksley, S.D., Young, N.D., Paterson, A.H. & Bonierbale, M.W. 1989. RFLP mapping in plant breeding: new tools for an old science. *Biotech.* 7: 257–264.

Toenniessen, G.H., O'Toole, J.C. & DeVries, J. 2003. Advances in plant biotechnology and its adoption in developing countries. *Curr. Opin. Biotech.* 6: 191–198.

Villegas, E. 1994. Factors limiting quality protein maize (QPM) development and utilization. *In* B.A. Larkins & E.T. Mertz, eds.. *Proc. Intl. Symp. on Quality Protein Maize, 1964-1994,* pp 79–88. Sete Lagoas, Minas Gerais, Brazil, Embrapa/CNPMS.

Werner, K., Friedt, W. & Ordon, F. 2005. Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). *Mol. Breeding* 16: 45–55.

Willcox, M.C., Khairallah, M., Bergvinson, D., Crossa, J., Deutsch, J.A., Edmeades, G.O., Gonzalez-de-Leon, D., Jiang, C., Jewell, D.C., Mihm, J.A., Williams, W.P. & Hoisington, D.A. 2002. Selection for resistance to southwestern corn borer using marker assisted and conventional backcrossing. *Crop Sci.* 42: 1516–1528.

William, H.M., Trethowan, R. & Crosby-Galvan, E.M. 2007. Wheat breeding assisted by markers: CIMMYT's experience. *Eyphytica* (In press).

Xie, C. & Xu, S. 1998. Strategies of marker aided recurrent selection. *Crop Sci.* 38: 1526–1535.

Xu, Y., Beachell, H. & McCouch, S.R. 2004. A marker based approach to broadening the genetic base of rice in the USA. *Crop Sci.* 44: 1947–1959.

Young, N.D. 1999. A cautiously optimistic vision for marker-assisted breeding. *Mol. Breeding* 5: 505–510.

Yousef, G. & Juvik, J. 2001. Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. *Crop Sci.* 41: 645–655.

Yu, K., Park, S.J. & Poysa, V. 2000. Marker-assisted selection of common bean for resistance to common bacterial blight: efficacy and economics. *Plant Breeding* 119: 411–415.

Zhang, T., Yuan, Y., Yu, J., Guo, W. & Kohel, R.J. 2003. Molecular tagging of a major QTL for the fiber strength in upland cotton and its marker-assisted selection. *Theor. Appl. Genet*. 106: 262–268.

Zhou, P.H., Tan, Y.F., He, Y.Q., Xu, C.G. & Zhang, Q. 2003. Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor. Appl. Genet*. 106: 326–331.

Zhou, W-C., Kolb, F.L., Bai, G.H., Domier, L.L., Boze, L.K. & Smith, N.J. 2003. Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breeding* 122: 40–46.

CHAPTER 20

Impacts of intellectual property rights on marker-assisted selection research and application for agriculture in developing countries

Victoria Henson-Apollonio

SUMMARY

Although the impact of marker-assisted selection (MAS) in commercial and public sector breeding programmes in developing countries is to date limited to a few crops and traits, the potential benefits of using markers linked to genes of interest in breeding programmes for improving the productivity of crops, livestock, forest trees and farmed fish is substantial. While more recent methods associated with the use of MAS are technically demanding and often expensive, most applications of basic MAS were initially described in the literature, and thus will likely have very few intellectual property (IP) restrictions associated with their use, irrespective of the agricultural sector involved. For example, isolating DNA, amplifying specific gene sequences from that DNA (with most available primers), separating fragments using gel/polyacrylamide electrophoresis and imaging of fragments with standard techniques are likely to be available without restriction to scientists and breeders in the developing world, even as part of a commercial service. Problems arise when there is a need to use or develop high-throughput modes, which require more sophisticated technologies. For high-throughput use, a breeder will want to use the most efficient techniques that are currently available. This means that the more advanced processes/methods, reagents, software applications/simulations and equipment, which provide the most effective means to exploit MAS fully, are most likely covered by intellectual property rights (IPRs) such as patent rights, confidential information (trade secrets) and copyrights, both in industrialized countries and also in many developing countries such as Brazil, China and India. In situations where breeders wish to use cutting edge technologies and the most efficient markers, care must be taken to avoid activities that may infringe IPRs when using MAS methodologies.

INTRODUCTION

Other chapters in this book describe the usefulness and applicability of MAS for developing germplasm with superior qualities, in a timely manner. Markers have been developed and used by plant and animal breeders (Dekkers, 2004), for fish and shellfish (Consuegra and Johnston, 2006) and for forest trees (Kellison, McCord and Gartland, 2004; Lee, A'Hara and Cottrell, 2005). Introduction of MAS to developing country scientists has been taken up by a variety of projects such as the Generation Challenge Programme (cgiar.org/exco/exco8/exco8_generation_report), supported by the Consultative Group on International Agricultural Research (CGIAR) and MAS jamborees sponsored by the Syngenta Foundation for Sustainable Development (syngentafoundation.org/pdf/Report%20Nairobi%20meeting%20.pdf and T. St. Peter, personal communication). MAS is also being used by many of the centres belonging to the CGIAR, notable examples being the International Center for Tropical Agriculture (CIAT), the International Potato Centre (CIP), the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and the International Institute for Tropical Agriculture (IITA), as well as the International Maize and Wheat Improvement Centre (CIMMYT), in programmes such as the Asian Maize Biotechnology Network (AMBIONET), and the International Livestock Research Institute (ILRI) in the areas of livestock production and health through the Biosciences Facility for east and central Africa (BecA).

In this chapter, a brief review of general intellectual property law is used to introduce a variety of aspects regarding intellectual property potentially associated with the use of the techniques, reagents and equipment that are necessary for implementing MAS. This intellectual property "primer" is followed by a description of specific cases and some recommendations regarding steps that should be taken by scientists and breeders in developing countries who are contemplating using MAS in breeding programmes to avoid restrictions or incurring risks of infringing the intellectual property rights (IPRs) of others.

INTELLECTUAL PROPERTY RIGHTS AND PUBLIC ACCESS TO INNOVATION

IPRs are awarded on the basis of national laws. There are, however, a few examples of regional cooperation institutions granting IPRs on a regional basis, such as the African Regional Intellectual Property Organization (ARIPO), the Gulf Cooperation Council (GCC) and the European Patent Office. In addition, under the Patent Cooperation Treaty (PCT), an international agreement administered by the World Intellectual Property Organization (WIPO) that facilitates patent filing, a single international application can be filed in a national PCT-receiving office, which can then subsequently be submitted to all PCT member national patent offices.[1] In addition, an example situation is given in Box 1 that illustrates the, perhaps unexpected, "far reach" of national patent law.

IPRs comprise original and novel assets that involve the use of human intellect. The awarding of such rights is intended to balance the needs of society to access and use the products of human ingenuity, with rewards for the endeavours going to the individuals from whom these intellectual assets originated. Obviously, there is a certain amount of tension in this

[1] See www.wipo.int/pct/en/texts/pdf/pct_paris_wto.pdf for a list of countries that are members of important IP international treaties, including the PCT.

balance between private rights and the needs of society (Murchie *et al*., 2006). Society is presumed to benefit from public disclosure in the form of patent disclosure requirements and copyrights, which are awarded to creative works that have been fixed (made tangible).[2] In addition, through the combination of the requirement of full disclosure in the written description (manifested in a patent application), and the time limitation over patents rights, inventions are put into the public domain when the rights expire. The pharmaceutical industry's experience with the success of "generics" is a testament to the value of "expired" inventions (CBO, 1998). In some specific cases, however, such as patents on certain drugs, rights may be extended for a certain period of time upon request to compensate for long delays in obtaining authorization for drug commercialization, e.g. the "Hatch-Waxman" act in the United States of America. The filing, prosecuting and maintenance of patents are business decisions that are put into place as a part of the strategy for bringing products to the consumer. An additional part of such a strategy can also include a plan for maintaining profitability when patent rights expire (Smyth, 2006). For example, depending upon the creativity of inventors, it may be that improvements allow for the filing of additional patents to cover these improvements, thus having the effect of extending patent rights for additional terms. This is a fact-based process, in that the improvement must meet the requirements of invention.[3]

The balance between public and private rights is considered by some to be tilted in favour of private rights, leaving elements of some societies wondering if IPR systems work at all except to protect the monopolies that they award (Epp, 2004). A number of civil society organizations are monitoring the potential effect of changing India's patent law to include patents over pharmaceutical products and agricultural chemicals (Sreedharan, 2007).

INTELLECTUAL PROPERTY LAW AS IT RELATES TO MAS

The standard steps employed in MAS generally include: selecting individuals to be tested; harvesting material; extraction of DNA from the material; polymerase chain reaction (PCR) amplification of the DNA to enrich for gene sequences/fragments associated with a particular trait or phenotype; separation of these fragments; visualization/identification of DNA fragments; and interpretation and utilization of the information. Each of these stages involves certain methods and the use of particular reagents and/or equipment associated with the particular methodological steps. For the purposes of this chapter, a series of tables (numbers 1–3) has been prepared to exemplify the types of intellectual property and associated IPRs that exist for materials and/or processes within each of these seven steps.

There is a general set of categories of IPRs that are awarded in most countries/jurisdictions. These include industrial or utility patent rights, plant variety protection/plant breeders' rights, copyrights, rights of appellation/geographic indications, trademarks and secrecy rights (trade secrets) associated with undisclosed or confidential information. Other types of patent rights can be awarded in many jurisdictions.

[2] Not that efficient and sincere disclosure is not without problems – see Fromer (2007).

[3] Note that concerns regarding the abuse of the patenting of incremental changes versus incremental improvements are often raised by a practice of patent time extension called "evergreening". For further discussion of evergreening from the view of generic pharmaceutical manufacturers, see Hore (2004).

For example, in addition to utility patents, two others types of categories of patents are available to inventors in the United States: a design patent for a new original or ornamental design for an article of manufacture, granted to protect the external appearance rather then the function of a product, and plant patents, awarded for the invention or discovery of a cultivated plant variety that can be asexually reproduced, (except via tubers, but including grafts and spores). Other countries have additional categories regarding subject matter (e.g. designs, plants) and also with respect to examination rigour and length of the patent rights grant (e.g. "short-term" patents in Belgium and the Netherlands (see e.g. www.ipr-helpdesk.org/docs/docs.EN/invencionesTecnicasBP.pdf), and innovation patents in Australia (www.ipaustralia.gov.au/patents/what_innovation.shtml).

Patent rights are awarded to inventions on the basis of criteria associated with usefulness (industrial applicability), originality (newness or novelty), and an "inventive step" (non-obviousness to persons with technical skills in the particular field where the invention is applicable). There are also rules governing the subject matter of the invention for utility patent rights to be awarded. For example, all countries' patent rights prohibit the awarding of patent rights for elucidating the "laws of nature". Thus, the fact that scientists have described laws of chemistry and physics, natural selection, or other such natural laws, does not render them as products of a person's intellect in intellectual property law. However, an innovation that applies one of these laws may well qualify for protection. Similarly, in many countries a new plant variety, a variety, type or breed of livestock used for food production, or computer software cannot be the subject of patent rights. Japan and the United States are notable exceptions in this regard. While the European Union (EU) (Directive 98/44/EC of the European Parliament and of the Council, 1998 on the legal protection of biotechnological inventions) does not permit the patenting of plant and animal varieties, it does allow patents for inventions concerning animals or plants the feasibility of which is "not confined to a particular plant or animal variety". The fact that the term "variety" is not well defined in the context of animal breeding means that the scope of this exemption is far from clear.

Irrespective of whether one is dealing with patent rights, plant breeders' rights (PBRs), copyrights, trademarks, trade secrets, etc., the type of IPR sought or awarded varies with the type of intellectual asset over which protection is being sought. It is also possible for one asset to be protected by several types of rights, depending upon the law in the applicable territory. For example, it is not unusual to have "double protection", i.e. for an invention to be patented and the product resulting from that invention to be trademarked. The trademark for Aspirin® for the formerly patent-protected acetylsalicylic acid is such a case in many parts of the world. It is not uncommon for a process or a piece of machinery to be treated in a similar fashion. This situation pertains to IPRs associated with MAS, two notable examples being "Selective restriction fragment amplification: a general method for DNA fingerprinting", a patented process paired with rights associated with the AFLP® trademark or the "Methods for genotyping by hybridization analysis" patent and the associated DArT™ trademark.

PATENTS

Patent rights are awarded on the basis of claims based on the inventor's description

to explain the new, non-obvious patentable subject matter in a way that clearly distinguishes its novel characteristics from all other available solutions. This explanation is called a patent "claim", and using the words of the patent drafter a claim will describe the "metes and bounds" (Gallagher, 2002) of the invention. Patent drafters are usually licensed patent agents, patent attorneys, scientists working for legal firms in this capacity or, rarely, the inventors themselves. Drafting patent claims is an arcane art that requires detailed knowledge of the scientific and technical basis of the invention as well as a current understanding of the state-of-the-art, regarding the judicial interpretation of claims, in the context of national patent law.

One patent can have many claims. In fact, patent law requires that every patent must contain at least one claim. Each claim is "directed to" an invention, ranging from its broad use, to the most narrow use for which an inventor may wish to seek rights. For example, a broad claim could be for the use of an enzyme class to perform a type of function (where this combination is not found in nature). A narrow claim could then specify the particular enzyme, the quantity of enzyme and/or the specific function. A distinction should be made between a patent application (often numbered in a different style such as the "WO" designation for PCT-filed patent applications), and an issued patent (generally numbered with a country prefix, e.g. CA 2172863, a patent issued by the Canadian Patent Office) to avoid confusion.

Patent applications contain claims that are untested and unexamined and these claims are therefore often very broad. During the patent prosecution process, the patent examiner seeks to limit claims to the new invention held by the applicant at the time the patent was filed. The claims are accompanied by written descriptions that would allow someone else familiar with technology in the same general area ("person having ordinary skill in-the-art" or "PHOSITA"), to understand how to make and carry out or "work" the claimed innovation. This useful written description accompanying claims is directed by law to provide "enablement", and is a required part of a patent disclosure, in order to make the invention "available to the public" (this is part of the social contract to balance private rights and public good). The written descriptions can also be important for interpreting the exact limits of patent claims. Patent rights are given to inventions that cover the reduction of ideas and concepts to practical use, and these rights may also extend to other treatments/variations that are of a nature sufficiently similar to be equivalent to the patented innovation. Such a "doctrine of equivalents", as it is called in patent lingo, means that ideas/concepts that are the basis of the useful innovation are a part of the patent claim coverage. Therefore, it is often stated that patents cover conceptual ideas as well as the practical application of the idea (see www.dwalkerlaw.com/patent.asp). This means that it is often difficult to discern whether a party is committing infringement without the interpretation of a court. Literal infringement, whereby the invention is practised exactly as it is described in a claim, can usually be identified without a problem. Equivalent infringement is often used as a strategic business tool by either the patent rights holder and/or the infringer. This confusion over the exact limits of patent claims can often lead to company mergers or buy-outs, just to minimize the risk associated with the IPRs (Fulton and Giannakas, 2001; Kattan, 2002).

BOX 1

Territoriality of patent rights

Developing country scientists and breeders should be aware that patent rights are only enforceable within the jurisdiction of the country or countries where the patent rights have been awarded. The caveat to this is that patent laws in most countries cover material that is imported into a country when patent rights exist on that material in the country where the importation would take place. The language that is included in such patent laws contains the terms: "making", "selling" or "using" within a country's boundaries. For example, if patent rights over the formula for a particular herbicide had been awarded in Country AA, but no patent rights over this same herbicide composition had been awarded in Country BB, then the herbicide could be made in Country AA only with the permission of the patent rights holder. However, the herbicide could be made in Country BB without permission of the rights holder in Country AA; no infringement would be possible in Country BB. If someone wanted to import the herbicide that was made in Country BB into Country AA, then the importer in Country AA would need to obtain permission (a licence) from the rights holder in Country AA.

The situation for Argentinian soybean containing a transgene covered by patent rights issued to Monsanto in Europe is a good illustration of the territoriality of patent rights. Monsanto holds plant breeders' rights over the variety, but does not have patent protection for the gene in Argentina. Many farmers in Argentina are growing herbicide resistant soybeans developed by Monsanto, (often using seed multiplied by companies that do not have a licence from Monsanto). The company has taken the strategy of preventing the importation of Argentinian-grown soybeans or *products* made from Argentinian-grown soybean into any country where Monsanto has patent rights by informing potential buyers of Argentinian-grown soybeans that they will be infringing Monsanto's patent rights if they bring such material into a country such as the United States or an EU country, where Monsanto has patent rights over the technology embedded in the seed or over the seed itself (Balch, 2006), and therefore also present in the soybean imported grain. Monsanto's patent covers the final product, that is the gene, and extends its protection to the seed and the grain containing the gene sequence. The European Commission (EC), in fact, recognizes the right of Monsanto to prevent import of the soybean grain, but not the soybean flour, where the gene sequence can no longer be expressed.

What, however, is the relevance of such action to MAS, where there is no technology embedded in the seed, remaining in the seed itself? Patent law is usually interpreted to cover any material where a patented technology was used to produce a product, even though such a product does not literally contain the technology. This means that in most situations, if patent-protected techniques, methods, processes or products are used in a MAS scheme, the resulting products are covered by these patent rights. Of course, this type of infringement can be very difficult to prove and therefore is rarely the subject of a legal suit, but the risk is present and occasionally is enforced (AsiaLaw, 2004). However, for developing country farmers who are not going to be exporting a product to an industrialized country, in actuality, the risk of an infringement is minimal (Binnebaum *et al.,* 2003). Nevertheless, the situation of using a patented invention without permission of the patent rights holder is not straightforward and, if such a course involves public resources, it should only be embarked upon on the advice of an IP counsel or an IP lawyer.

TABLE 1
Examples of patents relevant to MAS

Technique		Selected patent examples[1]	Public domain equivalent	Status of selected patent example	Implications
Harvesting DNA	Use of silica particles	US 5 234 809	Many, e.g. Doyle and Doyle, 1987 and Saghai-Maroof *et al.* 1984	In effect in the United States; related patents in effect in: Austria, Australia, Canada, Denmark, Germany, EU, Greece, Japan, Republic of Korea, Netherlands, South Africa and Spain	Licence needed; often supplied with reagents, kits and/or equipment such as thermal cyclers
	Equipment "Matrix Mill"	US 6 063 616	Many, e.g. Edwards, Johnstone and Thompson, 1991. Combination with centrifuge tube	In effect in the United States	If specialized equipment is used, licence may be needed. Likelihood that coverage would extend to developing country areas
Amplification of specific DNA seqs	Reagents	US 4 683 195	None	Expired in all countries (therefore in public domain)	Advance or improvement likely will require licensing, many even in developing countries
	Primers/genes	Primers for identifying Soybean Sudden Death Resistance US 6 300 541	Many e.g., Röder *et al.*, 1998. *gwm493* in wheat	In effect in the United States	Sequence(s) to be used should be checked by a patent searcher such as Gene-IT.com if breeding product is valuable and would be grown for export
	Equipment	Applied Biosystems Thermal Cycler US 5 656 493	Other equipment is available; contentious legal issues associated with many	In effect is the United States and most other developed countries, and a few developing countries including Brazil, China, Republic of Korea, South Africa	
Identification of marker genes	Reagents	Agarose, no applicable patents found	Polyacrylamide	No patent rights on traditional gel/acrylamide media	
	Equipment	Charge-coupled device imaging apparatus US 5 672 881	Cameras	Many systems that are no longer under rights protection	
MAS methods, in general	Methods and compositions	Use of selective DNA fragment amplification products for hybridization-based genetic fingerprinting, MAS, and high-throughput screening. US 6 100 030 QTL mapping in plant breeding populations. US 6 399 855	Numerous	In effect in the United States	Likely defensive patents. Could be problematic with imports to the United States

[1] There will inevitably be innovative improvements or technological advancement associated with each of these methods and materials, many of which will have been awarded IPRs to the inventor and/or the inventor's company.

Examples of published patents where rights have been awarded in the area of MAS include the basic PCR amplification process patents in the United States, US Patent nos. 4 683 195, 4 683 202 and 4 965 188, originally issued to the Cetus Company and then assigned to Hoffman-Roche in 1992, on the use of DNA polymerase based on the Taq polymerase enzyme isolated from the organism *Thermus aquaticus.* As these amplification patents expired worldwide in March 2006, when only the basic techniques and reagents covered by these patents are used, one does not now have to be concerned with infringement of these patents anywhere. However, the equipment used to control the reaction conditions may also carry IPRs on its own and most PCR techniques currently used are patented as improvements to the basic technology. For example, Applied Biosystems' PCR and real-time instrument patents and other PCR-related patents such as US Patent no. 5 656 493, are still in effect. A licence to these instruments and other patents may be needed in the United States in order to use their thermal cyclers to carry out PCR, although this is normally granted as part of the purchase price of the equipment and reagent kits. Table 1 contains additional examples of selected patents that are associated with MAS.

Another strategy that should be pointed out is the concept of "defensive" patents. Patent rights may be awarded in most jurisdictions over processes (actions/processes), and machines, manufactures and compositions of matter (things). Enforcement of patent rights, e.g. bringing a lawsuit against a person or forcing a licensing situation when a person is practising your invention (infringing your rights) without permission is less equivocal when the infringement involves making, using, possessing or selling an object or composition. However, the detection of infringement of methods claims is often much less straightforward. A patent owner would need to have insight into or gain access to how something was made or formed by the other party (potential infringer), in order to know whether his/her patented process or method was being used. This means that it can be even more costly and time-consuming to pursue potential infringers of methods claims than lawsuits involving infringement of making, buying or selling a patent-protected material or composition. Thus, sometimes a company or institution will decide to file a patent application, seeking rights over a method where such a filing will simply represent an attempt to preclude a competitor from preventing the company from carrying out a method, without concerns of infringement. Such a method or process patent would likely never be enforced except in blatant infringement and is only sought to provide insurance for the filing organization to lower the risk that the organization will be sued by someone else. The distinction between a patent that is filed defensively and one that is filed to prevent someone from practising the claimed invention can be very subtle. A discussion of patenting strategies including defensive patents can be found at www.271patent.blogspot.com/2006/09/valuing-patents-and-patent-paradox-why.html. This is an area of patent law that is always in flux and enforcement can be very complicated and expensive.

COPYRIGHTS

These rights are awarded for creative innovations that are "fixed" in a printed, video, audiotape or other recorded form. Copyrights only cover the form of the fixation, and not the ideas or concepts

TABLE 2
Examples of copyrighted software relevant to MAS

Technique	Use	Selected software examples	Licensing conditions	Comments
Analysis of QTLs	For use primarily in analysing animal pedigree associations	Loki www.stat.washington.edu/thompson/Genepi/Loki.shtml	Very liberal, freeware-type of licence www.stat.washington.edu/thompson/Genepi/license.shtml	To be downloaded only if licence is accepted by user
	List of open source or freeware	www.stat.wisc.edu/~yandell/qtl/software/	Open source or as freeware	Source code provided.
Analysis of fragment patterns	For use with ABI electrophoresis equipment	Genotyper®	Usually licensed with ABI equipment purchase. (Applera Corporation). Additional individual personal copies cost ~ US$1 500. Stand-alone copy costs ~ US$5 000. Software manual is also licensed with software	Source code is not provided; explicit prohibition in licence
Creation of binary table of fragment patterns	For use with Genotyper®	PeakMatcher http://crop.scijournals.org/cgi/content/full/42/4/1361	Licensed under GNU-GPL v 2	Source code is provided
Analysis of fragment patterns	For use with electrophoresis with fluorescently labelled markers	Genographer http://hordeum.msu.montana.edu/genographer/	Licensed under GNU-GPL v 2	Source code is provided
Genotyping software for linkage mapping applications	For use with ABI electrophoresis equipment	GeneMapper®	Licensed by ABI (Applera corp.) with equipment. Manual is licensed with software. Manual has own independent copyright	Source code is not provided; explicit prohibition in licence
Simulation of biophysical processes in farming systems	Predictive software	ApSim	See, www.apsru.gov.au/apsru/Products/APSIM/Access%20and%20Pricing%20Policy.pdf Also an annual licence fee	Reduced licensing fee (on a case-by-case basis for NARS)
Simulation platform for quantitative analysis of genetic models	Predictive software	QuGene Original Reference:http://bioinformatics.oxfordjournals.org/cgi/reprint/14/7/632.pdf	Now only available under licence from University of Queensland/CSIRO	Reduced licensing fee (on a case-by-case basis for NARS)

associated with the innovation (as is the case with patents). Although articles written about MAS, drawings of breeding schemes and the like would be products for which copyrights are awarded, it would be quite rare for someone to be concerned about infringing copyright in carrying out MAS. However, most MAS as currently practised, especially at high-throughput levels, involves the use of computer software to analyse the often complex data that result from marker detection. While software applications can be patented in a few countries, most jurisdictions only allow software to be covered by copyrights. (In many jurisdictions, there is ongoing discussion regarding whether software code is an appropriate matter to be covered by copyrights. While in Europe, the EC Directive on the Protection of Computer Programs (91/250EEC) has clearly established that in the EU, computer programs are protected on the same basis as literary works, other countries have a more checkered history [Starkoff, 2001].) Such copyrights are used as the basis for "Open Source" licensing of software. Most software used in conjunction with MAS must be licensed before it can be utilized in MAS breeding schemes or analysis.

The ethical aspects of copyright should also be understood. For example, breeders need to be respectful and careful when giving talks or other presentations to ensure that the material they use is original, or that the owner of the copyright has given permission for its use. Just because there is no "©" sign on an article, drawing, slide, picture, etc. does not mean that the material has not been copyrighted. Copyright is attached to almost any fixation with immediate effect. There is no need for an author or creator (or employer of the creator), to apply for copyright in most countries because of the conditions set forward in the Berne Convention for the Protection of Literary and Artistic Works (1886), which requires its signatories to protect the copyright on works of authors from other signatory countries in the same way it protects the copyright of its own nationals. A main principle of the Berne Convention, and incorporated into the WTO's Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPs), is the general principle of national treatment, "which requires each member state to accord to nationals of other member states the same level of copyright protection provided to its own citizens" (www.wipo.int/treaties/en/ip/berne/summary_berne.html). There are exceptions, e.g. publications that originate from the United States Federal Government cannot be covered by copyright, although sometimes copyright owners will register a copyrighted article with the government to take advantage of governmental assistance in infringement cases. Table 2 provides some examples of copyrighted materials that have relevance to the practice of MAS.

TRADEMARKS

These are registered marks given to an applicant as a result of a trademark application being made with a fee payment, and such an application withstanding a search by a trademark examiner for similar marks and use of marks (along with an opportunity for opposition to the awarding of the exclusive use by anyone in the public, based on use of the mark by someone else prior to the application to the trademark office). Trademark rights are different from patents, plant variety rights and copyrights, in that they are renewable, and thus, if national procedural rules are followed correctly, can likely last indefinitely. As

TABLE 3
Examples of trademarks relevant to MAS

Mark	Holder	Product	Use	Comments
AFLP®	KeyGene	Associated with the AFLP process/method and reagents	Creation of polymorphic markers based on difference in DNA sequence	Widely used system; developing country institutions often negotiate a low/no cost-licence on a case-by-case basis
DArT®	CAMBIA	Diversity array technology	Selection of markers based on variation from reference panels	Proprietary technology, often licensed under a BiOS licence
ABI®	Applera Corp.	Instruments such as capillary electrophoresis	Various electrophoresis equipment, sequencers, etc.	Widely used, associated with many patented technologies
Sybr®	Invitrogen	Fluorescent dyes	Visualization of DNA fragments	Widely used, patent on original dye in this series has expired in many jurisdictions
GeneChip®	Affymetrix	Microarray on glass substrate	Microarrays can be used for detection of nucleic acid sequences – DNA or RNA	Widely used methodology. Affymetrix one of the leaders in this field

mentioned earlier, "AFLP" is one example of a trademark. This means that in practice, when this method is referred to, the "®" symbol should accompany the term, i.e. the correct use of the term would be AFLP®. Another relevant example would be the Certified Angus Beef® protected by federal trademark law in the United States. In addition, the names of new markers, new varieties or types of crops, livestock, etc. would need to be checked by a professional trademark searcher if a breeder wished to be sure that no trademark infringement might occur by such naming. This is not precluding the fact that PBRs legislation requires the breeder to give its candidate variety a denomination that cannot be registered as a trademark, as it remains the generic denomination of the variety. Table 3 contains examples of trademarks that are often used as "brand" names, associated with products/processes used in MAS technologies. Commercial MAS practitioners need to be aware that use of a trademarked name in conjunction with a product requires the permission of the trademark holder.

TRADE SECRETS AND CONFIDENTIAL INFORMATION

These are not registered and, although considered to be non-statutory IPRs, they are protected by trade secret law in most countries. Crop breeders have used this approach for many years to protect the parent lines and information used to produce hybrid seeds for sale, and similar approaches are adopted in the poultry and pig industries. This type of IP is defined as commercially useful information that can be said to have the qualities of being any method, technique, process, formula, programme, design or other information that may be used in the course of production, sales or operations. It must also meet requirements such as not being known to persons generally involved in information of this type; having an actual or potential economic value due to its secretive and useful nature; and the owner has taken reasonable measures to maintain its secrecy. Infringement or non-authorized disclosure/use or misappropriation of a trade secret can result in criminal penalties. These rights might be of concern to scientists and breeders who are

working under conditions that require the use of confidentiality agreements or non-disclosure agreements (NDAs). Examples include MAS work being carried out by an employee of a company that requires employees to sign confidentiality agreements, or MAS carried out as part of joint work where breeders have been required to sign confidentiality agreements.

This is a very common type of protection used by commercial breeding companies involved in the development and use of markers and software in all sectors of agriculture. If a company becomes concerned that a trade secret risks being exposed, it may file a defensive patent application to ensure that a competitor will not obtain rights that would preclude use of its own trade secret. Obviously, when a patent application is filed on an invention that includes confidential information, the information will no longer be a trade secret. The applicant presumably would only resort to such a move if the possibility of "independent invention" were high, and thus the risk of disclosure in a patent application balances the risk of having the competition "know" of your trade secret. This will happen because of the way in which patent examiners normally decide if an invention is "new". Often such decisions are based upon the national IP law's definition of "new", as in the United States where there is a grace period of one year to file a patent application after an invention is made public and also where only use within the United States is considered to render an invention "not" new. A patent examiner cannot know that an invention has been used or described prior to the filing of a patent application if the invention is kept as confidential information. Therefore patent rights could be awarded to someone who actually copies a trade secret and companies must then consider filing for a patent or run the risk that a secret will be the subject of a competitor's patent.

Why would a company not simply file a patent application for each marker that it identifies? There are several strategic reasons. It is expensive to file for patent protection and, also, the applicant must disclose the invention and all of the specifics of the invention to satisfy the written description requirement of enablement. For a marker, this means that the applicant would need to disclose its nucleic acid sequence if it is known and, by wanting the rights over the use of the marker in MAS, also the trait(s) that is(are) associated with the presence (or absence) of detection of the marker, etc.

Obviously, it is impossible to list specific trade secrets that exist in MAS technology, although one indication of the existence of these can be a reference to a "personal communication" as, for example, in the case of the "15PICmarq" marker listed in Table 1 of the paper by Dekkers (2004). However, there are examples of information that is of the opposite nature, i.e. information that is publicly available and that can be used without permission because it is in the public domain such as information published by the United States Federal Government, or because no attempts are made to enforce rights. The company www.resgen.com, for example, sells kits comprising simple sequence repeat (SSR) primers mainly for use as MAS markers for many different species and based on sequences that have been published. These may therefore be covered by copyright, but these rights are not enforced.

CONTRACTUAL ARRANGEMENTS

An additional, "non-statutory" system of rights (Ricketson, 1984 as referenced

in Drahos, 2005), such as rights/requirements covered by conditions associated with a contract is often described as an IPR, although technically these types of rights or conditions are not the subject of IP law in most countries, but rather are a part of legal codes that deal with private rights. These requirements might be of concern to breeders working under conditions that require the use of contracts such as material transfer agreements (MTAs). Conditions that result from entering into agreements or contracts could carry a minimum level of awareness of the duties or responsibilities incurred by one agreeing to the terms. Other "non-statutory" rights could include contractual/legal terms, such as those included in a licence or a "Technology Use Agreement" (TUA). Enforcement and practice associated with contract law vary in all jurisdictions and can even vary at the local level. O'Connor (2006) has recently pointed out the degree to which MTAs are used to confer a licence to both patent rights and biological materials themselves. He refers to this arrangement as a "lease-licence" model wherein the IPRs and the physical property rights are "woven" together. Again, if the documents are read carefully, these conditions will not take anyone by surprise, in that they are a part of a contract or licence or other permission granted by an owner or provider of material. However, sometimes this permission may be agreed to in a manner that does not make a strong impression on a recipient. For example, the so-called "shrink-wrap" licence that accompanies software, or the "click-wrap" licence that covers software or other material downloaded from the Internet, may be too subtle for most people to be really aware that they have agreed to a licence. In agriculture, "bag-tag" or "seed wrap" licences exist that have the same sort of connotation (Kershen, 2004). Many courts have looked at the enforcement of these licensing/contract issues, with slightly varying results. The web site www.lex2k.org/shrinkwrap/shrinkwraprev.html describes individual cases and discusses these cases with regard to enforceability of "shrink-wrap" contracts in different jurisdictions and conditions.

EXAMPLES OF IPR PRACTICES ASSOCIATED WITH THE USE OF MAS AND RECOMMENDATIONS FOR SCIENTISTS AND BREEDERS

The type of formal IPRs most likely to cause a problem with the utilization of MAS are patent rights. Some examples of patents in this area are given in Table 1. Patents/patent applications are also listed in the paper by Concibido, Diers and Arelli (2004). Also, as mentioned in the preceding section, contractual arrangements/obligations may interfere with unfettered use of products and processes associated with MAS.

Patent rights have been awarded for most of the materials and methods that are involved in practising MAS within all fields of agricultural production. A careful researcher will choose methods and marker sequences that have been published and then carry out at least a cursory search of patent databases such as the European Patent Database (http://ep.espacenet.com) to make a first pass for determining the likelihood that the method and/or sequence(s) of choice are not covered by patent rights in the jurisdiction where they work. Depending upon the level of risk that one is willing to assume, for work that could result in a commercial product, more investigation is likely needed and perhaps the services of a patent information specialist (see www.piug.org/) or an IP lawyer will be required.

Most patents will be of concern primarily to those in developed countries, particularly the United States where many private companies have their base. For example, taking the company Pioneer, 209 US patents assigned to Pioneer are identified when the US Patent Database is searched for the terms "breeding" in the patent and "marker" in the claims. This is reduced to eight when the additional term "assisted" is searched in the claims of these 209. At the time of writing, Pioneer had 46 published US patent applications covering the "breeding"+"marker" category; reduced to one with the addition of "assisted". Monsanto, Bayer and Syngenta have utilized MAS practices for a number of years and accumulated patent portfolios and very likely many trade secrets in perfecting MAS techniques for their particular uses (Cahill and Schmidt, 2004). Monsanto announced in February 2007 that it would begin sharing its markers for soybean cyst nematode (SCN) resistance with academic and public institution researchers worldwide. According to the announcement, "Academic researchers and public institutions who request access will be given a royalty-free licence for using the rhg1 marker under a patent that was granted to Monsanto in December 2006 (US Patent no. 7 154 021)". It is of interest to note that the company, Genome and Agricultural Biotechnology, LLC, with five issued US patents and five US patent applications covering SCN inventions, has been sued for patent infringement in the use of SCN markers in conjunction with MAS (Genome and Agricultural Biotechnology had sought patent protection in order to establish "freedom-to-operate" testing services for material supplied by breeders who lack the facilities to perform MAS techniques for assessing the presence of particular disease-resistance alleles [www.siuc.edu/~psas/faculty/pubs/lightfoot_achv.htm]). As this situation indicates, persons wishing to establish their rights to use markers, by filing patent applications and even obtaining patent rights, need to understand that one cannot presume that an issued patent means that one then can practise the inventions, described in the claims, without concern that one may also be engaging in infringement of another patent or set of claims that have been allowed in other patents.

As of February 2007, a cursory search of the US Patent Database as an indicator of overall patenting activity related to MAS and plants revealed 372 issued patents and 112 published US patent applications. Of these 112 US patent applications, 79 were associated with plant breeding and 33 with animal MAS.

These numbers do not include most of the patents covering equipment, PCR and PCR-related technologies like AFLP®, such as US Patent no. 6 045994 that may be especially useful for generating markers. Also, analysis of the data indicates an increase in the number of applications submitted over the four years up to 2005, but most of these applications (58 percent) are for IPRs over specific plant varieties and sets of markers that allow identification of the germplasm variety. In recent years many patents have been granted that cover genes and markers associated with economically important traits in livestock species (Rothschild, Kim and Anderson, 2006; Barendse and Reverter-Gomez, 2007).

Potential commercialization of such inventions was predicted by Rothschild, Plastow and Newman (2004), as well as the associated development of inventions for methods covering breeding management

BOX 2

Representative claims that illustrate the breadth of patent claims over sequence information

US 6 235 972 , "Maize Rad23 genes and uses thereof" issued 22 May 2001

What is claimed is:

1. An isolated RAD23 polynucleotide comprising a member selected from the group consisting of:
   (a) a polynucleotide having at least 85 percent sequence identity to the polynucleotide of SEQ ID NO: 1; wherein the percent sequence identity is based on the entire region coding for SEQ ID NO: 2 and is calculated by the GAP algorithm under default parameters;
   (b) a polynucleotide encoding the polypeptide of SEQ ID NO: 2;
   (c) a polynucleotide encoding the polypeptide of SEO ID NO: 4;
   (d) a polynucleotide amplified from a Zea mays nucleic acid library using primers which selectively hybridize, under stringent hybridization conditions, to loci within the polynucleotide of SEQ ID NO: 1;
   (e) a polynucleotide which selectively hybridizes, under stringent hybridization conditions and a wash in 0.1.times.SSC at 60 degree C., to the polynucleotide of SEQ ID NO: 1;
   (f) the polynucleotide of SEQ ID NO: 1;
   (g) the polynucleotide of SEO ID NO: 3;
   (h) a polynucleotide which is complementary to a polynucleotide of (a), (b), (d), (e), or (f);
   (i) a polynucleotide which is complementary to a polynucleotide of (c) or (g); and
   (j) a polynucleotide comprising at least 75 contiguous nucleotides from a polynucleotide of (a), (b), (d), (e), (f), or (h); wherein the polynucleotides of parts (a), (d)-(e), (h)-(j) each encode monocot Rad23 polypeptides.

US 6 815 578 "Polynucleotide encoding MRE11 binding polypeptide and uses thereof" issued 9 November 2004

Claim 9. An isolated polynucleotide comprising of polynucleotide selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide having at least 95 percent sequence identity over its entire length to SEQ ID NO: 2; as determined by the GAP program under default parameters, wherein the encoded polypeptide binds to a MRE11 polypeptide; and

(b) a polynucleotide which is fully complementary to the polynucleotide of (a).

and breeding–related computer applications (Schaeffer, 2002).

Previous search of patent literature

As mentioned earlier, and irrespective of the agricultural sector in which they are operating, breeders and scientists should adopt a habit of checking online patent databases such as the database of the European Patent Office and the US Patent Database (www.uspto.gov) for patents and patent applications that may cover information and/or innovations relevant to their area of breeding and research.

It can be quite difficult to search for sequences and combinations of SSRs that

might be covered in patents and patent applications. It is beyond the scope of the non-professional patent searcher to state definitively whether or not a particular sequence is covered by patent rights. Searching patents for specific DNA sequence coverage is not quite as easy as it may seem because of the peculiarity of the language used in drafting patent claims. A few example claims taken from two US Patents, numbers 6 235 972 and 6 815 578 are reproduced in Box 2 to illustrate the complexity of this type of claim language. However, there are companies, such as Gene-IT, that have developed software to search for all possible matches that might occur in any patent (available in electronic form), and where unlicensed use would be considered an infringement. A good patent drafter will attempt to cover as much ground as possible when writing a patent claim as the broader the claim, the larger its technical spread over the landscape of that particular area of science/technology. This results in claims to a sequence and its uses being written so that the inventor claims the sequence and any sequences that are closely similar. Just how broadly a claim is written is a matter of how much the patent drafter/prosecutor can get a patent examiner to accept. Without the assistance of sophisticated computer software, it can be difficult to determine whether the use of a particular genetic sequence would infringe existing patents. Fortunately, however, biotechnology patents are now examined by biologists and molecular geneticists, instead of, as in the "early days", by chemists.

Copyright aspects

Others have thought that copyrights would be of little concern to the breeder or scientist interested in using MAS, in that copyright infringement might only occur if a material such as text, a design, photograph or video was copied and re-used without permission, such as in a publication or video that was to be distributed widely or sold. However, most results of marker testing need to be analysed by a computer program for the breeder to obtain maximum value from such testing. Most software is covered by (at least) copyrights and therefore must be licensed from the rights holder. Even software that is distributed under an "Open Source" type of licence is indeed licensed, and the conditions of the licence must be adhered to when the product is used and/or improved.

In addition, care should be taken by persons creating training materials that will be distributed widely or sold as part of a workshop to either refrain from using materials written and created by others or to obtain permission before use, especially if such use might be part of a course where participants pay for instruction or must buy the training materials, or where materials might be distributed in an electronic format.

Trademarks aspects

In general, the same is true for trademarks as for copyright. A minor point would be to remind authors that terms such as AFLP® and "Breeding by Design™", both trademarks of Keygene, Inc., should carry the "®" or "™" designation. In this regard, breeders would be primarily concerned with the correct use of their own trademarks, both by themselves and others. When naming varieties, etc., care should be taken to ensure that the trademark of another entity is not being infringed. Those responsible for creating names should therefore check public trademark databases such as the UK Trademarks Database (www.patent.gov.uk/tm/dbase/),

and the services of a professional trademark searcher or attorney should be sought before proceeding with the registration of a "new" trademark.

Plant breeders' rights aspects

Breeders using basic MAS protocols with non-proprietary breeding materials (e.g. germplasm that does not qualify as an "Essentially Derived Variety" [Wendt and Izquierdo, 2001]) generally do not need to be concerned with using materials covered by PBRs for breeding purposes.

Contractual aspects

It is very important that licences, contracts and agreements are monitored for restrictions as these often contain provisions dealing with IPRs that last until a contract expires or is renegotiated. Permission to use equipment and associated reagents is normally granted as a licence granted as a part of the purchase price. However, this type of licence may often contain limitations on the use of equipment, reagents and kits for non-research applications. As an example, and as stated in its legal information Web page, Applied Biosystems has an exclusive licence with Roche/Hoffman-La Roche for its PCR patents: "Applied Biosystems is the exclusive licencee of Roche Molecular Systems, Inc., and F. Hoffmann-La Roche, owner of the basic PCR process and reagent patents, for the field of research and development, and for applied fields such as quality assurance and control, environmental testing, food testing, agricultural testing (including plant disease diagnostics), forensics and identity testing in humans (other than parentage testing), and animal identity and breeding applications." This means that when a researcher buys (or has legal access to) and uses an Applied Biosystems machine, the rights to use this machine for certain specified purposes (rarely commercial) are included in the purchase agreement. Note, however, that the use of kits or other products of Applied Biosystems that involve any processes or reagents licensed from Roche/Hoffman-La Roche to carry out MAS is not specifically mentioned as a "field of use" in the terms of this licence. While it could be assumed that use for MAS is possible under the Applied Biosystems licence, if it was considered necessary to have the lowest probable level of risk associated with the use of equipment/reagents for MAS, then legal advice in the jurisdiction of the user should be sought.

An equipment or reagent licence could also contain provisions for what are called "reach-through" rights. These arise when improvements are made to an existing technology. When such innovations come about through use of the existing technology the rights to them may have to go back to the owner of the original existing technology. Such a transfer of sharing of the rights is called "reach-through rights". Some argue, for example, that the requirement in some Open-Source licences for improvements going back to the original creator of the software for distribution are a form of "reach-through".

Agreements to purchase and "package insert" licences should therefore be routinely checked to ensure that these sorts of licence are avoided.

MTAs can also cause problems, depending upon the conditions that are set down in such agreements. Laboratory personnel need to make sure that MTAs are only signed by persons authorized to do so and that efforts are made to check MTA language for provisions that restrict or interfere with the intended use of the germplasm that is produced using MTA-

associated materials. A practical explanation of MTAs is available in COGR (2003).

Breeders and scientists need to keep a file and/or database of all licences, package inserts, purchase agreements and MTAs as part of their routine record keeping. They also need to learn to reject documents that contain provisions that indicate an assertion of rights or include a restriction, to negotiate for terms that they require, or source replacement brands/materials. Contracts can be enforced long after patent rights expire.

Of all the types of IPRs/proprietary restrictions that could affect scientists and breeders in developing countries, licences and agreements have the most potential to impede the use of MAS technologies, unless a sophisticated, high-throughput laboratory is sought. MAS has considerable potential and relevance to developing country breeding systems for capturing desirable characteristics from widely disparate germplasm. IPRs should not hold this back.

REFERENCES

AsiaLaw. 2004. Process patent litigation and the collection of evidence in China. (available at www.asialaw.com/default.asp?Page=20&PUB=68&ISS0=10970&SID=433837).

Balch, O. 2006. Seeds of dispute: it's Argentina v Monsanto in the battle for control over GM soy technology. (available at www.guardian.co.uk/gmdebate/Story/0,,1715331,00.html).

Barendse, W. & Reverter-Gomez, A. 2007. A method for assessing traits selected from *longissimus dorsi* peak force, intramuscular fat, retail beef yield. WO2007012119 (available at www.wipo.int/pctdb/en/wo.jsp?LANGUAGE=ENG&KEY=07/012119&ELEMENT_SET=F).

Binnenbaum, E., Nottenburg, C., Pardey, P.G., Wright, B.D. & Zambrano, P. 2003. South-North trade, intellectual property jurisdictions, and freedom to operate in agricultural research on staple crops. *Economic Development & Cultural Change*, 51(2): 309–335.

Cahill, D.J. & Schmidt, D.H. 2004. Use of marker-assisted selection in a product development breeding program. *Proc. 4th Internat. Crop Sci. Congr.* (available at www.cropscience.org.au.).

CBO (Congressional Budget Office). 1998. How increased competition from generic drugs has affected prices and returns in the pharmaceutical industry. Washington, DC, Congress of the United States (available at www.cbo.gov/showdoc.cfm?index=655&sequence=0).

COGR (Council on Governmental Relations). 2003. Materials transfer in academia. (www.cogr.edu/docs/MTA_final.pdf).

Concibido, V.C., Diers, B.W. & Arelli, P.R. 2004. Review & interpretations: a decade of QTL mapping for cyst nematode resistance in soybean. *Crop Sci.* 44: 1121–1131.

Consuegra, S. & Johnston, I.A. 2006. Polymorphism of the lysyl oxidase gene in relation to muscle collagen cross-link concentration in Atlantic salmon. *Aquaculture Res.* 37(16): 1699–1702.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82: E313–328 (available at http://jas.fass.org/cgi/reprint/82/13_suppl/E313).

Doyle, J.J. & Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11–15.

Drahos, P.A. 2005. Intellectual property rights in the knowledge economy. *In* D. Rooney, G.E. Hearns & A. Ninan, eds. *Handbook on the knowledge economy,* pp. 139–151. Northhampton, MA, USA, Edward Elgar.

Edwards, K., Johnstone, C. & Thompson, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* 19(6): 1349.

Epp, T.D. 2004. Four-wheeling through the soybean fields of intellectual property law: A practitioner's perspective. *Washburn Law J.* 43: 669–679.

Fromer, J.C. 2007. Invigorating the disclosure function of the patent system. *NYU Law School, Public Law Research Paper Series,* 2 March 2007 (available at http://ssrn.com/abstract=967560) .

Fulton, M. & Giannakas, K. 2001. Agricultural biotechnology and industry structure. *AgBioForum*, 4(2): 137–151.

Gallagher, G.J. 2002. Recent development: the Federal circuit and claim construction: resolving the conflict between the claims and the written description. *North Carolina J. Law & Technol.* 4(1): 121–142.

Hore, E. 2004. Patently absurd: evergreening of pharmaceutical patent protection under the Patented Medicines (Notice of Compliance) Regulations of Canada's Patent Act. Canadian Generic Pharmaceut. Assocn. (available at www.canadiangenerics.ca/en/issues/patently_absurd_04.pdf).

Kattan, J. 2002. Evaluating patent infringement and validity in antitrust analysis (available at www.ftc.gov/opp/intellect/020514kattan.pdf#search=%22%22merger%22%2B%22infringement%22%22).

Kellison, R., McCord, S. & Gartland, K.M.A. 2004. *Forest biotechnology in Latin America*. Proc. Forestry Biotech. Workshop, 3–5 March 2004, Concepcíon, Chile (available at www.forestbiotech.org/pdffiles/ChlePDFfinal31Jan2005.pdf).

Kershen, D.L. 2004. Of straying crops and patent rights. *Washburn Law J.* 43: 575–610.

Lee, S., A'Hara, S. & Cottrell, J. 2005. The use of DNA technology to advance the Sitka spruce breeding programme. *Forestry Res. Report* (available at www.forestresearch.gov.uk/pdf/FR_report_2005-6_dnatech.pdf).

Murchie, B.J., Renaud, A.B., Horne, L.E.T. & Hussey, D.T. 2006. Canada's balancing act. (available at www.lexpert.ca/500/lb.php?id=120).

O'Connor, S.M. 2006. The use of MTAs to control commercialization of stem cell diagnostics and therapeutics. *Berkeley Technology Law J.* (available at : http://ssrn.com/abstract=921170).

Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P. & Ganal, M.W. 1998. A microsatellite map of wheat. *Genetics*. 149(4): 2007–2023.

Rothschild, M.F., Plastow, G.S. & Newman, S. 2004. Patenting in animal breeding and genetics. *In* A. Rosati, A. Tewolde & C. Mosconi, eds. *WAAP Book of the Year 2003: a review on developments and research in livestock systems,* pp. 269–278. Wageningen, Netherlands, Academic Publishers (also available at www.psas-web.net/documents/Info/3.5%20Rothschild.pdf).

Rothschild, M.F., Kim, K.S. & Anderson, L.L. 2006. Ghrelin alleles and use of the same for genetically typing animals. *US Patent Number* 7,074,562.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. & Allard, R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Nat. Acad. Sci.* 81: 8014–8018.

Schaeffer, L.R. 2002. Dairy cattle test day models: a case study. *In* M. Rothschild & S. Newman, eds., *Intellectual property rights in animal breeding and genetics*. Wallingford, UK, CABI.

Smyth, S. 2006. Implication and potential impacts from the expiry of patents on herbicide tolerant canola varieties. Saskatchewan Canola Development Commission (available at www.saskcanola.com/pdfs/scdc-patent-report.pdf).

Sreedharan, S.K. 2007. Coming of age – the Indian product patent regime. Intellectual Asset Management Magazine. Special report on "From Innovation to Commercialisation 2007" (available at www.iam-magazine.com/issues/article.ashx?g=e6c6910d-4788-4006-aed5-b9ad1aa69937

Starkoff, D. 2001. Copyright: Law and Practice in a Digital Age. Honour's Thesis, The University of Queensland (available at www.itee.uq.edu.au/~cristina/students/dbs/davidStarkoffHonoursThesis01.pdf).

Wendt, J. & Izquierdo, J. 2001. Biotechnology and development: a balance between IPR protection and benefit-sharing. *Electronic J. Biotech.* (available at http://ejb.ucv.cl/content/vol4/issue3/issues/01/index.html).

CHAPTER 21

Marker-assisted selection as a potential tool for genetic improvement in developing countries: debating the issues

Jonathan Robinson and John Ruane

SUMMARY

Marker-assisted selection (MAS) is a complementary technology, for use in conjunction with more established conventional methods of genetic selection, for plant and animal improvement. It has generated a good deal of expectations, many of which have yet to be realized. Although documentation is limited, the current impact of MAS on products delivered to farmers seems small. While the future possibilities and potential impacts of MAS are considerable, there are also obstacles to its use, particularly in developing countries. Principal among these are issues relating to current high costs of the technology and its appropriateness, given that publicly funded agricultural research in many developing countries is suboptimal and development priorities do not necessarily include genetic improvement programmes. Other potential obstacles to the uptake of MAS in developing countries include limited infrastructure, the absence of conventional selection and breeding programmes, poor private sector involvement and lack of research on specific crops of importance in developing countries. Intellectual property rights may also be an important constraint to development and uptake of MAS in the developing world. It is hoped that through partnerships between developing and developed country institutions and individuals, including public–private sector collaboration, MAS costs can be reduced, resources pooled and shared and capacity developed. With the assistance of the Consultative Group on International Agricultural Research (CGIAR) and international organizations such as FAO, developing countries can benefit more from MAS. These were some of the outcomes of a moderated e-mail conference, entitled "Molecular Marker-Assisted Selection as a Potential Tool for Genetic Improvement of Crops, Forest Trees, Livestock and Fish in Developing Countries", that FAO hosted at the end of 2003. During the four-week conference, 627 people subscribed and 85 messages were posted, about 60 percent coming from people living in developing countries. Most messages (88 percent) came from people working in research centres (national or international) or universities. The remainder came from people working as independent consultants or from farmer organizations, government agencies, non-governmental organizations (NGOs) or United Nations (UN) organizations.

INTRODUCTION

FAO, an intergovernmental organization with 189 Member Nations and one Member Organization, was founded in 1945 with a mandate to raise levels of nutrition and standards of living, to improve agricultural productivity, and to better the condition of rural populations. One of FAO's primary roles is to serve as a knowledge network sharing information on agriculture. It uses the expertise of its staff, agronomists, foresters, fisheries and livestock specialists, nutritionists, social scientists, economists, statisticians and other professionals, to collect, analyse and disseminate data that aid development. The information is made available using a number of different strategies, e.g. providing documents on the FAO Web site that can be freely downloaded, publishing hundreds of newsletters, reports and books, and hosting dozens of electronic fora. In this context, FAO has also been playing an active part in disseminating information and promoting information exchange regarding biotechnology. For example, in 2000 it established the FAO Biotechnology Forum (www.fao.org/biotech/forum.asp), with the aim of providing quality balanced information on agricultural biotechnology in developing countries and making a neutral platform available for people to exchange views and experiences on this subject.

At the end of 2003, the FAO Biotechnology Forum hosted a four-week long e-mail conference entitled "Molecular Marker-Assisted Selection as a Potential Tool for Genetic Improvement of Crops, Forest Trees, Livestock and Fish in Developing Countries". The conference was open to everyone and 627 people subscribed. Each of them received a ten-page document providing easily understandable background information on the conference theme, so that those with little knowledge of the area could understand what the theme was about. The conference was moderated (by John Ruane) and participants were requested to introduce themselves briefly in their first posting to the conference and to limit their messages to 600 words. During the conference, 85 messages were posted, each numbered in chronological order. Of the 627 subscribers, 52 (8 percent) submitted at least one message. Messages were received from each of the different world regions, 28 of the 85 messages (33 percent) were posted from Asia, 26 percent from Europe, 14 percent from Latin America and the Caribbean, 9 percent each from Africa and Oceania and 8 percent from North America. Messages were posted from people living in 26 different countries, the largest numbers from India (25 percent), followed by Australia (9 percent), United States of America (8 percent), United Kingdom (7 percent) and Peru (6 percent), with the remainder from Argentina, Austria, Benin, Brazil, Chile, Cyprus, Egypt, Finland, France, Germany, Ireland, Israel, Kenya, Madagascar, Mexico, Netherlands, Nigeria, Philippines, Spain, Syrian Arab Republic and Turkey. Fifty messages (59 percent) were contributed from people in developing countries and 35 (41 percent) in developed countries. The majority of messages came from people working in research centres (52 percent), including Consultative Group on International Agricultural Research (CGIAR) centres, and in universities (37 percent). The remainder worked as independent consultants or for farmer organizations, government agencies, NGOs or UN organizations.

This chapter summarizes the main issues that were discussed during the conference, based on the messages posted by the

participants. These included some general topics regarding MAS (such as its costs, its actual impact to date on products delivered to farmers, whether it should be a priority in developing countries and whether it was necessary to have an established breeding programme in place before introducing MAS), as well as some MAS-related issues that were more technical, such as which traits are suitable for MAS and the importance of tight marker-gene linkages. Other kinds of issues raised included intellectual property rights, public–private sector linkages, the differences in capacity between developing countries with respect to MAS and the role of the CGIAR and international organizations. Throughout the chapter, specific references to messages posted are provided, giving the participant's surname and message number. All the individual messages are available at www.fao.org/biotech/logs/c10logs.htm.

During the conference, contributions were not evenly spread across the four agricultural sectors of the conference. MAS for crop and livestock genetic improvement dominated the discussions, with issues relating to forest trees and aquaculture mentioned much less, possibly indicating differences in uptake of this relatively new technology among the four sectors. Nonetheless, many of the issues and concerns raised were general in nature and applicable across sectors. These issues included considerations of costs and gains, intellectual property rights and the benefits of partnerships to allow developing countries greater opportunities for developing and using MAS.

Murphy (1) began the conference with a request that MAS be viewed dispassionately as a potential tool for crop improvement to be deployed alongside conventional methods. Sokefun (64) referred to conventional selection methods as "soft" technologies and the newer technologies, such as MAS, as "hard" technologies, and suggested that the hard would not replace the soft technologies and that a fusion of both would achieve the best results. In contrast to more upstream technologies (including genetic modification, mutagenesis and protoplast fusion), which generate additional variation in plant populations, Murphy (1) described MAS as a "downstream technology" that, like conventional phenotypic selection, can be used to select the optimal variants in a population.

The conference discussion was balanced and the topic of the potential of MAS did not evoke a strong reaction among the participants, although many had reservations about it. There was consequently little indication of a substantial dichotomy of opinion whereby participants could be put into pro- and anti-MAS camps. This is in sharp contrast to many debates that have been held about genetic modification. As stated by Muralidharan (7), MAS differs from genetic modification in being more widely acceptable.

There was considerable agreement among the participants on the perceived opportunities and constraints associated with MAS and the usefulness and applicability of the technology in developing countries. Olori (21) thought that successful application of MAS in a well structured breeding programme in any developing country would yield the same benefits as in developed countries. However, as suggested by Montaldo (18) for genetic improvement in animals, it would be necessary to make case-by-case studies, taking into account not only biological issues, but also social, political and economic ones, before making recommendations on application of MAS.

MAIN THEMES DISCUSSED

Whether MAS should be a priority in developing countries

The general opinion was that MAS could be usefully applied for genetic improvement of plants and animals in developing countries, but that it would not necessarily represent a priority. Gianola (6) pointed out that in order for MAS to be taken up in developing countries, because of the scarcity of resources the returns to investment should be far superior compared with those for a developed country, given the significant opportunity costs. Africa was mentioned as facing major constraints to agricultural production, including drought stress, low soil fertility and pests, which were not easily and economically amenable to MAS. Koudandé (68) and Seth (26) stressed the importance of priority-setting in the context of national agricultural economies. Crop diversification and research on underutilized species were also mentioned as other possible priorities for addressing problems of the expanding human population (Priyadarshan, 11 and 71). Murphy (1) suggested that tremendous gains could be made in agricultural development without resorting to applications of biotechnology, by addressing issues of management and infrastructure. For example, in the case of Brazil, a priority might be improvements in the road system to facilitate export crops reaching the ports (Murphy, 1).

Costs of MAS

The cost associated with MAS was a common theme during the conference and several participants, including Collard (9), considered it to be the most important issue for developing countries. It was pointed out (e.g. De Koning, 13) that although costs associated with MAS can be high, conventional genetic improvement programmes can also be expensive. Gianola (2) called for an economic analysis of MAS in comparison with conventional methods, specifically requesting estimates of internal rates of return. He (6) also warned that there was a risk that some investments in MAS might be wasted given the advances being made in post-genomics. For Weller (4), "with respect to the economic questions, MAS is no different from any other technology that increases rates of genetic gain, but also increases costs", concluding that the investments required for MAS could be massive, but so also could the long-term economic gains. However, as pointed out by Montaldo (18), the economics of MAS is based on the value of the selected traits and most importantly, each case should be looked at individually. De Koning (13) highlighted the major economic benefits that could be gained by breeding livestock for resistance to trypanosomiasis.

Various stages in the MAS development and application process were regarded as being costly. Labour and DNA extraction were viewed by Williams (37) as representing the major costs, but Collard (45) considered equipment, consumables and infrastructure to be among the most costly items in a MAS programme. Genotyping (Toro, 67), marker development (El Ouafi, 77; Wallwork, 59) and patenting (Ganunga, 69) were other areas that represented large costs that might constrain MAS use in developing countries. It was suggested that farmers in the developing world could not be expected to pay for MAS (Chávez, 33), while Muralidharan (74) suggested that costs in a country like India would eventually be a lot cheaper than in developed countries.

Participants, including Buijs (58), pointed out that technologies become cheaper as knowledge accumulates and capacity is built up, citing the example of

tissue culture. Buijs (22) also felt that the costs of MAS should be put in perspective with those from other related research areas, pointing out that plant varieties or animals bred by MAS do not require costly safety regulations, in contrast to those bred using genetic modification. Toro (50) and Muralidharan (74) suggested that MAS would become cheaper due to automation/robotics, and Varshney (82) reported that microsatellite marker development has become cheaper as a result of bioinformatics. Many participants suggested that developing countries could make the best use of MAS through collaborative ventures (Olori, 21, 65; Acikgoz, 66; Saravanan, 73), formation of multidisciplinary teams (Sridhar, 76; William, 70; Muchugi, 49) and within national and regional frameworks (Montaldo, 18). Collaboration would spread resources and reduce costs.

Figures for the costs of genotyping mentioned in the conference ranged from US$4 per marker for MAS in pigs (Toro, 79) to under US$0.2 for durum wheat (El Ouafi, 77). Discussion of such exact figures for costs is at best indicative in the face of continuous changes in the world economy, particularly in exchange rates and purchasing power. Suffice to say that as costs are reduced, the value of MAS rises and it possibly becomes more widely applicable.

Putting MAS in context

Although MAS has generated a good deal of expectations, leading in some cases to over-optimism and in others to disappointment because many of the expectations have not yet been realized, participants in the conference aimed to consider MAS rationally and to put it in the context of the whole agricultural picture. As Murphy (1) wrote, MAS "should be viewed dispassionately as a potential tool for crop improvement to be usefully deployed alongside conventional phenotype selection for certain crops and for certain characters".

Good genetic improvement strategies were considered by many to be among the most important prerequisites for successful implementation of MAS. Montaldo (18) said that, with respect to livestock improvement, MAS would not substitute for choosing the right breeding objectives and the starting point of a programme incorporating MAS should be a sound breeding strategy founded on traditional selection methodology. Wallwork (59) thought that many of the criticisms of MAS (e.g. see De Lange, 57) stemmed from poor research and development strategies and not necessarily from shortcomings in the technology. El Ouafi (77) stated plainly that if a successful conventional breeding programme could not be established, MAS would not help, and Olori (21) suggested that the absence of "any real sense of the need for a genetic improvement programme" in developing countries would hinder application of MAS. Such practical strategic considerations balance the hyperbole and over-optimism that has sometimes been associated with MAS. De Lange (57) argued that because of its high costs and relatively moderate results to date, MAS seemed to be "yet another over-hyped gene technology" and questioned, like Ackigoz (66), whether MAS should be a primary consideration for developing countries. Bhatia (8) was among several participants to comment on this issue and believed that the hyperbole to some extent reflected fashion and vendor bias, as for all new technologies.

MAS in relation to conventional breeding programmes

The need for an established breeding programme to be in place for MAS to be

usefully introduced represented one of the main points debated in the conference. Many participants (e.g. Montaldo, 18) explicitly stated the need for a conventional programme to be operational prior to implementation of MAS and others inferred it. Notter (25), on the other hand, suggested that animal recording need not precede implementation of MAS, and he proposed they could be implemented together.

Referring to animal trypanosomosis in Africa, De Koning (13) commented that lack of routine recording of production and health traits, with limited national molecular research facilities, presented a structural problem to implementing a breeding programme using MAS. De Koning (20) also said that when livestock were mainly kept by smallholders, each with a handful of animals, there would be no routine recording. Makkar (52) too suggested that in the low input systems that characterize many developing countries, phenotype and pedigree information were often not available, and this would make it difficult to realize the value of MAS. Notter (25) proposed, however, that MAS (or related technologies) could act as a lever to promote implementation of animal recording. He also noted that "MAS without recording is unlikely to be very beneficial for most traits".

For crops, Singh (61) suggested that MAS should be an integral part of the breeding strategy, but Acikgoz (66) was critical of situations where scientists without any experience of traditional plant breeding programmes entered directly into MAS. Sridhar (76) and El Ouafi (77), while acknowledging the importance of MAS, both suggested that meaningful breeding programmes were necessary to make progress with MAS and Dulieu (23) doubted that traditional selection methods could easily be replaced by MAS. Priyadarshan (11) also believed that more basic biological knowledge about the intricacies of nature was needed to improve selection procedures for plants and Montaldo (18) pointed out that knowledge of genetic control of some important traits remained incomplete.

MAS in aquaculture in developing countries was only briefly discussed in the conference, although Priyadarshan (71) argued that aquaculture merited more emphasis. Martinez (63) suggested that, for aquaculture, application of DNA technologies and MAS was scarce even in developed countries because of the lack of integration between quantitative and molecular genetics, and that the only successful application in aquaculture was that described by Toro (50), who said that molecular markers could be used to assist classical genetic improvement programmes by constructing pedigrees needed for genetic evaluation in trees and fish where pedigree information was otherwise lacking. Martinez (63) noted, however, that economic analysis of this strategy compared with individually identifying fish using electronic devices was scarce. Krause (75) gave an example where molecular marker information could be used to reduce the costs of a fish breeding programme. Normally, electronically tagged back-up copies of nucleus breeding populations of fish are made as an insurance against loss of a deployed population. This is an expensive process that can be avoided by taking tissue samples from sires and dams that are analysed for the presence of established molecular markers if a nucleus stock is destroyed. This allows a nucleus stock to be regenerated relatively easily and cheaply, if and when necessary.

While the merits of applying MAS to genetic improvement of trees in developing countries were appreciated (e.g. Muralidharan, 7), participants suggested

there are many problems that detract from its usefulness. Principal among these is the poor state of current tree breeding in general, and in developing countries in particular. Simons (28) listed a number of problems concerning genetic improvement of tropical trees, including dioecy, undocumented origins and uncertainty of genetic control of traits. However, Galvez (10) mentioned that MAS had been used to assist in selection of coconut parents for breeding. Priyadarshan (11) considered MAS to be helpful for rubber improvement, at least theoretically, and Badr (47) seemed to be looking forward to MAS reducing the time needed for evaluation of fruit trees in Egypt, obviating the need for grafting to see the products of breeding efforts. Forest trees, perhaps more than other genetic resources used by humans, are at, or still very near, their wild state (Muralidharan, 7), which indicates that tremendous improvement can probably be made quite rapidly based on selection among existing genotypes. Muchugi (49) recognized the potential of MAS for tree species improvement, seeing it as a technique best placed to help select and upgrade tropical tree species where the first fruiting may take as long as twenty years.

Technical details of MAS use

There were several contributions to the conference regarding technical aspects of MAS, and how to use MAS effectively in genetic improvement programmes. Mota (14) raised the issues of molecular markers located far from the target gene, increasing the probability of recombination taking place between them, resulting in reduced efficiency of MAS and, secondly, of false positive marker-gene associations. Dulieu (23) also emphasized the importance of tight marker-gene linkage to minimize losses through recombination. Weller (15) acknowledged the importance of both issues raised by Mota (14) and proposed that the best solution to the problem of false positives is to employ the false discovery rate, to get an idea about the expected number of false positives. De Koning (16) supported the use of the false discovery rate and also referred to recent research results suggesting there were benefits in MAS from using a relaxed threshold for QTL (quantitative trait loci) detection. Mota (36) concluded that developing countries should only use MAS in their breeding programmes when there is complete linkage between the marker and the gene of interest, to avoid wasting precious resources. Dulieu (42) commented on this, pointing out the advantages of using flanking markers (i.e. where markers are located on either side of the gene of interest) in MAS.

Singh (44) described the usefulness of MAS in backcrossing programmes, by growing large BC1 populations (BC1 is the first backcross generation), rejecting 50–60 percent based on phenotype (conventional screening) and analysing the remainder with MAS. This could be repeated in the second backcross population, saving considerable time and resources. The usefulness of this approach was confirmed by Dulieu (53), and Sridhar (54) explained how three genes for rice bacterial blight resistance were pyramided into adapted germplasm using MAS in a backcrossing programme.

Which traits for MAS?

Referring to crop improvement, Murphy (1) noted that not all crops and traits were amenable to MAS. A Dutch perspective on the type of traits amenable to MAS to date was provided by De Lange (57), who indicated that single gene controlled traits had received most attention, but little

progress had been made with multiple gene traits. Makkar (52) stated that many MAS studies had adopted a single trait approach, pointing out that with a multitrait breeding objective, response for one trait often goes at the expense of another. He also suggested the utility of MAS when heritability for the trait was low. Singh (41) indicated that "breeders are not much thrilled about MAS for simply inherited traits, and not many QTL (especially the productivity related ones) with tightly linked markers are available".

Several other participants mentioned traits that would be amenable to MAS, including Priyadarshan (11) working with rubber trees, Williams (37) who provided the case of root nematodes and William (70) who mentioned work being done on barley yellow dwarf virus resistance in cereals, rust diseases, nematode resistance and root health. Rakotonjanahary (78) also suggested that MAS be used when conventional approaches to selection were difficult or impossible. For example, Reddy (62) proposed MAS be used for traits where it is difficult to get phenotypic data, like quality traits, and William (70) indicated that protein assays to develop quality protein maize were expensive compared with marker assays. Slaughter traits in livestock were also considered to be amenable to MAS as the desired traits are otherwise difficult to measure without killing the animal (Makkar, 52). Muchugi (49) suggested the potential usefulness of MAS in selecting for medicinal traits and growth rate in tropical trees.

Introgression of genes from wild into cultivated germplasm was proposed to be a good use of MAS (Bhagwat, 46). Notter (25) also commented on the opportunities molecular markers provide for screening populations of animals with favourable or unfavourable genotypes, giving as an example scrapie in sheep. Krause (75) mentioned other genetic examples, such as a sperm defect in pigs and the halothane gene implicated in low pork quality, that could be screened out using MAS. Sex-linked traits were also mentioned as being suitable for MAS (Makkar, 52).

Galvez (10) suggested that molecular markers could be also useful for work with transgenic crops, for characterizing GM plants and tracking movement of the transgene in the gene pool. William (70) also mentioned the use of MAS for transferring a desirable transgene, such as a gene from *Bacillus thuringiensis*, from one cultivar to another.

In addition to discussing traits considered amenable to MAS, brief mention was made of traits not considered amenable to MAS. It was realized that more progress had been made with single genes that were relatively easily transferred, but that there was potential for facilitating QTL transfer, although this was still relatively undeveloped. Traits that are highly influenced by the environment or production system, including crop yield (Priyadarshan, 11), were not considered easily amenable to MAS. Williams (37) pointed out that a major problem associated with MAS was lack of polymorphism at the DNA level, which would render a trait not amenable to MAS. Inadequate coverage of the genetic map with molecular markers was viewed by Dulieu (23) as an obstacle to applying MAS. He also detailed other conditions relating to the nature of the trait that should be considered for MAS to be efficient: single versus multigene, additive versus dominant, expressivity and penetrance.

Practical applications of MAS

Some participants considered the actual impact of MAS on genetic products deliv-

ered to farmers. Although documentation was limited, the current impact seemed small while the future impact was likely to be far more substantial.

Priyadarshan (11) indicated that biotechnology research had been supported actively for over 17 years in India, but was doubtful about the impact on varieties released to farmers. He believed that research on MAS and other biotechnologies had remained largely in journal articles and had not significantly boosted conventional plant breeding efforts on the ground. Kirti (12) lamented that there was no comprehensive documentation regarding the successful use of MAS for breeding new crop varieties or developing breeding material, as this information would be important for evaluating the technology. Collard (45), while noting that MAS had been successful in cereal crops in his country, Australia, said he was not aware of many examples of MAS-derived cultivars grown in Australia despite the wealth of publications from Australian institutions on the technology. Sridhar (48) suggested that, in India, most products of MAS are still in the hands of research institutions undergoing evaluation. He suggested that MAS products require a "fast track" evaluation system to expedite the release of promising germplasm.

According to Makkar (52), success in demonstrating genetic gain in the laboratory did not always equate with success under field conditions. However, some real successes were reported, including transfer of important resistance genes into adapted rice germplasm for Indian farmers (Sridhar, 35 and 54), indicating that more successes might be in the pipeline. Williams (51) said that molecular markers had been used for at least five years in Australia in some wheat and barley improvement programmes and that "it is likely that in Australia all breeding programmes with industry funding and probably also the private breeding companies are currently using MAS to some extent". However, the potential of the new technology has to be weighed against the success achieved using traditional methods. Acikgoz (66) pointed out that the Turkish rice cultivar Tokak was still being sold despite having been released in 1937, and questioned how much impact population genetics studies, popular 20–30 years ago, had on cultivar development, let alone the impact of biotechnology applications.

Buijs (58) mentioned tissue culture, once regarded as a modern, relatively expensive technology, which is now relatively inexpensive and widely used in developing countries. It will only be known retrospectively whether MAS evolves similarly to become a standard tool of the plant and animal breeder in developing countries.

Intellectual property rights issues

Some participants felt that intellectual property rights (IPRs) were an important constraint to development and uptake of MAS in the developing world. Corva (29) raised the issue of the use of licensed genomic technology by public institutions in developing countries, mentioning that many useful cattle markers were becoming available, but which were patented, and that there was therefore a demand for practical information about IPRs and violation of IPRs. Weller (30) pointed out that patents are only valid in the country where they are granted, that research tends to be exempted from patent restrictions and that there can be long delays between filing of patent claims and their eventual granting. Saravanan (31) argued strongly for the freedom of researchers to use patented biotechnology tools. Storlie (32) argued that farmers in the developing world should

be concerned about being constrained by "corporate patents on particular genes, which may require a company's authorization for possession and use". William (70) noted that development of useful markers for MAS was already a significant challenge in developing countries and felt that if their use was restricted due to IPRs "their use would be really limited". Both Williams (51) and Sarla (80) stressed that new genetic information has to be kept as much in the public domain as possible to ensure that there is equal access to it.

Fairbanks (60) described a case demonstrating how some of the limitations imposed by IP issues, including transfer of germplasm across international boundaries, could be overcome, while also avoiding some of the economic obstacles faced by scientists in developing countries. Microsatellite markers for quinoa were being developed at an American university in a joint programme with a Bolivian foundation, to be then sent to Bolivia for use by Bolivian scientists in their quinoa breeding and conservation programmes.

Differences in capacity between developing countries

From the conference it was clear that there is enormous diversity in terms of capacity, opportunities and constraints among developing countries that would have a bearing on development and application of MAS. There are substantial differences in factors including the state of public sector research, the involvement of the private sector in research, development and marketing capabilities, perceived priorities for development, the social and agricultural systems of the country, the state of educational systems and the degree to which information and technology remain in the public domain.

Many participants, including Buijs (22) and Corva (29), commented on developing countries lagging behind developed countries in uptake of new technologies, and Sokefun (3) expressed concern that a lack of resources should not result in the developing world being bypassed. Davila (81) suggested that developing countries like Brazil, where MAS can be used relatively easily, could help other developing countries with MAS development, through south-south cooperation. Roughly a quarter of messages posted in the conference came from India, and it was apparent that this is another developing country that has invested substantially in MAS, among other biotechnologies. Such are the trends in capacity and infrastructure there that it was indicated that Indian institutions might be able to provide MAS services more cheaply than in developed countries (Muralidharan, 74). This is an important consideration, as Bhatia (8) suggested that breeders should ask whether MAS-related analytical work could be outsourced. Reddy (62) believed that MAS would only be economical in developing countries like India.

Role of the CGIAR and international organizations

Collaboration between the developing and developed world was inferred to be the only way for the developing world to participate realistically in the development of MAS and avail itself of the opportunities it represented (Sokefun, 3; Galvez, 38). Fasoula (84) expressed the need for developing countries to play an active role in developing MAS, particularly in making the associations between markers and traits, although Koudandé (68) considered that for economic reasons developing countries could simply import required technology. Many other participants voiced the need for

international cooperation. One demonstration of the extent to which scientists from developing countries are contributing to research on, and application of, MAS is that many participants were from developing countries but studying and/or working abroad. Contributions came from national institutions hosting foreign researchers and also from centres of the CGIAR that promote collaborative research and training. Olori (65) described the many ways that developing country individuals and institutions are contributing to the development of MAS by participating in international agricultural research. Gianola (24), however, questioned the apparent altruism of developed countries in sponsoring collaborative MAS efforts, fearing that it might hide motives for developing biomedical applications from the results.

Partnerships between the CGIAR and national researchers led to some successes in MAS mentioned in the conference. Sridhar (35) reported on the collaboration between an Indian rice research institute and the International Rice Research Institute, and Wallwork (59) on cooperation between an Australian institution, the International Center for Agricultural Research in the Dry Areas and the International Maize and Wheat Improvement Center.

There was a strong call from many participants for the CGIAR and international organizations such as FAO to play an active role in the area of MAS development and application. For example, Murphy (1) suggested that the CGIAR and FAO should facilitate international collaboration in this area, while Priyadarshan (11) suggested that the CGIAR might manage a centralized facility for routinely doing MAS. Acikoz (66) envisaged a role for FAO in addressing issues of classical plant breeding at regional and national levels, which he saw as being more of a priority than MAS, while Muralidharan (74) thought FAO to be suited to playing the role of coordinator for MAS research among laboratories working on the same crop. Rakotonjanahary (78) proposed a similar role for FAO and the CGIAR as facilitators in the exchange of information and genetic material obtained from MAS. Sarla (80) suggested that FAO could play a catalytic role in marker-aided allele mining and facilitate capacity building for applying MAS, especially for crops of regional importance.

Public–private sector linkages

Various additional constraints to using MAS in plant and animal improvement programmes in developing countries were discussed in the conference. Notter (25) stated that the history of public funding in developing countries was not good and Fairbanks (60) commented that agricultural research in developing countries was not well coordinated. Australia has invested heavily in MAS in its breeding programmes but, as pointed out by Collard (45) regarding plant breeding, the major target crops have been cereals produced for export. Moreover, there has been considerable support from private industry for research and development of MAS. For example, the Grains Research and Development Corporation (GRDC) of Australia was set up to serve farmers and is maintained through a levy collected from them. In contrast, in the developing world, most important crops are usually produced for subsistence and there is often little private–public cooperation (Murphy, 1). Developing country farmers are unlikely to be able to support the activities of a dedicated research and development organization equivalent to the GRDC (Collard, 45). Similarly, Notter (25) pointed out that there

was a scarcity of private animal breeding initiatives in developing countries and little or no commercial sector. MAS, in his opinion, would not change this situation. Nicol (19) highlighted the importance of extension agencies in assisting uptake of commercially available DNA marker tests.

Koudandé (68) noted that in developed countries, most of the applied MAS in breeding is undertaken by companies, and wondered which companies in Africa would be wealthy enough to support MAS development and application. An additional factor is that MAS requires that molecular markers are available for particular crops and important traits, but most of the publicly available markers are for the major crops (Collard, 9), which are not necessarily of primary importance in developing countries. Some crops are also very region specific, such as black gram mentioned by Gopalakrishna (72), and are unlikely to be the target of research leading to development of MAS technologies. There seemed to be general support for a collaborative approach to MAS research and application, including public–private sector linkages, which would represent the best opportunity to facilitate development of, and access to, MAS in developing countries. Unfortunately, private sector contributions to this e-mail conference were limited and the discussion would have benefited from inputs by more of them.

REFERENCES

The author, number and title of messages referenced in the chapter – all messages are available at www.fao.org/biotech/logs/c10logs.htm

Acikgoz, N. (66). Is MAS a little luxurious for developing countries?

Badr, A. (47). Tropical fruit breeding

Bhagwat, A. (46). Polyploid crops

Bhatia, C.R. (8). Indicators of utility of MAS in plant breeding

Buijs, J. (22). MAS and other ongoing research

Buijs, J. (58). Locally adapted technology

Chávez, J. (33). MAS and animal breeding

Collard, B. (9). Cost of MAS - the biggest barrier

Collard, B. (45). Investment makes MAS feasible in developed countries

Corva, P. (29). Use of licensed genomic technology

Davila, A. (81). MAS and bioinformatics for developing countries

De Koning, D-J. (13). MAS for livestock in developing countries

De Koning, D-J. (16). Re: Crossing-over // false-positive markers

De Koning, D-J. (20). Re: MAS for livestock in developing countries

De Lange, W. (57). Experiences with MAS so far - Netherlands, plants

Dulieu, H.L. (23). Genetic conditions for MAS efficiency

Dulieu, H.L. (42). Re: When the marker is the gene

Dulieu, H.L. (53). Re: marker assisted backcrossing

El Ouafi, I. (77). Some of the main topics for discussion

Fairbanks, D. (60). Collaborative international research on MAS

Fasoula, D. (84). Re: Integration of molecular markers with plant breeding

Galvez, H. (10). MAS for tree crop improvement and transgenics

Galvez, H. (38). Re: When the marker is the gene

Ganunga, R. (69). Flanking markers/ Patenting

CHAPTER 22

Marker-assisted selection: policy considerations and options for developing countries

James D. Dargie

SUMMARY

Policy options for research, development and diffusion of the products of marker-assisted selection (MAS) depend on the development objectives and priorities of the agricultural sector, its various subsectors and cross-cutting activities dealing with science and technology (S&T), including biotechnology and the management of genetic resources. The policy agenda in each of these areas has shifted from the traditional focus of "raising productivity" to a broader agenda of improving rural livelihoods in both economic and non-economic terms in support of the Millennium Development Goals (MDGs). Securing financial commitments from national governments and donors to invest in MAS and related molecular approaches requires more active engagement by national agricultural research and extension systems (NARES) in the processes of revising poverty reduction strategy papers (PRSPs), and in developing policies and strategies for agricultural development, S&T and genetic resources. Agriculture and agricultural S&T are undergoing rapid change but few developing countries have either agricultural S&T or biotechnology policies. They need to develop these to build coherence across the agricultural sector including delineating the roles of public and private sector entities, and as a means to strengthen accountability with respect to priority setting, monitoring and evaluating the outcomes and impacts of both research and practical applications of MAS. Options are provided for developing and implementing MAS programmes and projects, for setting priorities and evaluating outcomes and impacts. Given the uncertain nature of technical change and the long time frames that often characterize translation of the research and extension services provided by NARES into sustainable improvements in productivity and livelihoods through genetic enhancement, it is concluded that greater emphasis needs to be placed on research to analyse systematically the critical paths involved in successfully transforming research outputs into development outcomes and impacts and that this can best be achieved using an innovation systems approach.

INTRODUCTION

This book provides a comprehensive description and assessment of MAS for increasing the rate of genetic gain in a wide range of agriculturally important species using DNA-based markers for both simple and quantitative traits. Its various chapters attest to the remarkable progress that has been made in researching this approach. This progress has only been possible through the determined pursuit of multidisciplinarity, i.e. by bringing together into teams the skills and knowledge of individuals who could: innovate around the suite of techniques provided by advances in molecular biology to isolate, multiply, identify and insert DNA sequences; produce innovations in electronics and engineering to miniaturize, automate and provide high sample preparation and analytical throughput; use statistical and computer science to analyse and manage the information (bioinformatics) obtained; extend knowledge of the mechanisms that regulate physiological processes in plants and animals; and use quantitative genetics in association with conventional and novel breeding and selection approaches. This research has contributed enormously to the processes of adapting the basic techniques and tools of molecular biology to study the genetic make-up of agriculturally important species at the molecular level, and to accumulating knowledge of the linkages between DNA sequences and in some cases genes and traits that are important for the livelihoods of farmers, foresters and fisherfolk.

Yet, while recognizing this admirable progress, for most species and most traits that are important for both large commercial enterprises in the industrialized world and more particularly for small-scale and resource-poor production systems that constitute the livelihoods of the majority of the rural poor in developing countries, MAS has still to deliver on its undoubted potential, and on the claims in academic and other circles that it would "revolutionize" the way advantageous varieties and breeding stock are developed. As a result, there is still a substantial mismatch between "the field" and the expectations of policy-makers, social scientists, community groups and non-government organizations (NGOs), etc. In fact, the reality is that, while the approach has certainly transformed laboratory operations, apart from its use by the private sector in backcrossing of transgenes into elite inbred lines of maize and other crops and some commercial applications in livestock, the impacts of MAS on rural livelihoods have to date fallen well short of expectations.

This chapter does not dwell on the scientific and technical issues underpinning MAS, nor does it challenge either the need for continued research or the unquestionably greater opportunities for scientific and technical breakthroughs and socio-economic benefits that will surely arise from sequencing and post-genomics research – provided levels of investment in gathering and analysing phenotypic data keep pace with molecular developments. Its focus is on the evolving political, policy and institutional settings (both nationally and internationally) within which agriculture and agricultural S&T institutions and extension services are operating in developing countries, and on some of the options open to governments and public sector institutions in these countries to engage more forcefully in MAS-related R&D and the diffusion of genetically improved products generated through this approach to producers. It argues that the challenges and opportunities for doing so cannot be

divorced from the policies and objectives underpinning country-based and donor-assisted strategies for achieving the targets set by the World Food Summit (WFS), the MDGs, and as described in national PRSPs and in national and regional programmes for food security. Policies and strategies for successful implementation of MAS are also inextricably linked to those for the sector as a whole and its various subsectors, and encompass cross-cutting issues like the management of S&T including modern biotechnology, genetic resources and other developments in the international policy and regulatory arenas that cross lines of national sovereignty. Policy considerations and options for MAS are therefore described within these broader frameworks.

The chapter begins by outlining the social and economic contexts within which the agricultural sector currently operates, the challenges it faces, the main political forces driving change, and both the processes and considerations involved in developing comprehensive agricultural development policies. It then goes on to discuss and provide options available to countries for formulating policies for agricultural research, S&T, biotechnology and genetic resources for food and agriculture (GRFA), arguing that in most countries there is substantial scope for greater "joined up" thinking and coherence of action in formulating, implementing and monitoring the outcomes and impacts of programmes and projects involving MAS. Based on the author's interpretation of the information provided by other contributors to this book, there then follows a section covering some general points that policy- and decision-makers should consider before embarking on MAS, and this is followed by sections dealing respectively with considerations for priority-setting and options for implementing MAS. The chapter concludes by looking at the future of MAS, stressing the need for greater effort in building political support, in setting priorities and better delineating the roles and responsibilities of different stakeholders, in fostering partnerships and in creating more effective delivery mechanisms.

CONTEXT

Hunger, poverty and agriculture

The number of people who go hungry each day in developing countries stands at around 820 million and around 24 percent of the people in developing countries are absolutely poor, living on less than US$1 a day (FAO, 2006). Hunger and poverty in the midst of plenty are the central challenges in today's global economy and society, but if the trends of the past decade are extrapolated forward, there will still be 582 million undernourished people by 2015 (FAO, 2006). This is well short of the target of 412 million that was set at the time of the WFS in 1996, although possibly on track to meet the somewhat less ambitious MDGs that were set in 2000. More than half of the 582 million will be in South Asia and East Asia, with 203 million and 123 million respectively, while sub-Saharan Africa will be home to 179 million hungry. The challenge is not only to provide food security in 2015 for the present 820 million malnourished, but for the additional 600 million people born over the coming nine years and the nine billion people projected to make up the world's population by the middle of this century.

The nature and causes of hunger and poverty are many, varying widely between and within countries; they are also evolving and often interlinked. Even so, the fact that they are most concentrated in rural

areas where people's livelihoods depend on agriculture (including fisheries and forestry) and the non-farm small and medium agro-industrial processing and servicing industries that are connected to it, means that investing in agriculture and more broadly in rural development must be at the heart of any strategy for hunger and poverty reduction. While the measures needed certainly go well beyond the issue of producing more food and agricultural products, achieving greater yields and higher value products from the same plot of land or enterprise must be a key ingredient for the great majority of developing countries.

How to do this at a lower cost to improve household access to food and the competitiveness of small-scale farmers while maintaining or improving producer incentives, the sustainability of farming systems, and the many services provided to societies by both managed and natural land and aquatic ecosystems, poses huge challenges. Particularly challenging is tackling situations where agricultural potential is low, resources poor and markets distant. Any investment in MAS needs to be justified on the basis of its potential to contribute in an effective and efficient manner to these challenges.

The evolving context of agricultural growth and policy

The situation facing rural producers, households and public institutions now is quite different from that of 20 years ago. Political support and consequently public sector investments in agriculture and rural development have fallen both nationally and from international donors and financial institutions. Privatization has been the overarching policy response, but often the private sector has failed to fill the gaps, leaving many producers with no or significantly reduced flows of the inputs and services critical for both production and access to markets such as technologies, extension and credit. Additionally, market access for poor producers has deteriorated due to greater integration of the global economy and other market distortions, and the need to conform to international sanitary, phytosanitary and food safety standards as well as to product accreditation schemes established by supermarket chains and others. Complicating the situation further are pests and diseases, natural resource degradation and climate change.

On the positive side, there is irrefutable evidence of a deepening political commitment within governments and the international community to tackle poverty, hunger and environmental degradation urgently and in a concerted manner. At the global level, this includes the Plan of Action that emerged from the WFS in 1996, the set of eight MDGs that followed the United Nations Millennium Summit in 2000, and the Plan of Implementation from the World Summit on Sustainable Development in 2002. Regionally, it includes the vision and strategic framework document for the New Partnership for Africa's Development (NEPAD) and its underpinning Comprehensive Africa Agriculture Development Programme (CAADP) (NEPAD, 2002).

Nationally, the most notable examples are: the development of revised country-driven PRSPs, which aim to link national public actions, donor support and the results needed to support the MDGs, and which provide the basis for World Bank and International Monetary Fund's concessional lending and for debt relief under the heavily indebted poor countries (HIPC) initiative; national development strategies, plans

and programmes; sector-wide approaches (SWaps) that are aligned with the MDGs; and national and regional programmes for food security that are supported by FAO and its donors. In one form or another, these documents describe national macroeconomic, structural and social policies and programmes to promote growth and reduce poverty, and map out plans for their attainment and priorities for both domestic and external assistance. With respect to the latter, the 2005 Paris Declaration on Aid Effectiveness provides a multidonor commitment to improve aid effectiveness through harmonization, alignment and managing for results (see, for example, www.oecd.org/document/18/0,2340,en_2649_3236398_35401554_1_1_1_1,00.html).

Significant within essentially all of these processes is the shift from the traditional agenda of "raising productivity" for agriculture to the broader agenda of improving rural livelihoods in both economic and non-economic terms. Also significant is the fact that despite their positive track records, agriculture and, even more so, agricultural research are not high on the list of priorities in country PRSPs. This points clearly to the need for agricultural ministries and NARES to engage more actively in the process of revising future PRSPs. However, success in elevating the priority given to agriculture within PRSPs will only be achieved by formulating and delivering agricultural R&D policies, programmes and activities that are coherent within the agricultural sector, with national PRSPs and with the MDGs. Noteworthy here is that in 2004[1] the Economic and Social Council of the United Nations (ECOSOC) underscored that most developing countries are unlikely to meet the MDGs without a clear political commitment to making S&T among the top priorities in their development agendas, and recommended that governments increase R&D expenditure to at least one percent of gross domestic product (GDP).

COMPREHENSIVE AGRICULTURAL DEVELOPMENT POLICIES

Defining these policies basically entails determining the broad-based objectives of the sector,[2] which, in the new context of development, also needs to include goals for enhancing social equity and natural resource sustainability. Essential to the process is "mapping the terrain". This requires evidence-based stocktaking of past trends within each subsector (e.g. with respect to production, marketing and legislation), and identifying barriers to realizing opportunities for expansion, strategic alternatives for moving forward based on an assessment of what the future is likely to hold, and the instruments and means for their implementation (e.g. through new legislation, administrative decrees, public and/or donor investment, and participation by the private sector and civil society).

Preparation then moves on to develop an integrated sector-wide package of policies to guide implementation, including an investment programme. Policies directed towards rural poverty reduction through agriculture must be based on (a) determining where poor people intersect most prominently with agriculture and the major risks they face (e.g. drought, salinity, disease outbreaks); (b) the types of production systems and commodities they produce;

[1] UN Economic and Social Council (ECOSOC) Resolution 2004/68 "Science and Technology for Development" (E/2004/INF/2/Add.3).

[2] In many countries, forestry and fisheries are separate sectors with ministries responsible for developing National Forestry or Fisheries Action or Development Plans

and (c) the linkages they have to markets, research and extension systems, etc. Some approaches for obtaining this information are described later.

From the standpoint of planning investments and strategies, it should set the scene for how the government intends to pursue reductions in poverty and food insecurity through agriculture, answering questions such as: will greater emphasis be placed on self-sufficiency and, if so, for what commodities; to what extent does the government intend to promote production by sourcing seeds and planting materials, fertilizers, breeding and broodstock, and feeds for livestock, fish and shellfish from abroad or relying on its own genetic resources and research and dissemination systems; does it foresee greater private sector involvement and, if so, in what areas and how will this be achieved; what is its attitude towards modern biotechnology – does it intend to pursue this and, if so for what purposes and how? For example, for internal political, external trade, cost and technical skill considerations, do investments in MAS appear more attractive than pursuing the development and/or importation of genetically modified organisms (GMOs),[3] and should investments in MAS and/or GMOs be given priority over conventional genetic selection approaches?

The value of having an agricultural development policy in place lies not only in the end result, i.e. a description of the course of action that a country intends to take to move the sector and its various subsectors forward over a given time frame. It also comes from the process itself, which if done with commitment to detail and rigour in analysing both past trends and future scenarios for the sector, and inclusiveness and transparency to ensure the broadest possible stakeholder participation and buy-in, leads to policies and strategies that are coherent within and between subsectors and between agriculture and rural development. It also has better prospects of securing consensus within the sector and endorsement for its implementation from other ministries with a stake in rural development.

There is clearly no ready-made model for conducting this process or for how the policy itself is implemented, monitored and the lessons learned are fed back for updating, but sound leadership and commitment are critical for preparing relevant and objective inputs through analysis and synthesis of information available within the ministry itself and from other relevant ministries. Ideally, information is also provided by research, extension, higher education and other service institutions and bodies in the public and private sectors including from civil society through documentation and/or organizing meetings at central, local and even community levels, and by outside advisers. In other words, both "top-down" and "bottom-up" approaches are essential to achieve balance and consensus with respect to goals and objectives.

Agreeing and implementing policies and strategies that are mutually supportive and where the "sum" is greater than the "parts", to meet the needs of the huge diversity of farming systems that exist in developing countries while targeting poor producers and consumers is clearly a huge challenge requiring negotiation, compromise and realism. It is also something that can neither be rushed nor "set in stone", and a fully integrated "rolling" policy that is updated at regular intervals represents the ideal. Developing an agricultural policy

[3] Referred to as LMOs (living modified organisms) in the Cartagena Protocol on Biosafety (2000) to the UN Convention on Biological Diversity (1992).

reform package in Honduras, for example, involved around 80 meetings over a year in which both Campesino and large-scale producer organizations participated, while in Guyana more than 100 meetings of civil society task forces were held over several years (www.fao.org/docs/up/easypol//354/agrc_pl_str_cnpt_prct_031EN.pdf).

POLICIES FOR AGRICULTURAL RESEARCH AND EXTENSION

The absence of systematic and comparative data on the benefits arising from the use of different technologies in agriculture precludes attaching priority to any one approach. However, the high rates of return and the reductions in both economic (Pardey and Beintema, 2001; Evanson and Gollin, 2003; Raizer, 2003) and non-economic poverty (Meinzen-Dick *et al.*, 2004) are impressive by any standards, and justify many of the past investments made in research and technology transfer. Consequently, it is hardly surprising that mobilizing and directing national institutions and skills towards capitalizing more forcefully on the opportunities available through S&T to meet the MDGs and other global, regional and national targets are consistently stated commitments within different sectors and themes. Most noteworthy in this regard are the plan to achieve the MDGs prepared by the UN Millennium Project and the underlying report of its Task Forces on Science, Technology and Innovation (UN Millennium Project, 2005a and 2005b), Africa's S&T consolidated plan of action (NEPAD, 2005), and the CAADP and InterAcademy Council reports on realizing the promise and potential of agriculture in Africa (NEPAD, 2002; IAC, 2004).

Yet, here again, judged by the content of current PRSPs and agricultural development plans, the vision of how S&T can contribute to enhancing the value of GRFA and thereby to achieving national economic and social objectives is invariably either missing, or the belief is projected that "on the shelf" technologies and existing knowledge only have to be adapted to local circumstances to meet the challenges ahead.

Unquestionably, policies and strategies promoting adaptive research and dissemination of existing technologies with a successful track record must have highest priority in the short term. However, priorities that will respond to the needs of small producers and rural households for new technologies in 10 or 20 years time (and requiring more upstream strategic and applied research) also need to be identified now because (and despite the claims of some scientists), even with the availability of more advanced R&D methods and tools, there is no reason to believe that the uptake of new agricultural technologies will continue to be other than slow and incremental (described by Pardey and Beintema, 2001, as "slow magic"). In this regard, the importance of countries having pluralistic, participatory, client-focused, decentralized and gender-sensitive advice on the processes of technology diffusion and adoption should be emphasized.

Maximizing the relevance and future contributions of new technologies to overall agricultural development also requires greater attention to planning and decision-making about the direction and management of the scientific techniques and tools as well as the genetic resources to which they will be applied in food and agriculture against the backdrop of current and likely future driving forces of change.

These forces include:

- *political policies*, where, as outlined previously, the new framework for

agricultural policy focuses on alleviation of rural poverty. By shifting future investments in research towards an emerging paradigm of "research for development", the research agenda is broadened to ensure functional linkages to national development policies and to include the wider dimensions of livelihood strategies in both planning and assessing the impact of projects and programmes. It also shifts past emphasis on input-based technology supply by scientists to demand and need-driven innovation systems involving many other actors. Political commitment is also crucial to ensure sustainability of funding;

- *advances in science, and most notably in the computing and biological sciences*, as these have provided new techniques and tools for researchers to locate better and therefore target production systems and communities most associated with poverty and food insecurity, and new technologies in the form of seeds, breeding stock, vaccines, etc. with the potential to increase productivity within agricultural systems and wider food chains and improve economic and social well-being. They are also helping to overcome barriers to wider social engagement in decision-making;
- *growing acceptance of the importance of optimizing system productivity* by better-managing the interactions among diversified farm enterprises and sustainable resource management and ensuring access to markets, rather than maximizing individual crop or animal performance;
- *expanded intellectual property rights (IPRs) for biological innovations* (see Chapter 20) and changed norms for accessing and sharing the benefits of genetic resources in general and plant genetic resources in particular, supported by international agreements, conventions and treaties (see later);
- *increased private investment in S&T in general*, and within agriculture, through both R&D directed primarily towards crop, livestock and fish genetic improvement, and the delivery of products through multinational and national seed and breeding companies and their franchises;
- *expanded public–private sector collaboration* in research, development and extension, in some countries supported by legislation;
- *increased public awareness* of the relevance of the uptake of new technologies and their significance for improving the livelihoods of rural people.

These changes have already been felt most forcefully in industrialized and large developing countries such as Brazil, China, India and South Africa, where the demand-pull created for products of R&D is greatest. However, their impact is increasingly spilling over into others, including the low-income food-deficit countries with much less capacity to benefit from or otherwise adjust to the new realities of conducting S&T in a globalized world. An important issue for all countries is therefore how to adapt their NARES to respond better to both the current and likely future needs of their agricultural sectors, and in so doing to consider their S&T "futures", one of which is clearly modern biotechnology. Nonetheless, judging by the content of both PRSPs and national agricultural development policies and strategies, few developing countries appear to have started along this road by producing an integrated agricultural research and extension policy. This situation is hardly conducive to obtaining political and financial support for R&D on approaches such as MAS, which, as

is clear from other chapters in this book, still remains largely in the laboratories and experimental stations of research institutes in both industrialized and developing countries.

NATIONAL S&T AND BIOTECHNOLOGY POLICIES

Like the comprehensive agriculture development policy, the rationale for having a national policy on S&T and, within that or separately, on biotechnology, is to provide a framework for government and key stakeholders to work together in a coherent and mutually supportive way to ensure that developments are captured for national benefit. The principles involved and mechanics of how it can be developed and managed are essentially the same as those described earlier for agricultural development policy planning, the main difference being in the breadth of government involvement – being cross-cutting issues, developing and implementing national S&T and modern biotechnology policies are clearly cross-sectoral responsibilities with coordination normally assigned to the Ministry of Science and Technology. The examples given in Box 1 illustrate options for pursuing the development of a biotechnology policy. More information on national biotechnology policies in individual countries is available at www.fao.org/biotech/country.asp.

While many countries have an overall S&T policy in place, and these and some other developing countries now have biotechnology policies (the most recent examples being Bangladesh, Kenya, Malaysia and Nigeria), the vast majority do not. Most national agricultural R&D institutions therefore lack the compass provided by the process of developing an overall national policy to guide the development and management of an agricultural biotechnology policy and from there, to formulate programmes and projects specific to the agricultural sector. This paralysis in policy-making only serves to promote supply-driven, at the expense of demand-driven, priority-setting and hence targeting of investments towards questionable needs. It also leads to fragmented and uncoordinated activities, and in some cases to delays in the adoption of technologies that could help improve the efficiency of agricultural research and provide products and services that directly or indirectly improve livelihoods. Indeed, the survey conducted by FAO on applications of MAS in the crop subsector (Chapter 2) illustrates well both the dearth of skills in priority-setting and coordination within many countries that adopt this approach and the complete lack of such activities in many others. While this can be explained to some extent by the relative novelty of biotechnology applications, as far as MAS is concerned, the paucity of information on the actual or potential economic and social benefits of the products arising from its application in the different agricultural subsectors is surely a major stumbling block to priority-setting and investment.

National agricultural research and biotechnology policies

Even in the absence of an overall national biotechnology policy, countries have a number of options for improving the strategic planning, monitoring and evaluation of modern biotechnology applications, including MAS, within their agricultural and wider rural development sectors.

The preferred approach is for the Ministry of Agriculture in association with other relevant ministries (particularly Higher Education) to champion the process

BOX 1

National biotechnology policies: Thailand and South Africa

In Thailand, the National Biotechnology Policy Framework (2004-2009) was prepared by a National Biotechnology Policy Committee chaired by the Prime Minister. This then led to the setting up of a national centre specifically devoted to biotechnology (the National Centre for Genetic Engineering and Biotechnology) under the National Science and Technology Development Agency. It is both a granting and research agency with its own research laboratories and is funded from a combination of government finances, revenue from services and commercial projects and competitive grants from national and international sources. It has major activities in agricultural biotechnology including: on genome mapping and marker-assisted breeding of rice; on cassava improvement where a database of cassava expressed sequenced tags (ESTs) is currently being developed and employed in the study of starch biosynthesis; and shrimp, with major projects on ESTs and genome studies for application in breeding, disease diagnostics and shrimp domestication. Noteworthy also is that through joint government-private sector funding, Thailand will host "Biotechnology Asia 007" with the focus firmly on agriculture.

In the case of South Africa, the National Biotechnology Strategy (2001) arose from a government request and the work of an interdepartmental committee led by the Department of Arts, Culture, Science and Technology with participation of the Departments of Agriculture, Health, Trade and Industry, and Environmental Affairs and Tourism. This committee set up an Expert Panel to provide specific inputs based again on "mapping the terrain" in terms of current applications, legislation and finance, and participation by all key stakeholder groups, etc. Arising from the policies proposed within the strategy document, a National Biotechnology Advisory Committee was established in 2006 and the Department of Science and Technology created a Biotechnology Unit. Since then, three Biotechnology Research and Innovation Centres and a National Bio-informatics Network have been established, interdepartmental cooperation has been promoted, and bilateral agreements have been signed. Again, agricultural applications of biotechnology receive high priority in this national strategy.

of establishing an open learning process for a national agricultural S&T policy dialogue including biotechnology, leading eventually to a planning document and a process of monitoring and evaluating outcomes and impacts. This could be achieved by establishing a national committee that would then define terms of reference and set up various task forces/working groups in a participatory and pluralistic manner to report on specific subsector and thematic issues.

In common with other planning procedures, the first step should involve a diagnostic study and analysis of existing S&T policies as well as of the national, regional and international S&T landscape. Bijker (2007) provides an excellent description of the criteria for building an S&T policy via a policy dialogue and a methodology for carrying out a diagnostic study. Essential for promoting a well thought-out policy and its effective management is the closest possible involvement of all

players with a stake in research, development and diffusion of genetic material (ministry personnel, representatives of NARES, private companies, NGOs, farmer's groups, etc.). However, its focus must be on developing new national agricultural S&T (including biotechnology) policies and strategies to: (a) support institutional reforms, including intensifying cooperation at national, regional and international levels; (b) strengthen national capacities; and (c) identify new funding mechanisms.

Within this process, countries need to identify priorities and appropriate levels of resources to assign to biotechnology in light of their socio-economic conditions and cultural contexts and, in situations of no-growth budgets, they need to decide on what is to be diverted from other important problems. A critical issue is also reaching agreement on the roles and responsibilities of public and private sector entities. Suggestions are given later for considering MAS within this overall context.

An alternative – or preferably as part of the process of preparing a national agricultural S&T policy – is to draw up specific subsector strategies. This has the advantage of focusing minds and resources within that subsector in a holistic manner, for example, in the case of the crop subsector, by bringing together stakeholders dealing with breeding, conservation and seed production/dissemination. Another possibility, very attractive from both the S&T angle and for avoiding the creation of new structures, is to cover modern biotechnology policy development and programming through existing structures for managing genetic resources (see below for more details on rationale). Least attractive and cost-effective, but unfortunately all too often the case, is for individual research institutions and universities to draw up and implement policies and programmes that lack coordination with others dealing with the same or closely related subject matter, in particular genetic resources management.

Whatever the path chosen and notwithstanding the need to ensure national ownership of the process, advice (if needed and requested) on the actual or potential role of MAS within the agricultural S&T landscape should be sought from independent sources. These include the Consultative Group on International Agricultural Research (CGIAR), FAO, the World Bank, the International Council for Science (ICSU), the InterAcademy Council (IAC), the InterAcademy Panel (IAP), regional academies such as the Federation of Asian Scientific Academies and Societies, and national academies.

National policies on genetic resources in food and agriculture

MAS needs access to DNA-based techniques, constructs, tools, databases, statistical packages, etc., and Chapter 20 describes the IPRs surrounding these which include patents, copyrights, trademarks etc., as well as providing suggestions to NARES about acquiring these technological resources. However, successful application of MAS also depends on accessing new breeding techniques and, as many national collections may lack sufficient diversity (e.g. to reduce vulnerability to pests and diseases), they may need to acquire genetic resources that are available in other countries within landraces, wild ancestors and relatives, parental and breeding lines, protected varieties, breeds and broodstock. Additionally, as knowledge grows of the linkages between phenotype and genotype, awareness increases of the potential value of genetic resources and, as participatory processes involving local communities

become more prevalent, so the demands for both germplasm exchange and sharing the benefits of the final products that are generated from R&D will increase. In fact, over the last 30 years, and due to a combination of the new possibilities opened up by molecular biology and some well publicized cases of "biopiracy", governments have increasingly come to appreciate the actual and potential value of genetic resources. This has resulted in an expansion of legally-binding global and regional instruments, and national laws, regulations and policy concerning issues of access, ownership and control of genetic resources and the sharing of benefits arising from their use or enhancement.

For the further pursuit and future success of MAS, policy- and decision-makers as well as individual scientists need to be aware of the requirements for international exchanges of genetic resources such as those described in the CBD (1992), the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (see Stannard *et al.,* 2004 and Bragdon, 2004) and its standard Material Transfer Agreement (MTA), and the World Trade Organization (WTO) Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs), particularly Article 27.3 (b), which states that while members may exclude plants and animals from patentability, if they choose to do so in the case of plants, they must provide an effective "*sui generis*" system of protection such as the 1978 and 1991 versions of the International Convention for the Protection of New Varieties of Plants administered by the International Union for Protection of Plant Varieties (UPOV), or a combination of the two (IPGRI, 1999; Le Buanec, 2004; Donnenworth, Grace, and Smith, 2004; FAO, 2005a; Tripp, Eaton and Louwaars, 2006).

They should also be aware that international exchange of germplasm carries with it the risk of introducing diseases and pathogens through plants and animals and their parts such as seeds and propagules, semen and embryos, and that sanitary and phytosanitary certificates are required to facilitate the safe exchange of genetic resources between, and under some circumstances, within countries. Familiarity is therefore needed with the WTO Agreement on the Application of Sanitary and Phytosanitary Measures (SPS), the WTO Agreement on Technical Barriers to Trade (TBT) and the instruments relevant to standard setting within these, including: the International Plant Protection Convention with its objective of preventing the introduction and spread of plant and plant product pests, and the Animal Health Code implemented by the World Organisation for Animal Health (OIE) that covers both livestock and fish.

This international policy and regulatory framework is both complex and continuously evolving. Hence, apart from the scientific and technical challenges involved in MAS, developing countries face formidable difficulties in crafting and implementing legal and regulatory frameworks that facilitate exchange of GRFA as well as the range of tools used in MAS for both research and commercial uses. It also challenges national policy-makers to keep abreast of the international policy-making processes and all that these imply in terms of both coordination between ministries of agriculture, trade, environment, and S&T, and in human and financial resources. However, the consequences of not being knowledgeable about these matters, and in particular about the appropriate national laws of countries from which genetic resources and scientific techniques

and tools are sought, could be serious for individual researchers and their institutions. All of these aspects must therefore be managed in a well coordinated, efficient and fair manner if countries are to realize fully the potential offered by MAS to contribute to national food security and agricultural development.

Most developing countries have activities dealing with specific aspects of plant genetic resources for food and agriculture (PGRFA) and substantial numbers have established cross-institutional programmes and coordination mechanisms including crop-specific bodies and networks to set priorities, evaluate progress and in general to promote the more effective use of genetic resources. In the case of livestock, and as noted in the draft report on the State of the World's Animal Genetic Resources for Food and Agriculture, presented at the Fourth Session of the Intergovernmental Technical Working Group on Animal Genetic Resources for Food and Agriculture, held at FAO headquarters in December 2006 (www.fao.org/ag/againfo/programmes/en/genetics/documents/AH473e00.pdf), apart from a few (mainly northern hemisphere) countries with well-developed commercial livestock enterprises, interest in the topic is often limited to isolated departments within institutes that are rarely involved in animal genetic resources-related activities. In the case of forestry, national programmes have been established with consultative fora and lead institutions often outside the Ministry of Agriculture, and essentially similar arrangements operate with fisheries. India, for example, has a National Bureau within the Indian Council for Agricultural Research (ICAR) devoted to fish genetic resources, which undertakes work with microsatellite markers for population genetic analysis and determining genetic variation among and within inland species.

With respect to national biodiversity programmes, by being Parties to the CBD, countries have committed themselves to establishing policies, legal and regulatory frameworks and programmes for conserving and using both wild and non-wild biodiversity in a sustainable manner and, in doing so, to establish national Biodiversity Strategies and Action Plans (BSAPs) through focal points invariably coordinated by Ministries of Environment.

At present, however, few countries emphasize the conservation and sustainable use of GRFA in their PRSPs and agricultural development plans, or have strategic and holistic roadmaps as to how these resources should be managed within a particular subsector, let alone how this will be accomplished across subsectors. Such deficiencies are not conducive to priority setting and reaching agreement on specific programmes and projects, and therefore jeopardize funding of the most critical needs and the attainment of national objectives. Since genetic resources are the raw materials to which molecular methods and approaches such as MAS are applied, these deficiencies inevitably also lead to ineffective and inefficient integration of modern biotechnology into national programmes.

To put management of GRFA on a better footing, including through the appropriate integration of MAS and related approaches, countries need to establish or strengthen existing organizational structures and programmes that respond to national development objectives. This means ensuring that: they are well linked to the wider policies and programmes drawn up for agriculture, biodiversity and biotechnology; that they take into account the perspectives both of public institutions

dealing with research, genebank operations and the supply of seeds or breeding stock, and those of wider stakeholder groups such as farmer and community groups, private sector entities, breed societies, etc.; and they recognize the interdependencies between national, regional and global policies and laws concerning access to these resources and sharing the benefits from their use.

It is beyond the scope of this chapter to deal with the setting up and coordination of systems or programmes for the management of genetic resources at the species, subsector, sector and wider levels. These aspects are covered comprehensively for PGRFA by Spillane *et al.* (1999), and the principles involved are equally relevant to the livestock, forestry and fisheries subsectors. However, while recognizing the essentiality of having a national biodiversity system or programme that is overseen, for example, by a high-level interministerial coordinating body for pursuing national development objectives and reporting through the CBD process, there is simply no substitute for specialized genetic resources knowledge within each of the agricultural subsectors to promote effective and efficient planning and implementation of MAS, including through awareness building and advocacy within national and international policy forums and interactions with donors.

MAS: GENERAL CONSIDERATIONS FOR POLICY- AND DECISION-MAKERS

One of the main take-home messages from the experts contributing to this book is that MAS can be demanding in its requirements for specialized equipment, consumables, electricity supplies, laboratory design and management, data handling and statistics, and Internet connectivity. Another is that MAS is a complement to and not a substitute for skills in conventional breeding and selection. Embarking on MAS should therefore never be considered as a paradigm replacing classical crossbreeding and phenotypic screening programmes, which in many developing countries are in any case limited in terms of species coverage and the availability of, for example, temperature and humidity-controlled greenhouses and growth chambers and field sites, and fragile in terms of staffing and funding levels (see, for example, Chapter 8).

Yet another message is that efficient and effective application of MAS requires well-qualified staff. First and foremost, it needs staff who have the knowledge to lead decision-making on when and when not to embark on MAS. This has to be done strictly on a case-by-case basis, bearing in mind that MAS may accelerate genetic progress in some traits better than others, and that the costs and benefits of using MAS in a production system need to be weighed up in the same way as any other input. It also needs leaders who give the "end product" rather than the "laboratory/research process" the main consideration, and staff with substantial design, technical, analytical and problem-solving skills and who are up to date with developments in the field. Furthermore, it demands a sustainable funding base. What should never be forgotten is the bottom line – namely, the investment made will ultimately be judged on the number of people benefiting from planting improved plant germplasm or keeping improved farm animals or fish.

Another key message is the absolute necessity of ensuring effective coordination between breeders and the people working in molecular biology laboratories. While it is not essential for all of these to be located physically within the same institution, policy- and decision-makers need to

know that investments in staff and infrastructure for the "molecular component" of MAS are wasted if they are not linked to the "breeding and selection" components.

Apart from countries with technologically advanced NARES (Type 1 and Type 2 described by Byerlee and Fischer, 2001), getting all of the above elements together is a big task for the vast majority of developing countries, particularly against the background of current and often deteriorating levels of public funding for agricultural R&D (Pardey *et al*., 2006). So big indeed that, while recognizing the need/opportunities for molecular MAS, they may consider, in the first instance, other very valuable applications of molecular-based techniques such as the polymerase chain reaction (PCR) for plant, livestock and fish disease diagnosis (see, for example, Viljoen, Nel and Crowther, 2005), estimating genetic distances between varieties, strains, lines and breeds, conducting variety and parentage testing (De Vicente, 2004; Chapters 14 and 17) and for GMO characterization and detection. These applications are not considered further here since they fall outside the core subject matter of this book. Also, while recognizing the increasing role of the private sector, the options described below for pursuing MAS are based on the assumption that the public sector will continue to be the major investor in R&D for small-scale producers and increasing the access of poorer sections of society to affordable food and agricultural products. Additional options are available through public–private partnerships and these are discussed later.

A final consideration is that, unlike the development and release of GMOs, MAS does not require the establishment and the enforcement of a specific legislative framework. Apart from avoiding the need for specific capacities in public administration, this certainly reduces the final costs of adopting MAS-derived varieties and breeds.

PRIORITY SETTING FOR MAS

Targeting the farming systems, species and traits linked most to poverty and hunger

Investing in MAS has to be based on striking an appropriate balance between needs and opportunities for combating hunger and poverty through genetic enhancement. Essential to that process is determining where the greatest concentration of poverty and hunger exists and the causal factors. There are essentially four approaches for pursuing this.

- *Poverty and hunger mapping*

Although still relatively new, this approach is gaining increasing acceptance in national and international development circles. One of the major challenges faced by all countries in targeting their development, and hence research efforts, towards the food insecure and poor lies in the diversity of their farming systems and socio-economic conditions. However, using a combination of survey and census information (e.g. household surveys), administrative data (e.g. markets, roads), geographical information systems (GIS) and small area estimation maps, it is becoming increasingly possible to develop correlations and maps that link population densities, welfare data and crop and livestock production and livelihood systems; in effect, to pinpoint where poor people live and the production and livelihood systems associated with high levels of poverty.

Increasingly, through programmes such as the Inter-Agency Programme on Food Insecurity and Vulnerability Information

and Mapping Systems (FIVIMS) which works both locally and internationally, Ministries of Planning are developing disaggregated poverty maps to help target their interventions for greatest benefits to the poor. Recent examples include the high resolution Kenyan poverty maps developed by the Bureau of Statistics within the Ministry of Planning with the assistance of the International Livestock Research Institute (ILRI), the Rockefeller Foundation, the World Bank and the World Resources Institute (WRI) (Ndeng'e *et al.,* 2003), and the International Rice Research Institute's work linking poverty and rice systems in Bangladesh (www.irri.org/science/progsum/pdfs/DGReport2000/FP1.pdf).

- ***Rapid rural appraisals***

These are systematic but semi-structured activities conducted by teams with both technical and social science backgrounds, usually as part of farming systems research (see below and Crawford, 1997). Their chief characteristics are that they take only a short time to complete, tend to be relatively cheap to carry out and make use of more "informal" data collection procedures. The techniques rely primarily on expert observation coupled with semi-structured interviewing of farmers, local leaders and officials. In substance, the techniques of rapid rural appraisals (RRA) are executed over a period of weeks, or at most months, rather than extending over several years. To date, RRA has mainly been used in the field of rural development as a short cut method to be employed at the feasibility stage of project planning.

- ***The farming systems approach***

This groups farm households with similar characteristics and constraints and therefore from a R&D perspective has the potential of promoting technology and knowledge spillovers. Unquestionably, the most authoritative study of the link between farming systems and poverty is provided by Dixon, Gulliver and Gibbon (2001). These authors describe 72 major farming systems throughout the developing world based on available natural resources, patterns of farm activities and household livelihoods, intensity of production and their relationship to markets. They also describe the needs of those living within them (with an average agricultural population of about 40 million inhabitants), the likely challenges they face and opportunities open to them in the next 30 years, and the relative importance of different strategies for escaping from poverty and hunger. In sub-Saharan Africa for example, of the 15 major farming systems identified, both they and the IAC (2004) gave priority to four systems based on the economic value of production and the extent of malnutrition, namely: the maize-mixed; the tree-crop based; the cereal/root crop based; and irrigated systems. However, NARES need to undertake similar priority assessments to complement such analyses.

- ***The "rural worlds" concept***

This categorizes rural people as capital intensive farmers, mixed commercial/subsistence farmers, the near or totally landless and those without any economic activity (OECD, 2006).

While each of these approaches has merits and limitations for targeting interventions based on geography and population, they all embrace the principle of engaging farmers and rural consumers/households in diagnosing problems and identifying possible solutions adapted to their particular circumstances.

Analysis of the needs and opportunities for MAS

As noted earlier and in an ideal world, the needs and opportunities for embarking on MAS should emerge first and foremost through the policy dialogue processes that lead to country priorities and objectives for agriculture and for agricultural S&T. In any case, careful analysis is needed to determine whether, given the current S&T, socio-economic and cultural landscape and government/community plans for the foreseeable future, the use of MAS will realistically contribute to hunger and poverty reduction. This requires a team of competent analysts to conduct an *ex ante* impact assessment that makes the best use of existing knowledge to determine whether:

- the principal barrier to sustainable intensification or diversification of the production system(s) as a whole could be overcome by introducing a new plant or animal genotype or by changing the environment, e.g. introducing better soil, water and nutrient management practices, draught power, vaccination, tsetse fly or other disease/integrated pest management practices. Also to be considered are the management changes that would inevitably be needed following the introduction of such genotypes. For example, increasing the prolificacy of local sheep or goats through MAS brings with it the requirement, *inter alia*, for an improved feed resource base. Does the system have the potential to provide this, and is there a market demand for the animals and their products? Foresight and total systems thinking are clearly required here;
- the species x trait(s) combination(s) required is not available in locally available germplasm (or breeds/broodstock) or in varieties/pre-breeding materials developed by and available from the International Agricultural Research Centres (IARCs), or other countries growing the same crops within similar production systems and located in similar agro-ecological zones;
- the species x trait combination cannot be developed more easily, and/or at less cost, through phenotypic selection. A number of chapters in this book provide excellent guidance on the factors that are important here, which include, *inter alia:* the species involved; the genetic complexity and heritability of the trait(s) required (the current focus for most crop and many animal species is heavily on disease and pest resistance); the availability of markers for the trait(s) in question and ability to scale up their usage, whether the trait is sex-limited (livestock); and the availability of reliable phenotypic data, etc.;
- there is already an existing national breeding programme(s) for the species in question;
- the national breeding programme(s) for the species in question has the infrastructure and levels of human and financial resources needed to sustain selection and breeding activities;
- national infrastructures and capacities in molecular biology match the scientific, technical and information requirements for effectively supporting MAS;
- professional legal advice is available concerning laws, agreements, licences, etc. for the acquisition and diffusion of both the tools or enabling techniques, and the starting- and end-products of MAS (see Chapter 20 and earlier in relation to access and benefit sharing of genetic resources);
- efficient mass propagation systems (e.g. seed multiplication or semen production programmes) are in place;

BOX 2

Centralized national centres of excellence and sectoral/subsectoral institutions

More technologically advanced developing countries such as Brazil, China, India and South Africa and others have established one or a number of cutting-edge centres for biotechnology working on both basic and strategic techniques and tools, and providing analytical and other support to national sectoral or subsectoral centres working on more applied and adaptive research projects. For example, the African Centre for Gene Technologies (ACGT) is an initiative by the South African Council for Scientific and Industrial Research (CSIR) and the University of Pretoria to create a national centre of expertise and a world-class platform in gene and genome analysis. Its focus is on using genetic markers to understand disease resistance in plants and nitrogen metabolism in cattle under harsh and arid environmental conditions. It supports the more downstream work of the Forestry and Agriculture Biotechnology Institute as well as various crop centres. ACGT is a member of the Southern African Network on Biosciences (SANBIO) and part of the NEPAD-sponsored African Biosciences initiative.

- adequate technical advisory services are able to support technically the dissemination of the improved variety or breed; and
- effective delivery, monitoring and evaluation strategies are in place to bring the products of MAS-related R&D to users and beneficiaries.

COUNTRY OPTIONS FOR IMPLEMENTING MAS

Countries with high-quality personnel and facilities for phenotypic evaluation and selection and in molecular biology

Individually or collectively and for a number of crop species, public institutions in these countries have the skills both to choose the appropriate parental and segregating materials and to apply routinely and with high throughput the full range of techniques available (including those requiring sequence information) to develop molecular markers. Through the establishment of centralized centres of excellence and sectoral/subsectoral institutions (Box 2), they have the potential to validate putative markers by combining their use with the detailed and comprehensive phenotypic information available on large numbers of lines for multiple traits to produce genetic linkage maps for identifying genomic regions controlling variations in simple and quantitative traits (QTL), and to use the right combination of trait-linked markers to improve the efficiency of parental selection and breeding programmes.

These countries tend to have concentrated their MAS activities on the introgression of a few traits (for instance those encoded by transgenes) and in a few crops, although markers themselves are being used for many of the non-MAS applications mentioned earlier. However, they also contribute effectively to global and regional crop genomic projects directed towards developing and validating genetic and linked markers and testing their usefulness for MAS in breeding programmes. They may also have some skills in applying

(some of) these methods and approaches to livestock, fish and especially forest species research. Generally, however, these efforts are of a small-scale experimental nature and, particularly in the case of livestock where most traits, even for disease resistance, are complex, they are unlikely to move beyond the research stage in the near term because of the large numbers of animals required, the limited amount of structured phenotypic data available and the long generation intervals of many animal species.

These countries have both considerable potential and many options to focus resources for MAS on poverty and hunger alleviation, including:

- mobilizing the techniques, tools, genetic resources and phenotypic data already available nationally and internationally (e.g. parental lines and segregating populations from international and other national programmes), tapping the vast and rapidly increasing molecular and genetic knowledge available internationally through collaboration with international and advanced agricultural research centres, and contributing to genomics consortia (e.g. the International Rice Genomics Consortium). Applications include extending non-transgenic and transgenic approaches by developing and validating markers based on fine genomic mapping of QTL (i.e. by identifying single nucleotide polymorphisms, [SNPs]) for more complex traits like drought, salinity and heat tolerance and nutritional quality in major food crops;
- pursuing MAS for both simple and complex traits in crops that although relatively minor and scientifically neglected are of tremendous importance to many poor households;
- recognizing the increasing importance of trees, livestock and aquaculture to their rural economies, strengthening efforts to characterize genetic diversity through both classical phenotypic and molecular marker approaches and then developing, validating and eventually using markers for improving economically important traits such as host resistance to diseases.

Countries with reasonable capacities for phenotype evaluation and selection and some capacities to apply molecular marker methods

These countries have less comprehensive breeding programmes and therefore can cover fewer species. They will likely have been relatively "late starters" in MAS and may not have the latest high-throughput equipment, which is invariably located in one or a number of institutes supporting a particular subsector. Neither of these features is a major limitation provided the country prioritizes its work appropriately. This means pursuing justifiable and doable genetic enhancements to the limited number of species for which it has an effective selection and breeding programme to: (a) provide the foundation for developing segregating populations from parental lines and for characterizing and validating markers for the trait(s) in question; and (b) evaluate populations in the environments for the traits that are prioritized.

Also, while recognizing the need to adapt specific molecular techniques to local circumstances, and markers for particular traits to their own genotypes, these countries should take full advantage of "lessons learned" with respect to both the molecular methods themselves and how best to integrate these into selection and breeding programmes. With the caveat that these conditions are satisfied, countries in this general category have the following options:

Strengthening existing national scientific and technical capacities and infrastructures in molecular laboratory(ies) and coordination with selection and breeding programmes

This can be achieved by:

- relying on their own germplasm and segregating populations and/or partnering with IARCs and advanced research institutes to obtain these. Using less sophisticated and largely "manual" sample preparation and analytical equipment even through to the point of sequencing since large-scale and high-throughput genetic analysers, accessories and other equipment are necessary only after the initial development and implementation of markers;
- taking advantage of the many kits, biological and other materials and the "how to do" and "what to avoid" protocols and manuals, statistical packages, bioinformatics freeware, software and analysis programmes. These, as well as specific markers that are available in the form of DNA clones for use as probes in restriction and amplified fragment length polymorphism (RFLP and AFLP) analysis, PCR primers for use as SSR (microsatellite) markers, and sequence information available in public databases that can be used to synthesize and clone specific markers are available commercially or for free from several of the IARCs belonging to the CGIAR, and from advanced research institutions and universities in developed and developing countries (e.g. Brazil, China and India). All of these resources help to avoid "reinventing the wheel" and to "short-cut" the process by assisting in getting round bottlenecks, e.g. the need to establish facilities and expertise in cloning; and
- taking advantage of the many opportunities available to upgrade scientific and technical expertise in molecular biology itself and in linking molecular and phenotypic approaches through species and theme-specific networks, workshops, training courses, scientific visits, etc. (see Box 3).

Using regional centres of excellence

While there is no real substitute for building national institutions and capabilities in MAS research, product development and transfer, countries with very limited infrastructures and skilled human resources can still engage in meaningful research by using the state-of-the-art analytical, bioinformatics and computing facilities located in regional centres of excellence (Box 4) and in advanced national institutes.

Countries with limited capacities in phenotypic evaluation and selection and no capacities to apply molecular techniques

These countries fall into the category of Type 3 NARES described by Byerlee and Fischer (2001). Unless the government commits itself to increasing substantially its level of investment in essentially all aspects of genetic resources management in one or a number of the different agricultural subsectors or at least one species, but particularly in selection and breeding programmes and in capacity-building for employing molecular techniques, their options are:

- to partner with institutes of the CGIAR system and other advanced institutions in developed and developing countries and import varieties and advanced breeding lines developed by these institutes through MAS that contain the needed trait(s). After testing these or their crosses with local varieties or landraces for

BOX 3

Regional networks

The East African Regional Programme and Research Network for Biotechnology, Biosafety and Biotechnology Policy Development (BIO-EARN) is supported by the Swedish International Development Cooperation Agency and involves Ethiopia, Kenya, Uganda and the United Republic of Tanzania. It has a Governing Board and a Programme Advisory Committee that provides technical advice to the Board on project proposals, progress and continuing funding. MAS is a priority theme within BIO-EARN, being used to look for resistance markers for plant viruses and fungi (including sweet potato, maize, banana and sorghum) and genotype variation in coffee and banana. The programme has contributed substantially to the improvement of scientific and technical capacities, research infrastructure, equipment and stocking of consumables. Connectivity at all BIO-EARN Network institutions has been achieved and this has not only greatly improved communication among network members, but also access to information from the Internet to keep abreast with new biotechnology developments in the world. BIO-EARN Ph.D. students have developed close links with each other through common workshops and annual BIO-EARN student meetings and created a strong basis for future regional collaboration.

BOX 4

Biosciences eastern and central Africa hub and network

Located on the campus of ILRI in Nairobi, Kenya, the Biosciences eastern and central Africa (BecA) hub is the first of four regional networks of centres of excellence in biosciences sponsored by the New Partnership for Africa's Development (NEPAD). The establishment of BecA is funded mainly by the Canada Fund for Africa (CFA) of the Canadian International Development Agency (CIDA). It consists of a hub that provides a common biosciences research platform, research-related services and capacity building and training opportunities, and a network of regional nodes and other participating laboratories distributed throughout eastern and central Africa for conducting research on priority problems affecting Africa's development. It has a Steering Committee and a Scientific Advisory Committee responsible for the quality and relevance of the programme. The genomics platform includes state-of-the art equipment for genotyping, DNA sequencing, transcriptomics and bioinformatics, and current activities include microsatellite and EST marker development, genetic linkage mapping, MAS, and fingerprinting for distinctness, uniformity and stability and plant variety protection. It currently supports work being conducted on MAS by the International Maize and Wheat Improvement Center (CIMMYT), the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), the International Institute of Tropical Agriculture (IITA), and ILRI and their national partners. Further information is available at www.biosciencesafrica.org.

their suitability for local environments, they can be released to producers;
- to increase awareness among policy- and decision-makers of the importance of improving GRFA through a multidisciplinary approach including molecular methods to their national economies and poverty reduction strategies. Opportunities for doing so include through advocacy within national policy dialogue processes and within FAO's Commission on Genetic Resources for Food and Agriculture, and by its country representatives and staff in regional and subregional offices during the processes of revising PRSPs and agricultural development policies; and
- subject to increased investment for high priority activities, training and capacity-building in selection and breeding procedures should precede the introduction of molecular approaches. Both should be initiated through close collaboration with international, regional and/or national centres.

MAS: OTHER POLICY CONSIDERATIONS AND OPTIONS

Few people would question the stark realities of doing any kind of R&D in the vast majority of developing countries and of getting the products generated from it to the rural poor and hungry. Conducting R&D directed towards MAS raises the bar considerably in terms of its requirements for organizational, scientific, technical and legal skills, as well as for physical infrastructure and financial resources. Funds, however, for public sector agricultural R&D in all but a handful of developing countries are becoming ever more scarce.

While data on spending and human resources for modern biotechnology applications in agriculture are not available, inflation-adjusted spending on agricultural R&D as a whole is now growing at much lower rates than in the 1970s and currently runs at around US cents 53 for every US$100 of agricultural output (Pardey *et al.,* 2006). In developed countries, public research funding actually fell by 6 percent per year during the 1990s, but is still running at the rate of US$2.36 per US$100 worth of agricultural output. This reflects a strong shift in funding priorities away from public R&D by both governments and donors.

However, the big differences between these groups of countries lie in two broad and interconnected areas. First, in their levels of private investment. In developing countries, this runs at between 8 percent (in Asia and Pacific, but in only a few countries) and 2 percent (in sub-Saharan Africa, with 66 percent of that being in South Africa), and by and large is devoted to export crops and conducted by locally-owned companies or affiliates of multinationals. In countries of the Organisation for Economic Co-operation and Development (OECD), such investments now form around 55 percent of their total agricultural R&D spending (Pardey *et al*., 2006), with 93 percent of that R&D being performed in these countries.

The second difference lies in the organization/orientation of their research. In developed countries, there is a much clearer division of labour between the public and private sectors. This generally conforms to the notions of "public goods" and profit/market-oriented R&D, although for MAS this demarcation differs across commodities and is often characterized by public–private sector research collaboration. For example, MAS-related R&D activities conducted by public sector institutions are very much oriented towards basic or strategic research to develop and validate new knowledge, methods and procedures

for variety, strain or breed selection through markers and quantitative genetics, attending to minor species and removing bottlenecks. However, for some crops such as maize (see, for example, Chapter 8), wheat, soybeans and cotton, and for some livestock (Narrod and Fuglie, 2000) and aquaculture species, the private sector is a significant player in both the upstream and applied molecular biology and quantitative genetics components.

This situation reflects the varying incentives provided for private agricultural research by a combination of income-driven demand-led market growth for the commodity(ies) or value chains in question, new technologies, changes in IPR regimes, market structure and the globalization of agricultural input markets, and public science and investment policies that have both supported private and undercut publicly-funded research. However, the impulses provided by effective demand-led market growth of commodity chains and by policies to promote private sector investment in R&D are much weaker in lower income countries, and governments that simply lack the cash are left to pick up the total bill. This, in turn, blurs the focus of the R&D conducted by their NARES which, rather than directing resources more towards science-oriented pre-product research (such as molecular marker development and validation for selection), attempt, often within one or two institutes, to cover the whole spectrum from strategic, applied and adaptive research, through to development and on to diffusion of products and services.

At the same time, the wider and interlinked contexts within which the agricultural sector now operates are increasingly requiring ministries, and the research institutes responsible for agriculture that are under them, to forward proposals for policies, legislation, programmes and projects that are not only sound, convincing and prioritized within and between subsectors, but also aligned with the needs perceived by other ministries, e.g. health, education, trade and the environment. Critically, in preparing plans for both domestic and donor finance, they must provide convincing evidence of engagement with those representing the interests of agricultural producers and other sectors of rural society including women's groups and the poor, private commercial and non-government organizations (in addition to involving their own officials and technical experts).

In other words, the pressure is real and growing both nationally and internationally for more "joined up" governance and greater participatory diagnostic and decision-making in helping to define, implement and assess the outcomes and impacts of public sector interventions, the expectation being that this will focus both minds and funds on tackling the problems of greatest relevance to the largest number of poor people in rural areas. This raises the issue of how NARES can better ensure that their agendas, including plans for using modern biotechnological approaches like MAS (that clearly requires long-term budgetary support), better meet the needs of the poor.

Besides macro and sectoral policies that provide appropriate price and market incentives to agricultural producers and service providers, developing countries have a number of options for creating the more conducive and enabling environment necessary for MAS research and the development and adoption of the products that emanate from it.

Building political support

The biggest policy gap in many developing countries is perhaps the lack of official

appreciation of the importance of S&T for meeting their socio-economic objectives through agriculture. Hence, the necessity for agricultural research and extension institutions to engage in dialogue during the processes of revising PRSPs, comprehensive agricultural development and related policies and strategies cannot be underestimated. This promotes learning and capacity development among and between policy- and law-makers, technical experts and civil society, as well as greater appreciation of poverty and its different dimensions and the trade-offs between different approaches to its amelioration. It results in stronger linkages between S&T, national poverty reduction and agricultural development objectives, and greater awareness among high-level policy-makers of the contributions that S&T can make to achieve these objectives. The merits of such engagement also include greater need-driven priority-setting, elimination of duplication, more informed decision-making both on the roles of the public and private sectors, and partnership identification. All of this helps to create more efficient and coordinated activities within and between the different agricultural subsectors and their supporting NARES. Well-conceived studies on the socio-economic impacts of crops and breeds developed through MAS would also assist decision-making on S&T investment allocation as this is hardly available in the literature (FAO, 2005b).

Creating S&T policies for driving stronger priority-setting and better delineating roles and responsibilities

Many developing countries continue to work with outdated, isolated and highly fragmented NARES, each with their own set of rules, fiscal arrangements and government oversight, and they have weak or non-existent linkages between public institutions, the IARCs and private firms. Some possess many of the essential components including cutting-edge equipment, but they are not maximizing their potential to develop MAS capacity. While substantial progress has been made by a number of countries to establish a single framework for managing agricultural R&D, including mechanisms for setting and evaluating priorities, in most developing countries these are rare. Making a case for MAS, unsupported by a well thought out (evidence/diagnostic-based) S&T policy framework that promotes mutually supportive actions by the different actors, is a recipe for continuing with a science and supply-driven research agenda, minimal interaction among the different institutions involved, underfunding and tinkering around the edges by some dedicated individuals.

There are, however, a number of options open to governments to reinvigorate their agricultural S&T systems and make way for new technologies:

- use wider or agriculture-specific S&T legislation to promote enhanced collaboration among public sector institutions and between public and private sector entities, and to establish a national funding agency and/or agricultural research council that is independent from any specific ministry and provides competitive grants for research and fellowship training within the public sector. Thailand, through its National Science and Technology Development Agency established under a S&T Development Act, and Brazil with its National Council and National Fund for S&T Development and its Sectoral Funds (Box 5), are two good examples for other countries to consider following, e.g. some African countries that under NEPAD have committed themselves to increasing

BOX 5

Brazil's sectoral funds

To promote high-quality research and development in Brazil's industrial sector, the national government established a programme of "sectoral funds" in which a percentage of corporate taxes are targeted to funding specific research and development objectives. The sectoral funds programme serves four major national objectives:

- stability of financial resources for medium- and long-term research and development;
- transparency in funding decisions, merit review and evaluation;
- reduction of regional inequalities; and
- interaction between universities, research institutes and companies.

The selection of strategic sectors, their respective shares of the funds' resources, the blend of basic and applied research, the required overall budget, and sources of support are all jointly decided upon by the indigenous academic community, private sector, and government. No new taxes are involved, just the redirection of already-established government levies. A comprehensive set of 14 funds has been established. It includes agriculture, biotechnology, informatics and university-industry research.

their S&T spending from 0.5 to 1 percent of GDP.

- use wider or specific agriculture legislation, tax breaks, public S&T funds and funds from donors to provide incentives for public and private sector involvement in MAS and public–private, civil society collaboration.

Governments and development agencies have shown increasing interest in partnerships as a mechanism to promote market-driven development. Byerlee and Fischer (2001) provide an excellent review on the subject of enhancing public–private partnerships for transfer of genes and constructs for GM crops, and some of the principles and mechanisms described are also relevant to MAS. However, outside of the Latin American region where Hartwich, Gonzalez and Veira (2005) conducted a study of 124 cases of such partnerships in agricultural innovation, including a number dealing with basic and applied research on plant breeding, the evidence that these make research more efficient or "deliver" more or better products to small-scale producers is not strong. Indeed, ensuring that these partnerships comply with public needs was found to be one of the major challenges in these partnerships.

Nevertheless, there are clear signs of such partnerships flourishing for both basic and applied MAS-related R&D activities in industrialized countries. With strong and enlightened leadership on both sides coupled with matching interests, opportunities should also be available for NARES to benefit both from the molecular technology platforms and from the selection and breeding experiences of industrialized country-based private sector entities (Chapter 8), as well as from cooperation with their developing country affiliates, local private companies, NGOs and producer organizations in taking the products from the laboratory to the field.

At the same time, the nature and scope of IPRs for genetic resources, the research

tools used in MAS and the breeding lines and varieties that are developed from it may be significant barriers to its furtherance by NARES, private sector entities and public–private partnerships. An excellent treatise of the influence of IPRs for plant breeding and the seed sector in developing countries, including the possible implications for MAS, is given by Tripp, Eaton and Louwaars (2006). By providing an empirical analysis of IPR developments in China, Colombia, India, Kenya and Uganda, this report provides government and institutional policy- and decision-makers with details of the different challenges faced by these countries in developing their IPR regimes, the options they have chosen through policies and legislation to develop these, and the lessons learned in implementing them. It concludes that while IPR regimes in developing countries require urgent attention, the supporting legislation and regulations should be the product of open debate among different stakeholders, and that, even if legislation is already in place, many countries will find that they have sufficient options for interpretation and application to warrant a thorough review of procedures and priorities. In terms of hunger and poverty reduction, the importance of segmenting markets into export and non-export crops, and into major and orphan crops, should not be overlooked for gaining preferential access to molecular tools, breeding lines and varieties (Spillane, 2000).

- make greater use of regional and bilateral agreements and organizations to foster international collaboration and obtain complementary assets for the furtherance of MAS.

Diplomatic level S&T agreements, knowledge exchange networks and research consortia (including those of a national and regional nature) can all build knowledge within and between molecular laboratories, genetic resources management programmes and organizations involved in product delivery. The benefits to developing countries of both the formal and informal networked world of collaborative research in molecular biology, genetic improvement and agricultural S&T in general are potentially enormous, providing sources of funding and making knowledge easier to access and researchers and policy-makers more interconnected (safeguards are needed, however, to minimize disadvantages/risks). Developing countries are not sufficiently linked to these resources, the CGIAR's Generation Challenge and the EU Research Framework Programmes and competitive grants with partners from individual or groups of developed and developing country research institutions funded by bilateral donors being just some examples. Their governments need to help them do so by providing funds for building broadband connections, establishing databases and information systems, and attending conferences. Also, some functions of a R&D system such as accessing IP-encumbered technology may be accessed virtually or even shared with neighbouring countries (Box 6).

Creating effective delivery strategies to bring the products of MAS-related R&D to users and beneficiaries

The channels through which the products of agricultural research reach producers have undergone major structural changes worldwide, and there is now a wide range of public, private and non-government organizations involved in providing extension services. At the same time, those responsible for funding and supporting R&D have come to realize that getting technology and

BOX 6

Partnerships for technology transfer: the African Agricultural Technology Foundation

The mission of the African Agricultural Technology Foundation (AATF) is to promote food security and reduce poverty. It is a not-for-profit foundation set up in 2002 with the help of the Rockefeller Foundation, the United States Agency for International Development (USAID) and the United Kingdom's Department for International Development (DFID) to identify opportunities for royalty-free transfers of technologies useful to resource-poor smallholder farmers in Africa. In pursuing its mission, it negotiates access to technologies, enters into contractual arrangements to facilitate their deployment and provides stewardship over their deployment. It is the responsible party for addressing the concerns about technology owners while protecting the interests of smallholders, handling intellectual property management, regulatory compliance, liability, licensing and freedom-to-operate assessments. In effect it is a "one-stop-shop" for structuring and accessing agricultural technologies and know-how. Among its priorities is genetic improvement of cowpea. This is being tackled through a Network for the Genetic Improvement of Cowpea for Africa, involving African and United States' universities, IITA, the Kirkhouse Trust, which is a United Kingdom charity, and Monsanto. Part of the project involves developing ready-to-use molecular marker kits for cowpea breeding teams in Africa. The markers are being selected from the cowpea SNP programme in IITA and the cowpea genome-sequencing programme, with polymorphisms detected using agarose gel systems.

knowledge to the field does not follow a linear and top-down transfer path that begins with research, moves on to development and production, and ends with the successful introduction of new products or processes. Instead, it involves continuous feedback loops between researchers, extension agents and farmers within which the development, fine-tuning and adoption of the products of research take place within a specific context. All too often, technologies lie "on the shelf", lost between research and its transformation into useful products, because of the lack of understanding of the functional linkages between research institutions, extension services and farmers.

Policy- and decision-making about supporting agricultural research and technology development therefore need to shift away from traditional and often laboratory-based research and the supply of technology *per se*, towards fostering an innovation systems approach to understand better the ways in which the producers and users of technology interact, and thereby identify and get round the obstacles faced in transforming research outputs into development outcomes and impacts. For example, why has artificial insemination, a technology central for driving improved genetics into many species of farm animals, been successful in some countries and localities and not in others and what are the implications of this for applying MAS?

Unrealistic expectations of what agricultural research and agricultural biotechnology in particular can do to alleviate hunger and poverty have had a negative

impact on the fiscal policies of national governments, financial institutions and donors for more than 20 years. It will be important, therefore, for policy-makers to provide incentives for getting the facts right before deciding on priorities and investments. They can do so by mandating that greater emphasis is placed on research to understand better the critical pathways involved in technical change, including the reasons for the long time frames between the research and extension efforts of their NARES and sustainable improvements in farm productivity through genetic enhancement. This can be achieved *inter alia* by requiring greater accountability, for example, through up-front specification of the R&D delivery strategy, and the introduction of monitoring and evaluation processes for research outputs and outcomes that use an innovation systems approach to promote information flow, and through this, to understand and improve current needs and priority assessments and levels of customer satisfaction.

REFERENCES

Bijker, W.E. 2007. Science and technology policies through policy dialogue. *In* L. Box & R. Engelhard, eds. *Science and technology policy for development: dialogues at the interface.* London, Anthem Press (in press).

Bragdon, S. 2004. International law of relevance to plant genetic resources: a practical review for scientists and other professionals working with plant genetic resources. *Issues in Genetic Resources* 10. Rome, IPGRI (available at www.bioversityinternational.org/Publications/pubfile.asp?ID_PUB=937)

Byerlee, D. & Fischer, K. 2001. Assessing modern science: policy and institutional options for agricultural biotechnology in developing countries. *IP Strategy Today* 1: 1–27 (available at www.biodevelopments.org/ip/ipst1n.pdf).

Crawford, I.M. 1997. Rapid rural appraisals, Chapter 8. In *Marketing research and information systems*. Rome, FAO (available at www.fao.org/docrep/W3241E/w3241e09.htm)

De Vicente, M.C. 2004. The evolving role of genebanks in the fast-developing field of molecular genetics. *Issues in Genetic Resources* 11. Rome, IPGRI (available at www.bioversityinternational.org/Publications/pubfile.asp?ID_PUB=986).

Dixon, J., Gulliver, A. & Gibbon, D. 2001. *Farming systems and poverty: improving farmers' livelihoods in a changing world*. Rome, FAO and Washington, DC, World Bank.

Donnenworth, J.M., Grace, J. & Smith, S. 2004. Intellectual property rights, patents, protection and contracts. *IP Strategy Today* 9: 19–33 (available at www.biodevelopments.org/ip/ipst9.pdf).

Evanson, R.E. & Gollin, D. 2003. Assessing the impact of the green revolution. *Science* 300: 758–762.

FAO. 2005a. *The legal framework for the management of animal genetic resources*, by A. Ingrassia, D. Manzella and E. Martyniuk. FAO Legislative Study 89. Rome (also available at www.fao.org/docrep/009/a0276e/a0276e00.htm).

FAO. 2005b. *Annotated bibliography on the economic and socio-economic impact of agricultural biotechnology in developing countries* (available at ftp://ftp.fao.org/sd/SDR/SDRR/bibliography1.pdf).

FAO. 2006. The state of food insecurity in the world 2006: eradicating world hunger-taking stock 10 years after the World Food Summit (available at www.fao.org/docrep/009/a0750e/a0750e00.htm).

Hartwich, F., Gonzalez, C. & Viera, L-F. 2005. Public-private partnerships for innovation-led growth in agrichains: a useful tool for development in Latin America? *ISNAR Discussion Paper* 1. Washington DC, IFPRI (available at www.ifpri.org/divs/isnar/dp/papers/isnardp01.pdf).

IAC. 2004. *Realizing the promise and potential of African agriculture.* Amsterdam, InterAcademy Council (available at www.interacademycouncil.net/CMS/Reports/AfricanAgriculture.aspx).

IPGRI. 1999. Key questions for decision-makers: protection of plant varieties under the WTO Agreement on Trade Related Aspects of Intellectual Property Rights. *Decision tools.* Rome (available at www.bioversityinternational.org/publications/pubfile.asp?ID_PUB=41).

Le Buanec, B. 2004. Protection of plant-related innovations: evolution and current discussion. *IP Strategy Today* 9: 1–17 (available at www.biodevelopments.org/ip/ipst9.pdf).

Meinzen-Dick, R., Adato, M., Haddad, L. & Hazell, P. 2004. Science and poverty: an interdisciplinary assessment of the impact of agricultural research. Washington, DC, IFPRI (available at www.ifpri.org/pubs/fpr/pr16.pdf).

Narrod, C.A. & Fuglie, K.O. 2000. Private investment in livestock breeding with implications for public research policy. *Agribusiness: an International Journal,* 16: 457–470.

Ndeng'e, G., Opiyo, C., Mistiaen, J. & Kristjanson, P. 2003. *Geographic dimensions of well-being in Kenya*, Volume 1. Nairobi, The Regal Press Kenya Ltd.

NEPAD. 2002. Comprehensive Africa Agriculture Development Programme (CAADP) (available at www.nepad.org/2005/files/progressrepport.php).

NEPAD. 2005. *Africa's science and technology consolidated plan of action* (available at www.nepadst.org/doclibrary/pdfs/ast_plan_of_action.pdf).

OECD. 2006. *Promoting pro-poor growth: agriculture*. Paris, OECD Publishing (available at www.oecd.org/dataoecd/9/60/37922155.pdf).

Pardey, P.G. & Beintema, N.M. 2001. *Slow magic: agricultural R&D a century after Mendel*. Washington, DC, IFPRI (available at www.ifpri.org/pubs/fpr/fpr31.pdf).

Pardey, P.G., Beintema, N., Dehmer, S. & Wood, S. 2006. *Agricultural research: A growing global divide?* Washington, DC, IFPRI (available at www.ifpri.org/pubs/fpr/pr17.pdf).

Raitzer, D.A. 2003. *Benefit-cost meta-analysis of investment in the international agricultural research centres of the CGIAR*. FAO, Rome, CGIAR Science Council Secretariat (available at http://impact.cgiar.org).

Spillane, C. 2000. Could agricultural biotechnology contribute to poverty alleviation? *AgBiotechNet* 2: 1–72.

Spillane, C., Engels, J., Fassil, H., Withers, L. & Cooper, D. 1999. Strengthening national programmes for plant genetic resources for food and agriculture: planning and coordination. *Issues in Genetic Resources* 8. Rome, IPGRI (available at www.bioversityinternational.org/Publications/pubfile.asp?ID_PUB=29).

Stannard, C., van der Graaff, N., Randell, A., Lallas, P. & Kenmore, P. 2004. Agricultural biological diversity for food security: shaping international initiatives to help agriculture and the environment. *Howard Law J.,* 48: 397–430.

Tripp, R., Eaton, D. & Louwaars, N. 2006. *Intellectual property rights; designing regimes to support plant breeding in developing countries.* Washington DC, The World Bank (available at http://siteresources.worldbank.org/INTARD/Resources/IPR_ESW.pdf).

UN Millennium Project. 2005a. *Investing in development: a practical plan to achieve the Millennium Development Goals.* London and Sterling, VA, USA, Earthscan (available at www.unmillenniumproject.org/reports/fullreport.htm).

UN Millennium Project. 2005b. *Report of the task force on science, technology and innovation*. London and Sterling, VA, USA, Earthscan (available at www.unmillenniumproject.org/documents/Science-complete.pdf).

Viljoen, G.J., Nel, L.H. & Crowther, J.R. 2005. *Molecular diagnostic PCR handbook.* Dordrecht, Netherlands, Springer.

This book provides a comprehensive description and assessment of the use of marker-assisted selection for increasing the rate of genetic gain in crops, livestock, forestry and farmed fish, including the related policy, organizational and resource considerations. It continues FAO's tradition of dealing with issues of importance to agricultural and economic development in a multidisciplinary and cross-sectoral manner. As such it is hoped that the information and options presented and the suggestions made will provide valuable guidance to scientists and breeders in both the public and private sectors, as well as to government and institutional policy- and decision-makers.

ISBN 978-92-5-105717-9

9 789251 057179

TC/M/A1120E/2/01.09/1500