CGRFA/WG-AnGR-6/10/Inf.7

October 2010

ENGLISH ONLY

منظمة الأغذية والزراعة للأمم المتحدة
联合国粮食及农业组织
Food and Agriculture Organization of the United Nations
Organisation des Nations Unies pour l'alimentation et l'agriculture
Продовольственная и сельскохозяйственная организация Объединенных Наций
Organización de las Naciones Unidas para la Agricultura y la Alimentación

Item 3.2 of the Provisional Agenda

COMMISSION ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE

INTERGOVERNMENTAL TECHNICAL WORKING GROUP ON ANIMAL GENETIC RESOURCES FOR FOOD AND AGRICULTURE

Sixth Session

Rome, 24-26 November 2010

DRAFT GUIDELINES ON MOLECULAR GENETIC CHARACTERIZATION

Table of Contents

DRAFT GUIDELINES ON MOLECULAR GENETIC CHARACTERIZATION

FOREWORD

Animal genetic resources for food and agriculture are an essential component of the biological basis for world food security. Hundreds of millions of poor rural people keep livestock and often rely on their animals to provide multiple products and services. In harsh environments where crops will not flourish, livestock keeping is often the main or only livelihood option available. Livestock currently contribute about 30 percent of agricultural gross domestic product in developing countries, with a projected increase to about 40 percent by 2030. The World Bank has estimated that it will be necessary to increase meat production by about 80 percent between 2000 and 2030. This will require more efficient animal production systems, careful husbandry of natural resources and measures to reduce waste and environmental pollution.

The *State of the World's Animal Genetic Resources for Food and Agriculture* provided for the first time a comprehensive country-driven global assessment of the roles, values and status of animal genetic resources. It clearly shows that the diversity of animal genetic resources continually decreasing and that potential of the remaining diversity for increasing food security and improving livelihoods is not being fully realized. As a global response to these shortcomings, the member states of FAO developed the *Global Plan of Action for Animal Genetic Resources*, which was adopted by the International Technical Conference on Animal Genetic Resources for Food and Agriculture held in Interlaken, Switzerland, in September 2007 and subsequently endorsed by all FAO member countries.

The *Global Plan of Action* contains four Strategic Priority Areas, which provide a basis for enhancing sustainable use, development and conservation of animal genetic resources throughout the world. Its implementation will contribute significantly to achieving Millennium Development Goals 1 (Eradicate extreme poverty and hunger) and 7 (Ensure environmental sustainability). The first of the Strategic Priority Areas is the characterization, inventory and monitoring of trends and associated risks of AnGR. The characterization of AnGR is necessary in order to properly assess value of breeds and to guide decision making in livestock development and breeding programmes.

The Interlaken Conference called on FAO to continue developing technical guidelines and technical assistance and to continue coordinating training programmes as a means to support countries in their efforts to implement the Global Plan of Action. The objective of these guidelines on *molecular characterization of animal genetic resources* are to help countries to plan and implement effective analyses of the genetic diversity of their AnGR, so the resulting information can contribute to the development of effective plans for sustainable use and conservation of AnGR. They provide countries with advice on how to:

- Plan molecular characterization studies;
- Collect and evaluate DNA;
- Organize and analyze molecular characterization data;
- Interpret the results; and
- Collaborate internationally to obtain the maximum utility of the information.

The preparation of the guidelines was initiated in accordance with Strategic Priority 2 of the *Global Plan of Action*, which addresses the need to develop international technical standard and protocols for characterization of animal genetics resources. The guidelines were discussed and validated at a series of workshops. In total, more 40 scientists, technicians and policy-makers from all regions had an opportunity to contribute to the process.

The guidelines will be refined and updated periodically as experience with their use in the field is accumulated and as technologies for molecular characterization advance. The assistance of the

National Coordinators for the Management of Animal Genetic Resources and their country networks will be particularly important to this process of revision.

ACKNOWLEDGEMENTS

These guidelines revise and build upon the *Secondary Guidelines: Measurement of Domestic Animal Diversity (MoDAD)*. They were developed in strict collaboration with the International Society of Animal Genetics (ISAG) – FAO Advisory Group on Animal Genetic Diversity and with GLOBALDIV, “A global view of livestock biodiversity and conservation”, a three year project funded by the European Commission in the framework of the AGRI GEN RES initiative. Paolo Ajmone-Marsan and Hans Lenstra, participants in both of these groups were involved in the planning, writing and review of the guidelines. Other members of the ISAG-FAO Advisory Group and of GLOBALDIV provided critical technical reviews. The guidelines were presented and evaluated at workshops held in Poland and Austria.

The guidelines were prepared under the supervision of Paul Boettcher, with the full support of the Chief of FAO’s Animal Genetic Resources Branch, Irene Hoffmann, and of present and former Animal Genetic Resources Officers and FAO volunteers: Badi Besbes, Beate Scherf, Dafydd Pilling, Roswitha Baumung and Joanna Marchewka. Administrative and secretarial support was provided by Kafia Fassi-Fihri and Silvia Ripani.

FAO would like to express its thanks to all these groups and individuals.

USER GUIDANCE

In 1993, a FAO working group proposed a global programme for characterization of AnGR, including molecular genetic characterization, and formulated the *Secondary Guidelines: Measurement of Domestic Animal Diversity (MoDAD)* (FAO, 1993) with recommendations for the molecular analysis of domestic animal diversity via a research program to be coordinated by the FAO.

The FAO has revised its original MoDAD guidelines in close consultation with the International Society for Animal Genetics (ISAG), through the ISAG/FAO Advisory Group on Animal Genetic Diversity, as well as scientists contributing to the GLOBALDIV project (www.globaldiv.eu) sponsored by the European Union. In this report, the role of the FAO has been redefined to mainly provide technical support to countries and to coordinate networking across countries, while the technical recommendations have been adapted accounting for the continuing progress of insight and advances in genetic technology. The broad objective of the revised manual is to provide guidance on performing studies on molecular characterization of AnGR.

The different sections of the guidelines target various audiences. For all stakeholders in management of AnGR we survey the accomplishments of the molecular approach after 16 years, summarizing realized benefits and future prospects. Next, for scientists we formulate recommendations for organizing a molecular survey of the genetic diversity. This starts with strategic choices to be made in the planning phase and continues with specific recommendations and technical pitfalls of sampling, genotyping and data analysis. This includes legal aspects of ownership and transfer of genetic resources as well as the translation of scientific results into genetic management. We emphasize standardization and integration of data into international analyses. Finally, various appendices provide a glossary of technical terms, examples of questionnaires to gather information about collection sites, animals and breeds during the process of sample collection, an example of a simple material transfer agreement, a summary of software that can be used to analyze molecular data and the standard ISAG/FAO Advisory Group panels of microsatellite markers for nine common livestock species.

ABBREVIATIONS USED

AnGR	Animal genetics resources (for food and agriculture)
AFLP	Amplified fragment length polymorphism
CNV	Copy number variation
EAAP	European Association of Animal Production
IAEA	International Atomic Energy Agency
ILRI	International Livestock Research Institute
ISAG	International Society for Animal Genetics
MHC	Major histocompatibility complex
MoDAD	Measurement of Domestic Animal Diversity
PCR	polymerase chain reaction
RAPD	random amplification of polymorphic DNA

1. RATIONALE FOR CHARACTERIZATION OF ANIMAL GENETIC RESOURCES

Domestic animal diversity is an important component of global biodiversity. About 40 species of domestic animals and poultry contribute to meeting the needs of humankind, providing meat, fibre, milk, eggs, draught animal power, skins, and manure, and are an essential component of many mixed farming systems. Within these species, more than 7000 breeds and strains (FAO, 2007b) constitute the animal genetic resources (AnGR) that are of crucial significance for food and agriculture.

The present pattern of diversity of AnGR is the result of a long and complicated history, particularly by the process of domestication, which took place 10,000 to 3,000 years ago, depending upon the species. Since then, domestic livestock have spread with human migrations and trading to all inhabited continents. Local adaptation, artificial selection, mutations and genetic drift then turned the genetic diversity captured with domestication into a vast array of differences in appearance, physiology and agricultural traits. During the last centuries this differentiation was accentuated by the emergence of breeds, more or less isolated populations that were subject to systematic selection. This development, and the degree of genetic isolation, was most pronounced in the temperate zones where the demands of food supply led to a rationalization of agriculture. The last 50 years saw the global spread of a few highly developed breeds, like the highly productive dairy Holstein Friesian cattle.

The global diffusion of these specialized breeds, which mostly originated in Europe, is endangering or even risking the extinction of many well adapted local breeds. This trend is particularly evident in marginal areas (Rege and Gibson, 2003) where local husbandry practices are also being abandoned (Köhler-Rollefson et al., 2009). As a consequence, erosion of diversity of AnGR has become a major concern (Hammond and Leitch, 1996; Hodges, 2006, FAO 2007a). The negative consequences of genetic erosion and inbreeding depression have been amply documented and is may be manifested by loss of viability, fertility and disease resistance and the frequent occurrence of recessive genetic diseases (Keller and Waller, 2002; FAO 2007b; Taberlet et al., 2008). According to the recently published State of the World’s Animal Genetic Resources for Food and Agriculture (FAO, 2007b), in the past century, approximately 10% of farm animal breeds have become extinct and an additional 15% are considered rare or endangered. Moreover, the situation is presently unknown for 34% of the breeds, most of which are reared in developing countries.

The FAO has a history of working with its member states in order to improve the productivity of livestock and the livelihoods their citizens while maintaining AnGR diversity (FAO 1990a,b,c). In order to outline specific priorities for AnGR management, the member countries of the FAO formulated the *Global Plan of Action for Animal Genetic Resources* (FAO, 2007a), which was adopted at the International Technical Conference for Animal Genetic Resources in Interlaken, Switzerland.

One of the Strategic Priority Areas of the *Global Plan of Action* is the characterization, inventory and monitoring of trends of AnGR diversity in order to properly assess value of breeds and to guide decision making in livestock development and breeding programmes.

The breed is the operational unit in conservation of genetic resources, but this concept covers groups of animals having genetic characteristics that depend on geographical area and species. Most breeds originating from industrialized countries are well-defined and phenotypically distinct and were usually genetically isolated throughout the course of their development. In contrast, Asian and African breeds most often correspond to local populations that differ only gradually according to geographical separation. In addition, breeds with different names may sometimes have a recent common origin, while in other cases their uniqueness has been eroded by crossbreeding.

In the absence of information about the genetic attributes of each breed available for a breeding programme, development of local breeds is often ignored in favour of the introduction of germplasm from exotic breeds, about which more information is generally available.Therefore, characterization of breeds both at the level of animal phenotypes and their interaction with production systems and at the genetic level is most essential.

2. MOLECULAR CHARACTERIZATION OF DIVERSITY

Since the beginning of the 1990’s, molecular data have become more and more relevant for the characterization of genetic diversity (Groeneveld et al., 2010). In 1993, a FAO working group proposed a global programme for characterization of AnGR, including molecular genetic characterization, and formulated the *Secondary Guidelines: Measurement of Domestic Animal Diversity* (MoDAD) (FAO, 1993) with recommendations for the molecular analysis of domestic animal diversity on a global scale via a research program to be coordinated by the FAO.

Although the MoDAD program recommended originally by the Working Group was not realized, the FAO MoDAD report succeeded in creating awareness of the necessity to monitor AnGR diversity and establishing a standard approach for molecular genetic characterization. In addition, the proposal of the programme helped motivate many nationally funded research projects as well as larger regional and international projects supported by organizations such as the European Commission, the Nordic Council of Ministers, the International Atomic Energy Agency (IAEA), International Livestock Research Institute (ILRI) and the World Bank, which together achieved many of the original objectives of MoDAD. Scientists in many countries have undertaken independent studies to characterize locally available breeds, while large-scale international efforts on breed characterization have built comprehensive molecular datasets for most livestock species. The study of genetic diversity of livestock at the molecular level has developed into a most active area of research, which for example receives considerable attention in scientific press and at conferences of organizations such as ISAG and the European Association of Animal Production (EAAP).

So far, most molecular work was based on the use of neutral genetic marker data, which served as a proxy or estimate of the likelihood of important functional genetic variation within breeds or breed groups. This has accomplished the following:

- The identification of the wild ancestral species of most livestock species and a localization of the site(s) of domestication.
- An assessment of the genetic constitution of breeds via quantitative measures of the diversity, admixture or subdivision, signalling inbreeding, introgression and local assortative mating.
- A partial reconstruction of the phylogenetic relationships of populations, unravelling the evolutionary history of species and population.
- Investigation of algorithms that derive a prioritization of breeds for conservation from molecular data. This would be directly relevant for programmes of conservation, re-stocking or utilisation of AnGR.
- The introduction of these data into the permanent scientific record and the opportunity to publicize the issues surrounding loss of livestock diversity in the international press, the interest of which may be attracted by the particularly well-done and intriguing studies.
- The establishment of an international network of organizations and institutions interested in molecular studies of AnGR.

Despite these accomplishments, the objective of a comprehensive characterization of livestock diversity that would guide the genetic management of AnGR has not been achieved and will require substantial investment in time and financial resources. Therefore, characterization programmes should be designed with maximum efficiency to yield the most valuable information. More specifically, in the near to medium term, the molecular characterization of AnGR should target the following objectives to overcome the limitations of the characterization data collected to date:

- A global view of diversity by integration of national or regional datasets. Too many projects were undertaken independently without coordination with other studies. This practice has very often led to using private panels of genetic markers, which precludes the joining of datasets and seriously decreases the impact of the study undertaken.

- Addressing the “white spots” on the current phylogeographic maps. Most studies have focused on the standardized breeds in industrialized countries with relatively less attention to many authentic local breeds, most notably the African and Asian livestock, which has remained largely free of systematic selection and harbours much of the original diversity.
- Identifying genomic regions involved in functional diversity: adaptation, disease resistance or productive traits.
- Determining how molecular data may best contribute to the optimal management of AnGR. Current studies may, via the discovery of unique molecular variants, via evidence for an independent development or conversely, via identifying breeds that are nearly identical genetically, suggest for some breeds a relative value for conservation. However there is no wide consensus on how to use molecular data for selection of AnGR for conservation. This shortcoming is in part because previous characterization studies have mainly considered neutral variation, which represents only a minor and non-functional fraction of the differences between breeds and individuals.

By recent technological development, the accomplishment of these objectives no longer relies exclusively on microsatellite genotyping, which after 1990 had revolutionized the science of molecular genetics. Dense marker panels of single nucleotide polymorphisms (SNP) are becoming available for most livestock species. This technology will require a new range of methods of analysis for inferring population structure and relationships through approaches such as network construction, model-based clustering, coalescent theory, population genomics, and identification of “selection signatures”, or regions of the genome subject to selection. In addition, whole genome sequencing has been completed or is under way for most major livestock species.

These new tools are likely to substitute microsatellites in many applications. The costs and benefits of re-examining characterized breeds must be considered. Models to link existing information on genetic diversity largely based on microsatellite markers to new information about to be generated with SNP markers may be needed.

In anticipation of the advent of personal genomics, as a more basic unit of conservation one may consider the DNA variation, which may have different modes of distribution: from a group of breeds with a special phenotype or native location within a specific geographic or agro-climatic area, to a specific breed, and to individual animals and their offspring.

3. HOW TO CARRY OUT MOLECULAR DIVERSITY STUDIES

In order to accomplish the above-mentioned objectives and to maximise the output of projects, both in scientific and agricultural terms, the following recommendations have been proposed. These recommendations are based on the combined experience gained during several projects.

3.1 Before you start

Ideally, molecular characterization should be undertaken as part of a comprehensive national programme for management of AnGR, the development of which is outlined in *FAO Guidelines for Preparation of National Strategies and Action Plans for Animal Genetic Resources*. For maximum efficiency, molecular characterization of AnGR should be done in concert with phenotypic characterization, which is addressed in *FAO Guidelines for Phenotypic Characterization of Animal Genetic Resources*.

3.1.1 Know your breeds

Collect and critically evaluate the available information on the breeds you want to investigate: scientific literature, breed handbooks, FAO Global Data Bank or other data banks, nonscientific literature and even anecdotal information. It is most relevant to identify the traditional rearing area and any evidence for genetic subdivision: different ecotypes, phenotypes, agroclimatic zones or isolated subpopulations.

3.1.2 Involve local experts

Most of the expertise on local breeds rests with the farmers and breeding societies, who should be informed on the objectives of the study. Coordination with national breeding societies and livestock research institutions is desirable, as they are expert on the breeds, are familiar with local circumstances and may serve as liaison with the owners. Inform also FAO National Coordinators for Animal Genetic Resources, because, as noted earlier, characterization studies should ideally be undertaken as part of an over National Strategy and Action Plan for AnGR.

3.1.3 Define the objectives

Ranging from an inventory of the pattern of diversity to a reconstruction of the history of breeds or formulation of specific guidelines for genetic management, objectives are most relevant for the sampling, choice of markers and data analysis.

3.1.4 Act locally, think globally

The data collected will invariably become more interesting if analyzed and evaluated in an international context. Combining results with other datasets requires the use of the same molecular markers, the use of common reference samples and, preferably, having one or more breeds in common.

3.1.5 Define the scope

Depending on the breeds and the objectives of the study, the following considerations may be relevant.

- Breeds most likely to be distinct from other breeds are those with a long history of genetic isolation, raised in a unique environment or having unique phenotypes.
- Priority should be given to local breeds, but common or economically important international transboundary breeds should be included as a reference.
- For regional transboundary breeds it will be useful to include populations across the borders or to collaborate with institutes who have studied those populations.
- For breeds that are of hybrid origin by introgression, upgrading or by the planned creation of a synthetic breed, it is essential to have data from parental breeds.

- For breeds having a recent history of intense selection and/or inbreeding, sampling of animals of previous generations, which may be available by cryopreservation of semen samples or from museum specimens, may be appropriate.
- For mammalian species, sampling of at least 10 males allows studies on Y chromosome variation, whereas samples of poultry species preferably should contain at least 10 female birds for eventually studying W-chromosomal variation (despite the fact that study of the W-chromosome is still problematic due to long repeated sequences and other structural obstacles).

3.1.6 Know the rules

Collection and exchange of DNA involves the transfer of potentially valuable genetic material and information, including intellectual property. Therefore, it is recommended that all collaborating parties prepare a Material Transfer Agreement (MTA) that spells out the terms of the exchange and the limits regarding the use of the genetic material during or after the study. When preparing the MTA, the provider of the genetic material must be informed about extent to which they can grant rights to the receiver for use of the material, again by consulting with appropriate national agencies. In any case, the project is carried out in close collaboration with the providers of the genetic material, who during all stages remain the proprietors of the genetic material. An example MTA is provided in Appendix 2.

Collaboration on a project for molecular characterization of AnGR may also involve the shipping or exchange of tissue or DNA across international borders. Therefore, this exchange of genetic material may be subject to legal and/or sanitary regulations. Many countries require particular permits for the importation of biological materials. In addition, some countries have regulations involving the export of genetic resources in order to prevent biopiracy. The scientists involved in the study should make themselves aware of these regulations by contacting the appropriate ministries of each country, as well informing themselves with regard to national obligations under the Convention on Biological Diversity. Both parties involved in the transaction should ensure that the exchange complies with the laws and regulations of both the exporting and importing countries with regard to the shipping of the samples.

3.1.7 Design the sampling

As noted earlier, molecular characterization will ideally be done in concert with phenotypic characterization and evaluation of the production system, according to *FAO Guidelines for Phenotypic Characterization of Animal Genetic Resources.* Even if molecular characterization is done independently, the following factors should be taken into account to ensure a genetically diverse sample:

- Consider the structure of the production system, geographic locations and pedigree relationships (even if these latter data are not formally recorded).
- Preferably sample in the production area that is closest to the site of the development of the breed.
- Preferably cover the different agroclimatic zones where the breed is found.
- Typically no more than 10% of any one herd or village population should be sampled and in any case no more than five animals should sampled from any herd.
- Do not sample animals with common grandparents.
- If there are indications of genetic subdivision, specify what has been sampled and try to cover the different types within the breeds.
- For studies on mtDNA and Y-chromosomal markers, sampling of animals with common maternal and paternal origin, respectively, should be avoided.

- In situations where suitable breeding infrastructure exists, sampling can be done in conjunction with breeding associations or artificial insemination organizations, which may also be reliable sources of pedigree information.

3.1.8 Choose the genetic marker

3.1.8.1 General considerations. In principle, the most advanced technologies should be chosen that are available for the species to be studied, because these technologies are generally the most informative. If the research objectives can be accomplished by an older technique, this approach can be justifiable (1) because of financial constraints or (2) if the results of the study have to be compared with those of other breeds for which only data from the older technique are available, and data with the advanced technology are not expected to be produced in the near future. Local technological expertise can be a consideration, but we stress that new expertise can be acquired and outsourcing is often a viable option.

The merits of the current marker systems are discussed below. We do not advise for diversity studies the use of formerly used marker categories such as major histocompatibility complex (MHC) or other protein polymorphisms or less-specific DNA markers such as RAPD, AFLP, etc., unless their use is essential for answering a specific question.

3.1.8.2 Microsatellites. So far, most studies (Groeneveld et al., 2010) have analyzed highly polymorphic microsatellite markers, which are repeated sequences of 1-6 base pairs. Variability is in terms of the number of repeated sequences observed. Microsatellites do not encode proteins and are thus assumed to be selectively neutral. Microsatellites are now available for most livestock species and have proven their value for studying variation within and across breeds. The most critical decision is the choice of the marker panel, for which the following should be considered:

The FAO and the ISAG/FAO Advisory Group on Animal Genetic Diversity, have proposed panels of 30 microsatellite markers for nine major livestock species (Appendix 7, www.globaldiv.eu/docs/Microsatellite%20markers.pdf). The many existing datasets from completed characterization studies that have used FAO markers allow new data to be compared with more breeds than for any other microsatellite panel, particularly for cattle, sheep and goats.

Ideally, all 30 markers should be used for characterization of populations. This approach not only yields more accurate data than with a subset of the markers, but also offers more opportunity to compare with results from previous studies undertaken with various subsets of the 30 markers.

As shown by too many examples, the use of markers other than the ISAG/FAO panels restricts the utility of the study to the regional or national level and obstructs efforts to obtain a global view of animal genetic diversity.

3.1.8.3 Single-nucleotide polymorphisms (SNPs). For cattle, sheep, chicken, pigs and horses, large panels of SNP have been developed for selection, parentage validation and other purposes and are being used for comparing breeds. The SNP have the following advantages relative to microsatellites.

- Automatic allele scoring is unambiguous and facilitates combination with other datasets.
- The high number of SNPs can allow a description of individual and breed relationships with unprecedented accuracy.
- Measurement of linkage disequilibrium reveals coalescence times and effective population sizes.
- Markers with high differentiation across breeds can reveal functional, as well as neutral, genetic variation (Kohn et al., 2006), which may lead to the identification of gene variants corresponding to specific phenotypes.
- When large panels of SNP are evaluated, the cost of genotyping on a per marker basis is much less than with microsatellites.

A caveat in these studies is the ascertainment bias, which occurs because the majority of SNP in most commercially available panels originate from the cosmopolitan highly productive breeds. As a consequence, diversity in other breeds, including those located close to the domestication centres, can be underestimated. Development of standard SNP panels for diversity studies that are not biased by ascertainment protocols would require a new SNP discovery effort in a larger group of breeds covering most of the existing diversity within a species.

3.1.8.4 Copy number variations (CNVs). Genetic studies of the human genome indicate the presence of variation in copy number of certain chromosomal segments, as well as a relationship between copy number and phenotypic variation. It is anticipated that this category of genetic variation will also prove to be relevant for studying the diversity of livestock.

3.1.8.5 Genome sequencing. New technologies of personal genomics, several of which have already passed the proof-of-principle stage, will expand further the scope of molecular studies and likely allow in the near future the affordable whole-genome sequencing of individual animals. Predictably, this will open new avenues of research that lead to new insights into diversity and the estimation of conservation values. Most notably, dense genetic maps allow the demarcation of "footprints" or "signatures" of selection, while the growing amount of knowledge on genotype-phenotype relationships will also reveal novel aspects of functional diversity. Clearly, this will ask for new software and hardware for extracting and storing meaningful information for the huge amount of DNA sequence. The building of a bioinformatic infrastructure will thus be a potentially limiting factor in the exploitation of this variation. .

3.1.8.6 Mitochondrial DNA (mtDNA) markers. These maternal markers have been instrumental in identification of wild ancestors, localization of domestication centres and reconstruction of colonization and trading routes (Bruford et al., 2003; Groeneveld et al., 2010; Ajmone-Marsan et al., 2010). Most studies with mtDNA target the hypervariable control region (D-loop), but complete mtDNA sequences add substantial information by establishing the relation between haplogroups (Achilli et al., 2008). A caveat is the artefactual amplification of nuclear copies of mtDNA, which can be minimized by using long range PCR amplifications and homologous primers that are complementary to their target regions without mismatches.

3.1.8.7 Y-chromosomal markers. Y-chromosomal variation is a powerful tool to trace gene flow by male introgression (Petit et al, 2002). It is the most powerful marker in human population genetics and is used more and more in domestic animal species.

3.2 In the field

3.2.1 Numbers that count

For reliable estimation of allele frequencies, at least 25 animals per breed should be typed, but at least 40 animals should be sampled to allow for possible losses, mistypings, missing values and genetic subdivision within breeds or various degrees of crossbreeding. If there is population subdivision, different subtypes or agroclimatic zones, sampling a larger amount of animals is recommended. On remote sites, remember that what is not collected is lost.

3.2.2 Collect samples

For this most crucial step, the following considerations are relevant:

- Almost all cells or tissues may be used for DNA analysis: blood, semen, hide, bone, tissue (e.g. ear tissue), plucked hair (only the root cells contain nuclei, but cut hairs can be used for mtDNA analysis) and feathers,
- High quality DNA is most easily obtained from samples of peripheral blood, organs or other tissues. Most convenient are blood samples, to be collected in an anti-coagulant (EDTA or Na-citrate). A protocol for blood collection is provided in Appendix 3.

- Collect enough material for present and future studies. For PCR-based applications, 10 ml of blood is adequate, but for high-density SNP typing and genomic sequencing, it is advisable to sample 50 ml or more. Note that poultry species have enucleated erythrocytes and, therefore, much less blood (~1 ml) is required.
- Transport of blood samples can be at ambient temperatures, but in tropical regions samples should be processed within 36 hours.
- For longer storage, samples can be placed in a room temperature preservative such as Queen's buffer (0.01. M Tris/HCL , 0.01 M NaCl, 0.01 M EDTA and 1 % n-laurosylsarcosine, pH 8.0, Seutin et al., 1991).
- Tissue samples of 1 cm squared should be minced to 1 mm squared pieces and placed in Queen’s buffer or 70% ethanol. Air-drying of ethanol-treated samples allows long term storage and the easy transport of samples. Alternatively, pieces of tissue may be directly dehydrated by placing in vials on crystals of silica gel.
- Hair samples should be desiccated as soon as possible and stored dry.
- FTA cards can be used for collection of genetic material with DNA to be amplified by PCR, but special protocols are required to obtain double-stranded DNA and the single-stranded DNA obtained with standard isolation protocols is not suitable for all other applications.
- Samples that are to be used for cloning, Southern blotting or genomic sequencing protocols require double-stranded DNA of high molecular weight.
- From each animal, duplicate samples should be taken and kept separate during subsequent transport and storage.
- Labelling of samples should be unambiguous and permanent. The labelling procedures should be developed and supervised by the responsible scientist of the project.
- Bank it: store all samples and document all relevant information unambiguously in such a way that they can be retrieved and understood, even by persons not involved in the sampling.

3.2.3 Also collect data

Essential is recording of the following information for each sample:

- Sample (and duplicate) number
- Date
- Location and GPS coordinates,
- Name of collector
- Breed
- Sex of animal
- Type of sample
- Any relevant phenotype
- Basic pedigree information
- Size of herd
- Digital photograph with measuring stick, showing any interesting morphological features
- Notes about any recent change in geographic location of the animal

An example of a sampling form is in *Appendix 4*.

As additional information, the form in Appendix 5 should be compiled once for each breed, to the extent that is possible based on the information available. This form addresses breed origins, farming practices, basic production information, and features of the breed such as productivity, disease resistance or adaptation to local conditions should also be recorded and follows the breed data sheet in DAD-IS.

3.3 In the laboratory

3.3.1 Extracting DNA

Several reliable protocols for DNA extraction are available. Older protocols are based on Proteinase K/SDS lysis of cells, organic extraction and alcohol precipitation. Salt precipitation avoids the organic solvents, but the long-term stability of the DNA samples is problematic. Now, convenient commercial kits based in the specific binding of DNA to resins are available for several kinds of tissues and generally perform well. It is recommended to test out before application to field samples any DNA extraction procedures that have not been used routinely. Consider the amount of DNA required by the different protocols.

3.3.2 Genotyping.

Protocols for genotyping are generally available and straightforward, but the following factors should be taken into account:

- Analyze in each experiment at least one reference sample in order to cross-validate successive genotyping experiments and include this in all experiments.
- For microsatellites, use the FAO recommended panel and include international reference samples in order to link your data with other datasets.
- Include blank extraction and amplification samples in order to check contamination of the reagents.
- Outsourcing the genotyping to dedicated custom service laboratories may very well ensure high quality and cost-effective results without requiring investment in new equipment and expertise. Outsourcing does not, however, lift the requirement of analyzing reference samples and critically checking the quality of the data.
- In collaborative projects with microsatellites, it is preferred that one laboratory performs all typings for a given marker in order to exclude laboratory-dependent scoring. If this is not feasible, it is most essential to share samples of reference animals in order to be able to standardize allele sizes.
- Multiplexing the PCR can reduce the costs, but results should be checked carefully to ensure it does not increase the percentage of missing genotypes. This applies especially to samples with low DNA concentration. As a compromise, PCR reactions can be carried out separately and be combined on the gel (multiloading).

3.4 Data analysis

3.4.1 Check the data

Remove uncertain scores and delete markers and animals with an excess of missing data. Also check for outliers. Be aware that erroneous genotypes may distort the results of the analysis. The following checks should be carried out in order to minimize the error rate (Pompanon et al., 2005):

- Matching duplicate samples, indicating errors during sampling or processing of samples
- Examining unusual alleles, which may result from clerical mistakes or incorrect interpretation of electrophoretic patterns.
- Check for an excess of apparent homozygosity in samples with low DNA concentration because of allele dropout.
- Comparing allele frequencies with data from breeds that are likely to share the most frequent alleles in order to detect inconsistent allele sizing.

- Check for absence of laboratory-dependent clustering of breeds, which may result from systematic differences in allele calling. One cause of laboratory-dependence may be the lab-dependent differentiation of microsatellite alleles that only differ by one bp in length.
- Linkage disequilibrium (LD). Markers in LD in all populations are probably genetically linked.
- Hardy-Weinberg (HW) equilibrium. Markers that in most breeds are not in HW may have null alleles or be linked to loci under selection, hence breaking the assumption of neutrality.

Always keep the original version of the data in which no corrections have been carried out, so data can be recovered if deleted in error.

3.4.2 Crunch the numbers

Choose the appropriate method for data analysis and the software to implement it.

3.4.2.1 General considerations. Considering the continual advances in computational population-genetic analysis (Labate, 2000; Excoffier and Heckel, 2006) with a clear shift towards Bayesian model-based approaches (Beaumont and Rannala, 2004; Chikhi and Bruford, 2005), no single strategy of data analysis can be recommended. Nevertheless, some basic analyses that are undertaken in nearly all studies of molecular diversity are described below. Although many algorithms are highly complex, it is most desirable that the experimental geneticist understands if, or why not, a specific approach is suitable for the dataset. In addition, output of the computer programs should never be accepted blindly, but should rather be evaluated critically for validation. The effect of changing program parameters should be tested. Simulated datasets can be used to verify the validity of the algorithms. Results of the data analysis should be translated to biological, genetic and/or historic phenomena, mechanisms or events and, whenever possible and appropriate, utilized in the development recommendations for management of AnGR.

3.4.2.2 Within-breed analysis. Expected heterozygosity or allelic richness within breeds indicates the influence of drift on breed diversity, where decreased heterozygosity is associated with increased drift. Differences between expected and observed heterozygosity as well as departure from Hardy-Weinberg equilibrium indicate non-random mating or the existence of population substructures. The presence of inbreeding can be tested by F statistics (Weir and Cockerham, 1984), in particular by testing if the F_{IS} parameter is significantly larger than zero. Genetic subdivision may be explored further by model-based clustering (Pritchard et al., 2000). Depending on the results, it may be decided to exclude from further analysis markers not satisfying the expectation of neutrality or outlier individuals (e.g. unrecorded crossbreds).

If introgression from other breeds is suspected, the degree of introgression can be estimated by special programs for this purpose (Chikhi and Bruford, 2005). Introgression can be confirmed by mtDNA and/or Y-chromosomal analysis of the parent and introgressed breeds. The availability of large SNP datasets may lead to the identification of SNP alleles or haplotypes that are diagnostic for a given breed and then also for introgression of genes from this breed.

Another parameter relevant for diversity is the effective population size (Chikhi and Bruford, 2005). Sequence data are suitable for the mismatch analysis, which may reveal recent population expansions (Schneider and Excoffier, 1999). Coalescence analysis of haplotypes within or across populations estimates the age of the most recent common ancestor (MRCA, Drummond et al., 2005).

3.4.2.3 Analysis of breed relationships. Breed formation has led to a partitioning of the total diversity in a within-breed and an among-breed component. These components and others, e.g. the component due to the geographic location of breeds, can be quantified by

AMOVA analysis (Excoffier et al., 1992) and reflect history and breeding practices. Typically, 50 to 90% of the total diversity corresponds to the within-breed component, depending upon the group of breeds sampled and the sources of variability considered.

As an essential step in data analysis, we recommend the calculation of genetic distances among breeds based on gene frequency data, followed by visualization of relationships in trees, networks (Huson and Bryant, 2006) or spatial plots.

There is no wide consensus on the relative merits of the many estimators of genetic distance for analyzing within-species populations such as domestic animal breeds. Nei's (1972) standard genetic distance D_s has been used most commonly in studies of natural populations in evolutionary genetics and has the advantage that it is linear in time (Takezaki and Nei, 1996). Distance measures based on Wright's F_{ST} statistic, which may be estimated via the D_R distance (e.g. Reynolds et al., 1983; Laval et al., 2002) may be more appropriate for short-term evolution such as the divergence between livestock breeds. Another popular distance is the Nei D_A distance (Laval et al., 2002). Although the correlations among various distance measures have been found to be generally very high (Chakraborty and Tateno, 1976), it should be checked if essential conclusions depend critically on the choice of the genetic distance measure used. However, genetic distances within species as estimated by neutral markers cannot adequately measure divergence times, because distance measures also depend on population size and become exaggerated if genetic drift in small populations or inbreeding distorts the allele frequencies.

The distance estimates are then used in a clustering analysis involving methods that allow for unequal rates of evolution, such as the neighbor-joining method (Saitou and Nei, 1987; Takezaki and Nei, 1996) or NeighborNet graphs (Bryant and Moulton, 2004). Another widely used analysis method is the construction of coordination plots, either on the basis of the genotypes (principal component analysis), on the basis of genetic distances (principal coordinate analysis) or by multidimensional scaling, decomposing the diversity patterns in different clines that each may represent a different geographic domain and/or historic event.

Different algorithms of spatial analysis (Chikhi and Bruford, 2005) allow the explicit introduction of geographic information and the testing of phylogeographic hypotheses.

For analysis of mtDNA and Y-chromosomal haplogroups, which do not recombine, reduced median networks (Bandelt et al., 1995) offer insights in the dispersal of molecular variants via maternal and paternal transmission, respectively. These analyses often permit direct inferences of migration, introgression and expansion events.

Unsupervised model-based clustering by the *Structure* software (Pritchard et al., 2000) and other programs operates on individual genotypes and uses a Bayesian approach. Depending on the preset number of clusters, the program may identify clusters of related breeds, clusters of individuals of the same breed or clusters that correspond to subpopulations within breeds. For each individual the proportion of the genome derived from the inferred clusters are calculated, which may reveal qualitatively introgression events. Prior information on ancestral populations can be introduced in the dataset and be used for supervised clustering.

This list of analysis options is in no way exhaustive. The choice of the programs for data analysis depends on the data and the hypotheses to be tested (see Appendix 6). Much of the software needed to perform the analyses listed above can be obtained on the internet. In general, it is free of charge, but proper citation in publications is expected.

3.4.3 What are the data trying to tell us? Place the results in a historic perspective. Keep in mind that genetic events – migration, introgression, admixture, crossbreeding, population bottlenecks, and selection – have happened at different times, which may complicate the pattern of diversity. Consider alternative explanations and do not interpret according to preconceived ideas.

3.4.4 Conservation priorities. Several theoretical approaches to conservation have been published, such as the Weitzman (1992) approach based on genetic distances or minimizing marker-estimated kinships (Eding & Meuwissen, 2001; Caballero & Toro, 2002) . However, the usefulness of the currently available algorithms is still a matter of debate (European Cattle Genetic Diversity Consortium 2006, Toro et al., 2009). The use of genetic marker information in prioritization of breeds for conservation has been reviewed by Boettcher et al. (2010) and is discussed in the *FAO Guidelines for In Vivo Conservation of Animal Genetic Resources.*

3.5 Publish it

3.5.1 Let the world know

Publish your findings in a scientific journal. Open-access journals are recommended because of their wide diffusion and free accessibility.

3.5.2 Share the credit

Properly acknowledge contributors of samples and/or data.

3.5.3 Share the data

After publication, deposit your data in a public database and/or comply with requests to make datasets available.

3.6 Translate the results

If appropriate, formulate recommendations for genetic management and conservation and disseminate these to breeding organizations and government agencies such as the National Advisory Committee on AnGR, assuming that one exists, as well as the FAO National Coordinator for AnGR. Below are a number of examples that show how molecular observations are relevant for genetic management:

- **Original diversity.** Cattle, sheep and goat breeds from near the domestication sites in Southwest Asia have high neutral genetic diversity, as measured by observed heterozygosity and number of alleles of microsatellite markers. This observation is assumed to indicate that the breeds also have a large amount of functional diversity in genes influencing phenotypic traits. Functional diversity is valuable for achieving present or future breeding objectives and, therefore, breeds with high functional diversity (i.e., as measured by neutral genetic markers as a proxy) should receive a high priority for conservation.
- **Unique origin.** DNA analysis of Indonesian zebu breeds revealed a large contribution of the exotic banteng species to both the autosomal DNA and maternal lineages (Mohamad et al., 2009), which very well may have contributed to the adaptation to local conditions. Breeds with a unique species origin may receive a greater priority for conservation, especially if the species of unique origin is now extinct.
- **Separate history.** Both mtDNA and microsatellites indicate that the Italian Chianina cattle have an ancient origin and evolved separately from other Podolian breeds. DNA information was critical in making this discovery. Similar results have been found for other breeds with a long history of separation from other breeds (e.g., Soay sheep, Jersey cattle). Separate evolution may favour the presence of special combinations of genes and alleles and thus indicate a uniqueness that should be preserved.
- **Crossbreeding.** In contrast, Russian Black-and-white cattle is one of several livestock breeds with a documented recent history of crossbreeding, in this case with the Holstein-Friesian. Microsatellite data indicate that the Russian breed is now very similar to the

Holstein. Although one cannot exclude the possibility that some of the original germplasm of the Russian Black-and-white breed has been retained in the current population, its value for conservation has clearly been reduced. The amount of original and exotic germplasm in a breed subject to crossing will depend on the timing and amount of crossing undertaken and the diversity in production environments between the breeds involved.

- **Consanguinity.** Kinships can be estimated from molecular data. This is particularly valuable when pedigree data are missing or unreliable. In these case the mating of animals having very similar genotypes, and therefore high kinship, should be avoided. Kinship can also be minimized among animals selected for a conservation programme, in order to maximize genetic variability within the conserved population.
- **Inbreeding depression.** One of the most inbred horse breeds is the Friesian horses. A high incidence of retained placenta is one of the several inbreeding problems in this breed, for which crossbreeding ("new blood") may be the only solution. Molecular analysis can quantify the degree of inbreeding and help to identify which breed is related to the inbred breed and might thus be suitable for crossbreeding.
- **Unique phenotype.** Many highly developed breeds with unique phenotypes e.g. Scottish Highland cattle, several British sheep breeds, Tauernschecken goats) tend to have low degrees of molecular diversity as detected by a panel of microsatellites (Wiener et al., 2004; Cañón et al., 2006; Peter et al. 2007;). This result emphasizes that phenotypic and molecular diversity should be considered as separate and complementary criteria for conservation. A breed with a unique phenotype may be of value for conservation, even if molecular diversity is small – or vice versa.
- **Carriers of genetic defects.** Many genetic defects can be diagnosed by associated molecular markers. These can be used to screen the population and identify carriers of the undesirable alleles. The use of carriers for mating should be preferably avoided when possible, especially with each other.
- **Functional variation.** For several livestock species, gene variants have been found that control coat color, a trait that is usually relevant for the identity of a breed. Genome sequencing is likely to reveal sequence variants that are also involved in phenotypic variation, for instance by changing the coding sequence of expressed genes, gene deletion, altering regulatory sites, etc. An inventory of such mutations per breed may be weighted in conservation decisions.
- **Complementary concepts.** A logical strategy is to conserve both functional (based on known gene effects) and neutral variation for yet unknown effects.

In general, genetic diversity measured with neutral markers should not be the sole criterion to conservation. Obviously, decisions on choice of breeds should also take into account traits of economic value, specific adaptive features, distinct phenotypes, role of breed in local production systems, and availability of resources and infrastructure in the region where a breed is located.

3.7 International coordination

3.7.1 Bank the data

It is recommend that after publication, information on breed samples, allele frequencies and additional collected information be incorporated into a central and eventually into national data banks. These data banks should have open access to all investigators interested in livestock. Investigators should be encouraged to contribute in a standard format to the data bank.

The Joint FAO/IAEA Division on Nuclear Techniques in Food and Agriculture has constructed a web-linked database for sharing of molecular genetic data for several species (www.globalgenomic.com). Registered users can upload their data as well as browse and download the data from other studies of different breeds from around the world.

3.7.2 Bank the samples

Only a small fraction of the biological and DNA samples from each animal will be used for the molecular assays proposed here for the estimation of genetic parameters. The remaining DNA will be a valuable resource for future use.

This DNA should therefore be preserved in an Animal Genetic Resources DNA Bank. As insurance against loss, the DNA from each animal should be kept as duplicate samples in different sites, and two laboratories should provide the necessary maintenance and storage facilities, and, while respecting rules in the source country and conditions imposed by the original owners of the animals, to make DNA samples available to other laboratories working on the molecular genetics of domestic animals.

Furthermore, a set of reference samples should be used for each species. The ISAG/FAO Advisory Group for Animal Genetic Diversity coordinates this activity. Scientists that would like to obtain reference samples should contact the responsible officer at FAO (DAD-IS@fao.org).

3.7.3 Develop appropriate panels of SNPs

The development of SNP panels specifically designed for characterization of the diversity of livestock species is proposed. These panels would likely decrease the cost of genotyping and increase the ability to standardize and share results across laboratories and countries. Preliminary research is required to develop these panels, however, to ensure that they are free of ascertainment bias and accurately estimate genetic variation for all populations.

4. CONCLUSIONS

The FAO and the ISAG/FAO Advisory Group on Animal Genetic Diversity recommend:

1. That current activities to genetically characterize the genome and establish the genetic relationships among the breeds of each domestic animal species should be continued and completed as a matter of urgency and should be complemented with phenotypic characterization.

2. That the above recommendations are taken into account during setup and execution of studies of the diversity of animal genetic resources, while monitoring closely the advances in molecular technology and bioinformatics.

3. That particular attention should be given to standardization of results from existing and planned studies for integration into a global analysis of AnGR diversity.

4. That breeds from “white spots” on the current phylogeographic map and samples relevant for joining datasets are analyzed.

5. That new frameworks for international cooperation are established to create and distribute reference samples of DNA for standardization and to develop a centralized database to store and provide access to data.

6. That FAO National Coordinators any National Advisory Committees for AnGR and are made aware of all diversity projects at any geographic level, so that results can contribute to the planning and development of national conservation and sustainable use activities and so that FAO can help facilitate coordination among projects, exchange information and promote funding.

7. That new genomic tools for characterization of diversity are to be implemented that avoid ascertainment bias and that methods are developed for combining datasets generated by established and new technologies, respectively.

REFERENCES

Achilli, A., Olivieri, A., Pellecchia, M., Uboldi, C., Colli, L., Al-Zahery, N., Accetturo, M., Pala, M., Kashani, ,B.H., Perego, U.A., Battaglia, V., Fornarino, S., Kalamati, J., Houshmand, M., Negrini, R., Semino, O., Richards, M., Macaulay, V., Ferretti, L., Bandelt, H.J., Ajmone-Marsan, P. & Torroni, A. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology,* 18: R157-R158.

Ajmone-Marsan, P., Garcia, J.F. & Lenstra, J.A. 2010. On the origin of cattle: how aurochs became domestic and colonized the world. *Evolutionary Anthropology*, in press.

Bandelt. H-J., Forster, P., Sykes, B.C. & Richards, M.B. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141: 743–753.

Beaumont, M.A. & Rannala, B. 2004. The Bayesian revolution in genetics. *Nature Reviews Genetics* 5: 251-261.

Bensch, S. & Akesson, M. 2005) Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, 14: 2899-2914

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., Joost, S., Garcia, D., Colli, L., Ajmone-Marsan, P. and GLOBALDIV Consortium. 2010. Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics*, 41 (Suppl 1): 64-77.

Bruford, M.W., Bradley, D.G. & Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4: 900-910.

Bryant, D. & Moulton, V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21: 255-265.

Caballero, A. & Toro, M. A. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3: 289-299.

Cañón, J., García, D., García-Atance, M.A., Obexer-Ruff, G., Lenstra, J.A., Ajmone-Marsan, P., Dunner, S. & the ECONOGENE Consortium. 2006. Geographical partitioning of goat diversity in Europe and the Middle East. *Animal Genetics,* 37: 327-334.

Chakraborty, R. & Tateno, Y. 1976. Correlations between some measures of genetic distance. *Evolution,* 30: 851-853.

Chikhi, L. & Bruford, M.W. 2005. Mammalian population genetics and genomics in: *Mammalian Genomics (ed. Ruvinsky A, Marshall Graves J). CABI Publisher* Ch. 21, pp 539-583.

Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22: 1185-1192.

Eding J.H. & Meuwissen T.H.E., 2001. Marker based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics*, 118: 141–159.

European Genetic Cattle Diversity Consortium 2006. Marker-assisted conservation of European cattle breeds: an evaluation. *Animal Genetics*, 37: 475-481.

Excoffier, L., Smouse, P.E. & Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131: 479-491.

Excoffier, L. & Heckel, G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, 7: 745-758.

FAO. 1990a. *Animal genetic resources. A decade of progress, 1980-1990.* by J. Hodges. Animal Production and Health Paper. Rome.

FAO. 1990b. *Manual on establishment and operation of animal gene banks.* by J. Hodges. Animal Production and Health Paper. Rome.

FAO. 1990c. *Review of regional animal gene banks and recommendations from Hannover Workshop on associated topics raised by the Tenth Committee on Agriculture.* by J. Hodges. Animal Production and Health Paper 80: 51-57.

FAO. 1992. *The management of global animal genetic resources.* Animal Production and Health Paper 104. Rome.

FAO. 1993. *Secondary Guidelines: Measurement of Domestic Animal Diversity (MoDAD).* Rome, Italy.

FAO. 2007a. *Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration.* Rome. (available at http://www.fao.org/docrep/010/a1404e/a1404e00.htm).

FAO. 2007b. *The State of the World's Animal Genetic Resources for Food and Agriculture,* edited by B. Rischkowsky, B. & D. Pilling. Rome (available at http://www.fao.org/docrep/010/a1250e/a1250e00.htm).

Groeneveld, L.F., Lenstra, J.A., Eding. H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Jianlin, H., Finlay, E.K., Groeneveld, E. , Weigend, S. & the GlobalDiv Consortium. Genetic diversity in livestock breeds. 2010 *Animal Genetics*, 41 (suppl. 1): 6-31

Hammond, K. & Leitch, H. 1996. The FAO global programme for the management of farm animal genetic resources. In: *Beltsville Symposia in Agricultural Research. XX. Biotechnology's Role in the Genetic Improvement of Farm Animals*. (Eds. R. Miller, V. Pursel & H. Norman) American Society of Animal Science, Illinois, 24-42.

Hodges, J. 2006. Conservation of Genes and Culture: Historical and Contemporary Issues. *Poultry Science*, 85: 200–209.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies *Molecular Biology and Evolution*, 23: 254-267.

Keller, L.F. and Waller, D.M. 2002. Inbreeding effects in wild populations. *Trends in Ecology and Evolution* 17: 230-241.

Köhler-Rollefson, I., Rathore, H.S. & Mathias, E. 2009. Local breeds, livelihoods and livestock keepers' rights in South Asia. *Tropical Animal Health and Production*, 41: 1061–1070.

Kohn, M.H., Murphy, W.J., Ostrander, E.A. & Wayne, R.K. 2006. Genomics and conservation genetics. *Trends in Ecology and Evolution*, 21: 629-637.

Labate, J.A. 2000. Software for population genetic analyses of molecular marker data. *Crop Science*, 40: 1521-1528.

Laval, G., SanCristobal, M. & Chevalet, C. 2002. Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genetics Selection Evolution*, 34: 481-507

Mohamad, K., Olsson, M , Van Tol, H.T.A., Mikko, S., Vlamings, B.H., Andersson, G, Rodríguez-Martínez, H., Purwantara, B., Paling, R.W., Colenbrander, B. and Lenstra, J.A. 2009. On the origin of Indonesian cattle. *PLoS ONE*, 4: e5490

Nei, M. 1972. Genetic distance between populations. *American Naturalist*, 106: 283-292.

Nei, M. 1987. *Molecular Evolutionary Genetics*. New York, Columbia University Press.

Peter, C, Bruford, M., Perez, T., Dalamitra, S., Hewitt, G. Erhardt, G. & the ECONOGENE Consortium 2007. Population structure of 57 European and Middle Eastern marginal sheep breeds. *Animal Genetics*, 38: 37-44.

Petit, E., Balloux, F. & Excoffier, L. 2002. Mammalian population genetics: Why not Y? *Trends in Ecology and Evolution*, 17: 28-33.

Pompanon F, Bonin A, Bellemain E & Taberlet P. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, 6: 847-859.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959

Rege, J.E.O. & Gibson, J.P. 2003. Animal genetic resources and economicdevelopment: issues in relation to economic valuation. *Ecological Economics* 45: 319–330.**Reynolds, J., Weir, B.S. & Cockerham, C.C.** 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105: 767-779.

Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, 152: 1079-1089.

Saitou, N. & Nei, M. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406-425.

Seutin, G., White, B.N. & Boag, D.T. 1991. Preservation of avian blood and tissue samples for DNA analysis. *Canadian Journal of Zoology*, 69: 82-90.

Taberlet, P., Valentini, A., Rezaei, H.R., Naderi, S., Pompanon, F., Negrini, R. & Ajmone-Marsan, P. 2008. Are cattle, sheep, and goats endangered species? *Molecular Ecology*, 17: 275-84.

Takezaki, N. & Nei, M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144: 389-399

Toro, M.A., Fernández, J. & Caballero, A. 2009. Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120: 174-195.

Weir, B.S. & Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution,* 38: 1358-1370.

Weitzman, M.L. 1992. On diversity. *Quarterly Journal of Economics*, 107: 363–405.

Wiener, P., Burton, D. and Williams, J.L. 2004. Breed relationships and definition in British cattle: a genetic analysis. *Heredity,* 93: 597–602.

APPENDIX 1: GLOSSARY OF TECHNICAL TERMS

Ascertainment bias Systematic distortion in estimates of molecular genetic parameters (such as allelic frequencies) due to irregularities in the process undertaken for the identification of the markers. For instance, many SNP in large panels were selected according to their high minor allele frequency in cosmopolitan breeds and can underestimate the relative diversity in other breeds.

AMOVA Analysis of molecular variance, estimation of the portioning of diversity over different hierarchical levels: within breeds, between breeds within regions, between regions, etc.

Bayesian analysis Estimation of a likelihood distribution of model parameters on the basis of the likelihoods of parameter values in the absence of data (the prior) and the likelihoods of the observed data given different values of the model parameters. These estimations depend on a specific model and are often achieved by a strategy (like the Multiple Chain Monte Carlo simulations, MCMC) to explore different plausible values of the parameters (the ‘parameter space’).

Coalescence analysis Estimation of the divergence times of individual DNA sequences since their descendance from a hypothetical most recent common ancestor (MRCA), often used to infer present and past effective population sizes.

CNV Copy number variation, a type of structural variation in the genome resulting from differences in the copy number of chromosomal fragments of up to several megabases in length. CNV can be used as a genetic marker and has been associated with differences in human phenotypes.

Effective population size (often abbreviated N_e) Hypothetical population size that would generate observed values of diversity parameters for a given population if mated randomly and not subject to forces such as selection and migration. The N_e corresponds the number of breeding animals per generation and is usually smaller than the actual population count. It may be calculated separately for males and females.

Genetic distance a measure of the genetic differences between two populations (or species) calculated on the basis of allelic frequencies in both populations.

Genetic marker Sequence of DNA that is variable (polymorphic) within a species. The different variants are called alleles, such microsatellite loci that differ in the number of repeat units or the two different nucleotides of a SNP.

Hardy-Weinberg equilibrium (HWE) Ratio for a given marker and population of the numbers of homozygote and heterozygote genotypes as predicted by random mating in a large population in the absence of selection, migration and mutation.

Introgression Movement of a particular allele or set of alleles from one population (i.e. breed) to another, usually by either deliberate crossbreeding or casual contact between neighbouring populations.

Linkage disequilibrium (LD) Distribution of multilocus genotype combinations in a population for a given pair of markers that is incompatible with independent inheritance, thus indicating genetic linkage of the loci.

Locus A distinct region of DNA (often a gene) in the genome.

Microsatellite Tandem DNA repeat of a 2 to 5 bp unit. In most cases, the repeat unit is the dinucleotide CA. The number of repeats of a given microsatellite is often polymorphic within populations, in which case the microsatellite may serve as genetic marker. Also known as STR (simple tandem repeat) or SSR (simple sequence repeat).

Mismatch analysis Calculation of distribution of the number of sequence differences ("mismatches") in pairwise comparisons of individuals from two different populations.

MtDNA Mitochondrial DNA, widely used in phylogenetic studies because of its variability, lack of recombination and maternal inheritance.

Multiplex PCR Carrying out simultaneously in one reaction the amplification of several different loci by using different pairs of primers.

Nucleotide any of the four types of molecules that make up the structural units of DNA (and RNA). For DNA, these molecules are adenine, cytosine, guanine and thymine and are often denoted by their first letter (i.e. A,C, G, and T, respectively).

PCR (polymerase chain reaction) Method for amplifying DNA segments that uses cycles of denaturation, annealing to primers, and polymerase-directed DNA synthesis.

Phylogeny Evolutionary history of a taxonomic group.

Phylogeography Geographic pattern of the genetic variation of a breed or species.

Primer is a strand of nucleic acid that serves as a starting point for DNA synthesis in PCR. The sequence of the primer is complementary to the sequence of the initial portion of the DNA to be synthesized.

Principal component analysis (PCA), principal coordination analysis (PCOORDA) Two different methods for analysis of a set of variables, such as allele frequencies, by calculation of a set of coordinates that each correspond to a weighted combination of the variables in such a way that each coordinate displays as much as possible variation (optimal variance). Visualizing the distribution of individuals or breeds in a plot of the first two or three coordinates then generates a figure in which the distances between the points reflect the differences according to the dataset and visualizes at least partially the pattern of diversity. The coordinates potentially correspond to phylogeographic clines.

SNP Single nucleotide polymorphism, resulting from a point mutation and most often corresponding to a biallelic (having two different alleles) marker

Structural variation DNA sequence variation based on copy number variations (CNVs: deletions, duplications and largescale copy number variants) and on insertions, inversions and translocations.

APPENDIX 2: EXAMPLE[1] MATERIAL TRANSFER AGREEMENT

MATERIAL TRANSFER AGREEMENT (MTA) for genetic material for genotyping

This Material Transfer Agreement is made by and between,

____

Name of provider of genetic material ("Provider")

Mailing Address ____
Other contact information – i.e. telephone and fax numbers, email address

____

and

____

Name of recipient of genetic material ("Recipient").

Mailing Address ____
Other contact information – i.e. telephone and fax numbers, email address

____

The parties have agreed as follows:

1. Provider agrees to transfer to Recipient the following (biological) material ("Material"):

Description of the genetic material including type (e.g. DNA, blood, tissue) amount (i.e. number of samples) and other information (e.g. means of preservation)

2. This Material will be used by Recipient solely in connection with the project described as follows:

Description of the project, including assays to be performed (e.g. molecular characterization using the FAO/ISAG panel of microsatellites), use of the data, context in a larger project and project sponsors - ("Research Project").

3. This Material will only be used for research purposes by the Recipient in its laboratory. By requesting the material and signing this agreement, the Recipient is considered responsible for appropriate handling of the material and guarantees that suitable containment conditions are available and will be applied in the Recipient's laboratory. This Material will not be used for commercial purposes, such as production or sale of products or services, for which a commercialization license may be required. Recipient will promptly, after termination of the Research Project, inform Provider of the results of the Research Project.

4. To the extent permitted by law, Recipient agrees to treat in confidence, for a period of *XXXX* years from the date of its disclosure, any of the Provider's written information about this Material that is stamped "CONFIDENTIAL" (hereinafter "Confidential Information"), except for information that was previously known to Recipient or that is or becomes publicly available through no fault of Recipient or which is lawfully disclosed to Recipient without a confidentiality obligation or that is independently developed by Recipient or its

---

[1] Example kindly provided by the International Livestock Research Institute, Kenya, Nairobi.

affiliated companies without the benefit of any disclosure by Provider. Recipient may publish or otherwise publicly disclose the results of the Research Project, provided that in all such oral presentations or written publications concerning the Research Project, Recipient will acknowledge Provider's contribution of this Material unless otherwise requested by Provider.

5. This Material is considered proprietary to Provider. Recipient therefore agrees to retain control over this Material, and further agrees not to transfer the Material to other people not under its supervision without prior written approval of Provider. Provider reserves the right to distribute the Material to others and to use it for its own purposes. When the Research Project is completed, the Material will be disposed of as mutually agreed upon by Provider and Recipient.

6. This Material IS BEING SUPPLIED TO RECIPIENT WITH NO WARRANTIES, EXPRESS OR IMPLIED, INCLUDING ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Provider makes no representations that the use of the Material will not infringe any patent or proprietary rights of third parties. Recipient agrees to hold harmless and indemnify Provider for all liabilities, demands, damages, expenses and losses arising out of or as a result of Recipient's use of the Material for any purpose.

7. Nothing in this Material Transfer Agreement shall or may be construed as granting Recipient any right or license to the Material for any use other or further than the evaluation described here above.

8. This Agreement shall be governed and construed in accordance with the laws of *(the country where the 'Research Project' was conducted*). All disputes arising out of or in connection with this Agreement shall be settled in first instance by the relevant court of *(the country where the 'Research Project' was conducted)*.

RECIPIENT	**PROVIDER**
Place:	Place:
Date:	Date:
By:	By:
Title:	Title:

APPENDIX 3: PROTOCOL FOR SAMPLING OF BLOOD FOR DNA[2]

Description

This protocol describes collection of blood from the jugular vein into a vacutainer EDTA containing tubes (3 per animal). The protocol applies to medium to large species. Please, check national legislation to confirm that you are allowed to collect and transport the samples.

Personnel

Experienced sampler (e.g. a vet), and ideally at least two assistants (one to hold animal and second to assist with blood sampling)

Equipment

Scissors for cutting hair or wool

Protective clothing

Marker pen

Disposable gloves, possibly shoe covers

Consumable Materials

For 2 animals, including sufficient extra reserves in parentheses

70% Alcohol for disinfecting (in a spray bottle, if possible. 95% is NOT suitable)

Cotton or tissue for wiping the area

2 (+4) needles

1 (+1) needle holders

6 (+4) pcs of 10ml vacutainer EDTA containing tubes

In addition

Box to transport and hold the equipment/consumables.

Notebook, a pen and a marker for documentation.

Procedure

- Constrain the animal for sampling. Cut some wool or hair if necessary for finding jugular vein. Spray the area with 70% alcohol and wipe the area dry with tissue.
- Prepare the needle. Take needle and needle holder. Remove the protective cap from the shorter end of the needle. (This is the appendix going into the vacutainer tube.) Without removing the other protective cap, screw the needle to the needle holder (see instructions on the box). Reserve all the vacutainer tubes required for the animal at hand.
- Remove the protective cap from the needle, and jab the needle in the jugular vein (wing vein for poultry). Place the vacutainer tube to its place in the needle holder. Vacuum should draw the blood relatively quickly, if the needle is in the correct place. When tube has enough blood (75% filling is typical) change the tube, without withdrawing the needle from its position in the vein. Gently and repeatedly invert the tube 4-5 times to mix blood with EDTA, and use the marker pen to write the animal ID, breed and date on the tube.
- After you have collected enough blood from the first animal, change the needle and repeat the steps 1-3 with the second animal.
- Remember to collect as much information about the animal as possible. Owner of the animal, Animal ID, breed, sampling site and date, age of the animal, animal origin and

---

[2] Adapted from a manual prepared for the ICARDA/ILRI/IAEA training course on molecular characterization of small ruminant genetic resources of CWANA, 1 – 13 April 2007, ICARDA, Aleppo, Syria

pedigree information as well as known, short description of appearance, major diseases (or lack of them) or other observations by the owner. A digital photograph of side of the animal (with a size scale) would contain much information.

After collecting blood at the field site, all necessary precautions must be taken to avoid exposure to extreme temperatures (heat or cold). Specifically, exposure to direct sunlight or storage in a vehicle in the sun or freezing must be avoided. (If kept without extreme temperatures, samples can keep for about 2 days. Within this period the samples must be taken to the laboratory for centrifugation and eventual DNA extraction.

APPENDIX 4: EXAMPLE QUESTIONNAIRE TO BE FILLED DURING SAMPLING

(Questions in **bold** are generally considered to be mandatory)

Animal code ____________________

Farm ID ____________________

Species code ____________________

AA = Anser anser domesitcus (greylag goose)
AC = Anser cygnoides (swan goose)
AP = Anas platyrhynchos (mallard duck)
BB = Bubalus bubalis (water buffalo)
BF = Bos frontalis (gayal)
BG = Bos grunniens (yak)
BI = Bos indicus (zebu)
BJ = Bos javanicus (banteng, Bali cattle)
BT = Bos taurus (taurine cattle)
CB = Camelus bactrianus (Bactrian camel, two-humped)
CD = Camelus dromedarius (dromedary, one-humped camel)
CH = Capra hircus (goat)
CM = Cairina moschata (muscovy duck)
EA = Equus asinus (donkey)
EC = Equus caballus (horse)
GG = Gallus gallus (chicken)
LG = Lama glama (Llama)
MG = Melea gallopavo (turkey)
OA = Ovis aries (sheep)
SS = Sus Scrofa (pig)
VP = Vicugna pacos (alpaca)
VV = Vicugna vicugna (vicuña)

Species name ____________________

Country ____________________

Number of the sample ____________________

Official Animal ID (if available) ____________________

Animal and sampling information

Sex of animal: female ☐ male ☐

Year of birth of the animal: ____________________ (YYYY)

Place (locality) of birth of the animal: ____

Date of collection: _______________________________________ (DD.MM.YYYY)

Breed's full name: ____

Collector's name: ____

B Collector's institution:____

Address of the farm and telephone number (if available)

Country of the farm: ____

Province/county of the farm: ____

Region of the farm: ____

Closest town to the farm: ____

International phone code: ____________________(4 digits - ex: 0033, 0041, ...)

Area phone code: __________________________

Phone number: ____________________________

Type of biological material: blood ☐ tissue ☐ hair ☐ other (specify) ☐

GPS coordinates _____________________________

APPENDIX 5: BREED QUESTIONNAIRE

(To be completed once per breed – complete all questions that are relevant and for which information is available)

COMMON NAME OF BREED: ____________________

NAME OF SPECIES: ____________________

Transboundary or brand name			
Local breed name			
Main location			
Breed society?	Circle: Yes No Year established:		
Description of origin and development			
Population size	Year: N° of animals:		
N° of reproductive animals	*Males* in natural service		
	Males used for AI		
	Breeding females		
	Trend in breeding females		
Females mated pure (%)			
Adult size (male/female)	Withers height (cm)	M	F
	Live weight (kg)		
N° of farmer/breeders			
Main uses (e.g. meat, milk)			
Typical management conditions	*Type*		
	Housing		
	Feeding		
Conservation activities	In situ: Y / N Ex situ: Y / N Cryo: Y / N		
Performance comparison	*Relative to which breed:*		
much higher in:	*(e.g. milk yield)*		
higher in:			
equal in:			
lower in:			
much lower in:			
Estimated endangerment status (check one)	☐ *Critical (<100 head)* ☐ *Endangered (<1000 head)* ☐ *Vulnerable (<3000 head)* ☐ *Not at risk*		

APPENDIX 6: SOFTWARE FOR GENETIC ANALYSIS

This is a list of recommended genetic analysis programs with a short specification. For a more complete list with more detailed descriptions we refer to Excoffier & Heckel (2006). Most programs can be downloaded freely from the internet, along with detailed instruction manuals. The url are given for convenience and were valid at the time the guidelines were prepared. In the event that the url are no longer valid, the software can likely be found by using the software name in a search engine.

Multipurpose programs

ARLEQUIN. Windows program for calculation of various summary statistics. Requires its own format. Calculates proportion of diversity contained at different hierarchical levels and the significance of the differentiation between populations or groups of populations.
http://cmpg.unibe.ch/software/arlequin3

CONVERT. Easy program that converts data from an Excel format or from the **GENEPOP** format to files suitable for **GDA**, **GENEPOP, ARLEQUIN, POPGENE, MICROSAT, PHYLIP** and **STRUCTURE**.

http://www.agriculture.purdue.edu/fnr/html/faculty/Rhodes/Students%20and%20Staff/glaubitz/software.htm

FSTAT Windows program for various summary statistics. Requires its own format or the GENEPOP format. Calculates the significance of the diffentiation between populations.
http://www2.unil.ch/popgen/softwares/fstat.htm

GENEPOP Windows program for various summary statistics. Requires its own format, which is used by several other programs . Carries out exact tests of HW and LD equilibria.
http://genepop.curtin.edu.au/

GENALEX Estimates of variability based on allele and genotypic frequencies, genetic distances, Principal Component Analysis, Formatting of data for other software. Runs as a Microsoft Excel addin. http://www.anu.edu.au/BoZo/GenAlEx/

GENETIX Ordination method that recovers structure in the data. Estimates linkage disequilibrium using permutation test. Nei's GST, Fis Bootstrapping to test if Fis is different fom zero. Mantel test (i.e. between genetic and geographic distances). Transformation in ARLEQUIN files. (in French) www.genetix.univ-montp2.fr/

MICRO-CHECKER To explore microsatellite data. Estimates basic summary statistics and indicates possibility of null-alleles, mistyped alleles. http://www.microchecker.hull.ac.uk/

MICROSATELLITE TOOLKIT. Convenient Excel microsatellite data handling tool. Format requires sample names in which letters indicate breed and numerals the individual; easily transformed to Structure format. Exports to MICROSAT, ARLEQUIN, GENEPOP and FSTAT formats. Check errors in the dataset (missing figures, large gaps between alleles, non-unique sample labels, duplicate samples), converts two column per marker and one line per sample to one column per marker and two lines per individual). Provides summary statistics (observed and expected heterozygosity, number of alleles) and allele frequencies.
http://www.animalgenomics.ucd.ie/sdepark/ms-toolkit/

POPULATIONS. Windows or Linux program for calculation of several genetic distances and file conversion. Accepts GENEPOP or GENETIX file formats and converts to several formats, including the LEA format. http://bioinformatics.org/~tryphon/populations/

POWERMARKER Comprehensive Windows program for microsatellites, SNPs and other biallelid data with flexible input options. It provides several summary statistics, genetic distances and phylogenetic trees with bootstrapping. http://statgen.ncsu.edu/powermarker/

Genetic distances, trees and plots

BEAST Performs Bayesian MCMC analysis of molecular sequences, inferring rooted, time-measured phylogenies using strict or relaxed molecular clock models. Provides a framework for testing evolutionary hypotheses without conditioning on a single tree topology.. http://beast.bio.ed.ac.uk/Main_Page

DISPAN (Genetic Distance and Phylogenetic Analysis) Calculates various basics population genetics and diversity statistics, genetic distances and constructs phylogenetic trees. http://homes.bio.psu.edu/people/Faculty/Nei/Lab/dispan2.htm

MEGA Calculation of a wide variety of population genetics statistics and convenient tree recontruction program by the most common algorithms except the Baysesian method. http://www.megasoftware.net/

MRBAYES Command-line operated for handling nexus sequence files for Bayesian tree reconstructions. http://mrbayes.csit.fsu.edu/

NETWORK Constructs Median-joining networks of haplotype data. Generates evolutionary trees and networks from genetic and other data. http://www.fluxus-engineering.com/sharenet.htm

PAUP Command-line operated comprehensive package handling nexus files for tree reconstruction according to the most common algorithms (not cost free). http://paup.csit.fsu.edu/

PHYLIP Command-line comprehensive package requiring its own file format for tree reconstruction according to the most common algorithms, but offering fewer options than PAUP. http://phylip.com

SPLITSTREE. Constructs neighbor-joining tree, SPLITSTREE graphs and NeighborNet graps. Aceepts nexus files. Many graphical output options. http://www-ab.informatik.uni-tuebingen.de/software/splitstree4/welcome.html

TREECON Draws phylogenetic trees http://bioinformatics.psb.ugent.be/software/details/TREECON

Population assignment and cluster analysis

ADMIXTURE. Produces results for unsupervised clustering (i.e., without prior population information) comparable or identical to **STRUCTURE** but faster by means of a more efficient algorithm. http://www.genetics.ucla.edu/software/admixture/

BAPS. Comparable to STRUCTURE but with increased flexibility in the definitions of levels at which genetic structure may exist. http://web.abo.fi/fak/mnf/mate/jc/software/baps.html

GENELAND. Clustering program that can make use of both geographic and genetic informations to estimate the number of populations in a dataset and delineate their spatial organisation. http://www2.imm.dtu.dk/~gigu/Geneland/

INSTRUCT. Joint inference of population structure and inbreeding rates, eliminating the assumption of Hardy-Weinberg equilibrium and especially applicable in cases of self-fertilization or inbreeding. http://cbsuapps.tc.cornell.edu/InStruct.aspx

CLUMPP. Accepts the output of **STRUCTURE** or other clustering programs in order to align the output of different runs. http://rosenberglab.bioinformatics.med.umich.edu/clumpp.html

DISTRUCT. Reads in tables of genomic components from the STRUCTURE output and files of options set by the user in order to provide graphical output of the Structure clustering. http://rosenberglab.bioinformatics.med.umich.edu/clumpp.html

PARTITION Infers population structure. Assigns test K value of the number of partitions. Less sensitive than Structure in the case of departure from Hardy Weinberg Equilibrium. http://www.genetix.univ-montp2.fr/partition/partition.htm

SPAIDA AND SPAIGN Assigns individual animals to genetic clusters based on spatial autocorrelations. http://notendur.hi.is/~snaebj/programs.html

STRUCTURE. Popular and user-friendly program for an informative visualization of patterns of diverity. Reconstructs model-based subdivision of individual genotypes into a user-specified number of clusters (k) by optimizing of Hardy-Weinberg equilibrium and minimizing linkage disequilibrium within clusters; estimates proportion of individual genomes derived from the inferred clusters. Clusters may correspond to ancestral components, but also to a relatively homogeneous breed or group of breeds. Optionally allows for admixture, linkage between markers, codominant markers, incorporation of prior population information and incorporation of spatial information. http://pritch.bsd.uchicago.edu/structure.html

WHICHRUN Assigns individuals to populations based upon maximum likelihood theory. http://www-bml.ucdavis.edu/whichrun.htm

EIGENSOFT Analyses population structure by combining statistical genetics with principal components analysis (**EIGENSTRAT**) to explicitly model ancestry differences between cases and controls along continuous axes of variation http://genepath.med.harvard.edu/~reich/Software.htm

Special purpose programs

2-MOD Jointly estimates of drift and migration. http://www.rubic.rdg.ac.uk/~mab/software.html

ADMIX Admixture program. It assumes only a single admixture event and not recurrent admixture and does not allow for drift. http://web.unife.it/progetti/genetica/Isabelle/admix2_0.html

BOTTLENECK Calculates basic descriptive statistics and detects bottlenecks and various statistical anomalies such as heterozygosity excess. http://www1.montpellier.inra.fr/URLB/bottleneck/pub.html

DLIK Drift model program. Uses coalescent theory and Monte Carlo simulation to sample of gene frequencies to estimate the number of generations from when a population was founded and the effective population size. http://www.rubic.rdg.ac.uk/~mab/software.html

FDIST2 Reads in allele frequencies in it own format and plots for the dataset the confidence intervals of the population subdivision (F_{ST}) as depending on expected heterozygosity (H_e). Plotting the F_{ST} and H_e for each marker then identifies outliers that may indicate selection. http://www.rubic.rdg.ac.uk/~mab/software.html

GENECLASS2 Reads reference and test data in GENEPOP or FSTAT format and calculates scores of assignment of individuals from the test dataset to breeds of the reference dataset. http://www1.montpellier.inra.fr/URLB/ (web page in French)

IBD (Isolation by distance) Calculates basic statistics such as allelic frequencies, heterozyogsity and genetic distances and relates geographical and genetic distances. Easy to use. The software can be implemented on line at (http://ibdws.sdsu.edu/~ibdws/) or downloaded and installed on a personal computer (http://www.bio.sdsu.edu/pub/andy/IBD.html).

LAMARC Estimation of recombination rate, migration rate, effective population sizes and exponential growth rates. http://evolution.gs.washington.edu/lamarc/lamarc_prog.html

LDNE. DOS program with graphic interface. Reads in data in GENEPOP or FSTAT formats and estimates effective population size based on linkage disequilibrium. http://fish.washington.edu/xfer/LDNE/

LEA AND PARALLEL LEA (Likelihood estimation of admixture) Allows genetic drift happened in both parental and hybrid populations. Still admixture occurred ones. http://dm.unife.it/parlea

LOSITAN Reads in GENEPOP format and plots confidence intervals of the population subdivision (F_{ST}) as depending on expected heterozygosity (H_e). Plotting for each marker F_{ST} and H_e then identifies outliers that may indicate selection. http://popgen.eu/soft/lositan

MOLECULAR EVOLUTION Provides links to an array of specialized genomics, phylogenetics and population genetics software, including modules that carry out Bayesian assessment of coalescence models, estimate recombination rates, calculate and draw phylogenetic trees. Excellent for testing population-genetic hypotheses and finding the most likely reason for the scenario observed. http://www.molecularevolution.org/software

MOLKIN Window program, reading in GENEPOP files and calculating diversity contributions, molecular coancestry, genetic distances and allelic richness corrected for sample size by rarefaction. http://www.ucm.es/info/prodanim/html/JP_Web.htm

MSVAR Detects past population growth or decline using microsatellite frequencies. http://www.rubic.rdg.ac.uk/~mab/software.html

RANNALA SOFTWARE Multiple packages that perform various specific functions, including LD mapping, data simulation, and detecting migration by using multilocus genotypes. www.rannala.org

SPAGEDI Characterizes the spatial genetic structure of individuals or populations based on genetic marker data. Estimates genetic distance and other basic statistics. http://ebe.ulb.ac.be/ebe/Software.html

Software for managing large SNP datasets

MACH Performs haplotype reconstruction and infers missing genotypes of unrelated individuals for large sets of genome-wide SNP data. http://www.sph.umich.edu/csg/abecasis/MACH/download

PLINK Clearly documented high-density SNP handling and analysis progam. Requires special two-file format. Outputs to Structure and GENABEL. Performs a variety of data handling operations and calculations, such as allele-sharing between individuals and coordination analysis. http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml

GENABEL R package (www.r-project.org) to perform whole-genome association analysis. Imports input files from PLINK and MACH software. Some data editing capabilities and nice graphic features are also available. http://mga.bionet.nsc.ru/~yurii/ABEL/GenABEL/

APPENDIX 7: ISAG/FAO RECOMMENDED MICROSATELLITE MARKERS

Species covered

- Cattle
- Buffaloes
- Sheep
- Goats
- Horses
- Donkeys
- Camelids
- Pigs
- Chicken

CATTLE

Name(s)	Chromosome	Primer sequence (5' -> 3') Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Multiplex[1] group
INRA063 (D18S5)	18	ATTTGCACAAGCTAAATCTAACC AAACCACAGAAATGCTTGGAAG	55-58	X71507	167-189	1
INRA005 (D12S4)	12	CAATCTGCATGAAGTATAAATAT CTTCAGGCATACCCTACACC	55	X63793	135-149	2
ETH225 (D9S1)	9	GATCACCTTGCCACTATTTCCT ACATGACAGCCAGCTGCTACT	55-65	Z14043	131-159	7
ILSTS005 (D10S25)	10	GGAAGCAATGAAATCTATAGCC TGTTCTGTGAGTTTGTAAGC	54-58	L23481	176-194	5
HEL5 (D21S15)	21	GCAGGATCACTTGTTAGGGA AGACGTTAGTGTACATTAAC	52-57	X65204	145-171	4
HEL1 (D15S10)	15	CAACAGCTATTTAACAAGGA AGGCTACAGTCCATGGGATT	54-57	X65202	99-119	1
INRA035 (D16S11)	16	ATCCTTTGCAGCCTCCACATTG TTGTGCTTTATGACACTATCCG	55-60	X68049	100-124	5
ETH152 (D5S1)	5	TACTCGTAGGGCAGGCTGCCTG GAGACCTCAGGGTTGGTGATCAG	55-60	Z14040, G18414	181-211	3
INRA023 (D3S10)	3	GAGTAGAGCTACAAGATAAACTTC TAACTACAGGGTGTTAGATGAACTC	55	X67830	195-225	7
ETH10 (D5S3)	5	GTTCAGGACTGGCCCTGCTAACA CCTCCAGCCCACTTTCTCTTCTC	55-65	Z22739	207-231	7

HEL9 (D8S4)	8	CCCATTCAGTCTTCAGAGGT CACATCCATGTTCTCACCAC	52-57	X65214	141-173	3
CSSM66 (D14S31)	14	ACACAAATCCTTTCTGCCAGCTGA AATTTAATGCACTGAGGAGCTTGG	55-65	...	171-209	1
INRA032 (D11S9)	11	AAACTGTATTCTCTAATAGCTAC GCAAGACATATCTCCATTCCTTT	55-58	X67823	160-204	4
ETH3 (D19S2)	19	GAACCTGCCTCTCCTGCATTGG ACTCTGCCTGTGGCCAAGTAGG	55-65	Z22744	103-133	3
BM2113 (D2S26)	2	GCTGCCTTCTACCAAATACCC CTTCCTGAGAGAAGCAACACC	55-60	M97162	122-156	7
BM1824 (D1S34)	1	GAGCAAGGTGTTTTTCCAATC CATTCTCCAACTGCTTCCTTG	55-60	G18394	176-197	2 or 8
HEL13 (D11S15)	11	TAAGGACTTGAGATAAGGAG CCATCTACCTCCATCTTAAC	52-57	X65207	178-200	4
INRA037 (D10S12)	10	GATCCTGCTTATATTTAACCAC AAAATTCCATGGAGAGAGAAAC	57-58	X71551	112-148	1
BM1818 (D23S21)	23	AGCTGGGAATATAACCAAAGG AGTGCTTTCAAGGTCCATGC	56-60	G18391	248-278	3
ILSTS006 (D7S8)	7	TGTCTGTATTTCTGCTGTGG ACACGGAAGCGATCTAAACG	55	L23482	277-309	3
MM12 (D9S20)	9	CAAGACAGGTGTTTCAATCT ATCGACTCTGGGGATGATGT	50-55	Z30343	101-145	5
CSRM60 (D10S5)	10	AAGATGTGATCCAAGAGAGAGGCA AGGACCAGATCGTGAAAGGCATAG	55-65	...	79-115	1

ETH185 (D17S1)	17	TGCATGGACAGAGCAGCCTGGC GCACCCCAACGAAAGCTCCCAG	58-67	Z14042	214-246	5
HAUT24 (D22S26)	22	CTCTCTGCCTTTGTCCCTGT AATACACTTTAGGAGAAAAATA	52-55	X89250	104-158	6
HAUT27 (D26S21)	26	TTTTATGTTCATTTTTTGACTGG AACTGCTGAAATCTCCATCTTA	57	X89252	120-158	2
TGLA227 (D18S1)	18	CGAATTCCAAATCTGTTAATTTGCT ACAGACAGAAACTCAATGAAAGCA	55-56	...	75-105	2 or 8
TGLA126 (D20S1)	20	CTAATTTAGAATGAGAGAGGCTTCT TTGGTCTCTATTCTCTGAATATTCC	55-58	...	115-131	8
TGLA122 (D21S6)	21	CCCTCCTCCAGGTAAATCAGC AATCACATGGCAAATAAGTACATAC	55-58	...	136-184	7
TGLA53 (D16S3)	16	GCTTTCAGAAATAGTTTGCATTCA ATCTTCACATGATATTACAGCAGA	55	...	143-191	7
SPS115 (D15)	15	AAAGTGACACAACAGCTTCTCCAG AACGAGTGTCCTAGTTTGGCTGTG	55-60	X16451	234-258	8

[1] The notation indicates multiplex group. The multiplexing was developed by K. Moazami- Goudarzi, INRA, Jouy-en-Josas. Indicated below are multiplex group and lane, (8 reactions, 2 reactions loaded per lane), marker and label (FAM, HEX or TET), respectively. The annealing temperature for all reactions is 55°C. For markers TGLA227 and BM1824, two alternatives (S1 and S2) are shown. See also http://diagnostics.finnzymes.fi/bovine_genotypes.html for other multiplex combinations.

Group	Lane	Marker	Label
1	1	CSRM60	FAM
		CSSM66	FAM
		HEL1	HEX
		INRA063	HEX
		INRA037	TET
2	1	INRA005	FAM
		HAUT27	HEX
		TGLA227 S1	TET
		BM1824 S1	TET
3	2	ETH3	FAM
		BM1818	HEX
		TGLA53	HEX
		ETH152	TET
		HEL9	TET
		ISLT006	FAM
4	2	HEL5	FAM
		HEL13	FAM
5	3	INRA035	FAM
		ILST005	FAM
		MM12	TET
		ETH185	TET
6	3	INRA032	TET
		HAUT24	HEX
7	4	INRA023	TET
		TGLA122	TET
		BM2113	FAM
		ETH225	HEX
		ETH10	FAM
8	4	SPS115	HEX
		TGLA126	TET
		TGLA227 S2	FAM
		BM1824 S2	HEX

BUFFALO

Name(s)	Chromosome	Primer sequence (5’ -> 3’) Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)
CSSM033	17(17)	CACTGTGAATGCATGTGTGTGAGC CCCATGATAAGAGTGCAGATGACT	65	**U03805**	154-175
CSSM038	11(10)	TTCATATAAGCAGTTTATAAACGC ATAGGATCTGGTAACTTACAGATG	55	**U03817**	163-187
CSSM043	1p(27)	AAAACTCTGGGAACTTGAAAACTA GTTACAAATTTAAGAGACAGAGTT	55	**U03824**	222-258
CSSM047	3q(8)	TCTCTGTCTCTATCACTATATGGC CTGGGCACCTGAAACTATCATCAT	55	**U03821**	127-162
CSSM036	1p(27)	GGATAACTCAACCACACGTCTCTG AAGAAGTACTGGTTGCCAATCGTG	55	**U03827**	162-176
CSSM019	1q(1)	TTGTCAGCAACTTCTTGTATCTTT TGTTTTAAGCCACCCAATTATTTG	55	**U03794**	131-161
CSRM060	11(10)	AAGATGTGATCCAAGAGAGAGGCA AGGACCAGATCGTGAAAGGCATAG	60	**AF232758**	95-135
CSSM029	9(7)	CGTGAGAACCGAAAGTCACACATTC GCTCCATTATGCACATGCCATGCT	55	**U03807**	174-196
CSSM041	21(22)	AATTTCAAAGAACCGTTACACAGC AAGGGACTTGCAGGGACTAAAACA	55	**U03816**	129-147
CSSM057	9(7)	GTCGCTGGATAAACAATTTAAAGT TGTGGTGTTTAACCCTTGTAATCT	60	**U03840**	102-130

BRN	11(10)	CCTCCACACAGGCTTCTCTGACTT CCTAACTTGCTTGAGTTATTGCCC	60	**X59767**	121-147
CSSM032	1q(1)	TTATTTTCAGTGTTTCTAGAAAAC TATAATATTGCTATCTGGAAATCC	55	**U03811**	208-224
CSSM008	Unknown	CTTGGTGTTACTAGCCCTGGG GATATATTTGCCAGAGATTCTGCA	55	**U03796**	179-193
CSSM045	2q(2)	TAGAGGCACAAGCAAACCTAACAC TTGGAAAGATGCAGTAGAACTCAT	60	**U03830**	102-122
CSSM022	4q(5)	TCTCTCTAATGGAGTTGGTTTTTG ATATCCCACTGAGGATAAGAATTC	55-60	**U03806**	203-213
CSSM046	11(10)	GGCTATTAACTGTTTTCTAGGAAT TGCACAATCGGAACCTAGAATATT	55	**U03834**	152-160
CSSM013	5p(29)	ATAAGAGATTACCCTTCCTGACTG AGGTAAATGTTCCTATTTGCTAAC	55	**U03841**	162-172
ETH003	3p(19)	GAACCTGCCTCTCCTGCATTGG ACTCTGCCTGTGGCCAAGTAGG	65	**Z22744**	96-192
CSSM061	Unknown	AGGCCATATAGGAGGCAAGCTTAC TTCAGAAGAGGGCAGAGAATACAC	60	**...**	100-126
BMC1013	3p(19)	AAAAATGATGCCAACCAAATT TAGGTAGTGTTCCTTATTTCTCTGG	54	**G18560**	217-239
DRB3	2p(23)	GAGAGTTTCACTGTGCAG CGCGAATTCCCAGAGTGAGTGAAGTATCT	50-55	**M30012**	142-198
CSSM062	Unknown	GTTTAAACCCCAGATTCTCCCTTG AGATGTAACAGCATCATGACTGAA	55	**...**	124-136

CSSME070	3p(19)	TTCTAACAGCTGTCACTCAGGC ATACAGATTAAATACCCACCTG	50-55	**AF004364**	119-139
ETH121	2q(2)	CCAACTCCTTACAGGAAATGTC ATTTAGAGCTGGCTGGTAAGTG	59	**Z14037**	182-198
ILSTS033	13(12)	TATTAGAGTGGCTCAGTGCC ATGCAGACAGTTTTAGAGGG	55	**L37213**	126-138
ILSTS005	11(10)	GGAAGCAATGAAATCTATAGCC TGTTCTGTGAGTTTGTAAGC	55	**L23481**	173-186
ILSTS030	2q(2)	CTGCAGTTCTGCATATGTGG CTTAGACAACAGGGGTTTGG	55	**L37212**	146-158
ILSTS008	15(14)	GAATCATGGATTTTCTGGGG TAGCAGTGAGTGAGGTTGGC	58	**L23483**	168-176
RM099	3p(19)	CCAAAGAGTCTAACACAACTGAG ATCCGAACCAAAATCCCATCAAG	60	**G29087**	87-119
HMH1R	21(22)	GGCTTCAACTCACTGTAACACATT TTCTTCAAGTATCACCTCTGTGGCC	60	**D10197**	169-187

[1] Cattle chromosome assignments in parentheses

[2] No multiplex developed.

SHEEP

Name(s)	Chromosome	Primer sequence (5’ -> 3’) Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Name(s)
OarFCB128	OAR2	ATTAAAGCATCTTCTCTTTATTTCCTCGC CAGCTGAGCAACTAAGACATACATGCG	55	**L01532**	96-130	...
OarCP34	OAR 3	GCTGAACAATGTGATATGTTCAGG GGGACAATACTGTCTTAGATGCTGC	50	**U15699**	112-130	...
OarCP38	OAR 10	CAACTTTGGTGCATATTCAAGGTTGC GCAGTCGCAGCAGGCTGAAGAGG	52	**U15700**	117-129	...
OarHH47	OAR 18	TTTATTGACAAACTCTCTTCCTAACTCCACC GTAGTTATTTAAAAAAATATCATACCTCTTAAGG	58	L12557	130-152	...
OarVH72	OAR 25	GGCCTCTCAAGGGGCAAGAGCAGG CTCTAGAGGATCTGGAATGCAAAGCTC	57	L12548	121-145	...
OarAE129	OAR 5	AATCCAGTGTGTGAAAGACTAATCCAG GTAGATCAAGATATAGAATATTTTTCAACACC	54	L11051	133-159	...
BM1329	OAR 6	TTGTTTAGGCAAGTCCAAAGTC AACACCGCAGCTTCATCC	50	G18422	160-182	...
BM8125	OAR 17	CTCTATCTGTGGAAAAGGTGGG GGGGGTTAGACTTCAACATACG	50	G18475	110-130	...
HUJ616	OAR 13	TTCAAACTACACATTGACAGGG GGACCTTTGGCAATGGAAGG	54	M88250	114-160	...
DYMS1	OAR 20	AACAACATCAAACAGTAAGAG CATAGTAACAGATCTTCCTACA	59		159-211	...
SRCRSP9	CHI12	AGAGGATCTGGAAATGGAATC GCACTCTTTTCAGCCCTAATG	55	L22201	99-135	1

OarCB226	OAR 2	CTATATGTTGCCTTTCCCTTCCTGC GTGAGTCCCATAGAGCATAAGCTC	60	L20006	119-153	1
ILSTS5	OAR 7	GGAAGCAATGAAATCTATAGCC TGTTCTGTGAGTTTGTAAGC	55	L23481	174-218	1
ILSTS11	OAR 9	GCT TGC TAC ATG GAA AGT GC CTA AAA TGC AGA GCC CTA CC	55	L23485	256-294	1
ILSTS28	OAR 3	TCCAGATTTTGTACCAGACC GTCATGTCATACCTTTGAGC	53	L37211	105-177	2
SRCRSP5	OAR 18	GGACTCTACCAACTGAGCTACAAG GTTTCTTTGAAATGAAGCTAAAGCAATGC	56	L22197	126-158	2
MAF214	OAR 16	GGGTGATCTTAGGGAGGTTTTGGAGG AATGCAGGAGATCTGAGGCAGGGACG	58	M88160	174-282	2
SRCRSP1	CHI13	TGCAAGAAGTTTTTCCAGAGC ACCCTGGTTTCACAAAAGG	54	L22192	116-148	3
MAF33	OAR 9	GATCTTTGTTTCAATCTATTCCAATTTC GATCATCTGAGTGTGAGTATATACAG	60	M77200	121-141	3
MCM140	OAR 6	GTTCGTACTTCTGGGTACTGGTCTC GTCCATGGATTTGCAGAGTCAG	60	L38979	167-193	3
OarFCB20	OAR 2	AAATGTGTTTAAGATTCCATACAGTG GGAAAACCCCCATATATACCTATAC	56	L20004	95-120	...
OarFCB193	OAR 11	TTCATCTCAGACTGGGATTCAGAAAGGC GCTTGGAAATAACCCTCCTGCATCCC	54	L01533	96-136	...
OarFCB304	OAR 19	CCCTAGGAGCTTTCAATAAAGAATCGG CGCTGCTGTCAACTGGGTCAGGG	56	L01535	150-188	...
OarJMP29	OAR 24	GTATACACGTGGACACCGCTTTGTAC GAAGTGGCAAGATTCAGAGGGGAAG	56	U30893	96-150	...

OarJMP58	OAR 26	GAAGTCATTGAGGGGTCGCTAACC CTTCATGTTCACAGGACTTTCTCTG	58	U35058	145-169	...
MAF65	OAR 15	AAAGGCCAGAGTATGCAATTAGGAG CCACTCCTCCTGAGAATATAACATG	60	M67437	123-127	...
MAF70	OAR 4	CACGGAGTCACAAAGAGTCAGACC GCAGGACTCTACGGGGCCTTTGC	60	M77199	124-166	...
MAF209	OAR 17	GAT CAC AAA AAG TTG GAT ACA ACC GTG G TCA TGC ACT TAA GTA TGT AGG ATG CTG	63	...	...	...
BM1824	OAR 1	GAGCAAGGTGTTTTTCCAATC CATTCTCCAACTGCTTCCTTG	58	...	...	...
INRA063	OAR 14	ATTTGCACAAGCTAAATCTAACC AAACCACAGAAATGCTTGGAAG	58	...	...	...

[1] Only a portion of markers were assigned to multiplex groups. Multiplex groups and proposed dye labels:

⇒ SRCRSP9 (FAM) + OarCB226 (HEX) + ILSTS5 (NED) + ILSTS11 (FAM)

⇒ ILSTS28 (NED) + SRCRSP5 (FAM) + MAF214 (HEX)

⇒ SRCRSP1 (NED) MAF33 (HEX) + MCM140 (FAM)

GOAT

Name(s)	Chromosome	Primer sequence (5’ -> 3’) Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Name(s)
SRCRSP5	CHI21	GGACTCTACCAACTGAGCTACAAG TGAAATGAAGCTAAAGCAATGC	55	**L22197**	156-178	1
MAF065	OAR15	AAAGGCCAGAGTATGCAATTAGGAG CCACTCCTCCTGAGAATATAACATG	58	**M67437**	116-158	2
MAF70	BTA4	CACGGAGTCACAAAGAGTCAGACC GCAGGACTCTACGGGGCCTTTGC	65	**M77199**	134-168	...
SRCRSP23	unknown	TGAACGGGTAAAGATGTG TGTTTTTAATGGCTGAGTAG	58	**...**	81-119	2
OarFCB48	OAR17	GAGTTAGTACAAGGATGACAAGAGGCAC GACTCTAGAGGATCGCAAAGAACCAG	58	**M82875**	149-173	3
INRA023	BTA3	GAGTAGAGCTACAAGATAAACTTC TAACTACAGGGTGTTAGATGAACT	58	**X80215**	196-215	2
SRCRSP9	CHI12	AGAGGATCTGGAAATGGAATC GCACTCTTTTCAGCCCTAATG	58	**L22200**	99-135	3
OarAE54	OAR25	TACTAAAGAAACATGAAGCTCCCA GGAAACATTTATTCTTATTCCTCAGTG	58	**L11048**	115-138	3
SRCRSP8	Unknown	TGCGGTCTGGTTCTGATTTCAC GTTTCTTCCTGCATGAGAAAGTCGATGCTTAG	55	**L22200**	215-255	1
SPS113	BTA10	CCTCCACACAGGCTTCTCTGACTT CCTAACTTGCTTGAGTTATTGCCC	58	**...**	134-158	3

INRABERN172	BTA26	CCACTTCCCTGTATCCTCCT GGTGCTCCCATTGTGTAGAC	58	**...**	234-256	3
OarFCB20	OAR2	GGAAAACCCCCATATATACCTATAC AAATGTGTTTAAGATTCCATACATGTG	58	**L20004**	93-112	2
CSRD247	OAR14	GGACTTGCCAGAACTCTGCAAT CACTGTGGTTTGTATTAGTCAGG	58	**...**	220-247	2
McM527	OAR5	GTCCATTGCCTCAAATCAATTC AAACCACTTGACTACTCCCCAA	58	**L34277**	165-187	2
ILSTS087	BTA6	AGCAGACATGATGACTCAGC CTGCCTCTTTTCTTGAGAG	58	**L37279**	135-155	2
INRA063	CHI18	GACCACAAAGGGATTTGCACAAGC AAACCACAGAAATGCTTGGAAG	58	**X71507**	164-186	2
ILSTS011	BTA14	GCTTGCTACATGGAAAGTGC	58	**L23485**	256-294	3
ILSTS005	BTA10	GGAAGCAATTGAAATCTATAGCC TGTTCTGTGAGTTTGTAAGC	55	**L23481**	172-218	...
SRCRSP15	Unknown	CTTTACTTCTGACATGGTATTTCC TGCCACTCAATTTAGCAAGC	55	**...**	172-198	...
SRCRSP3	CHI10	CGGGGATCTGTTCTATGAAC TGATTAGCTGGCTGAATGTCC	55	**L22195**	98-122	...
ILSTS029	BTA3	TGTTTTGATGGAACACAG TGGATTTAGACCAGGGTTGG	55	**L37252**	148-170	...
TGLA53	BTA16	GCTTTCAGAAATAGTTTGCATTCA ATCTTCACATGATATTACAGCAGA	55	**...**	126-160	...

ETH10	CHI5	GTTCAGGACTGGCCCTGCTAACA CCTCCAGCCCACTTTCTCTTCTC	55	**Z22739**	200-210	...
MAF209	CHI17	GATCACAAAAAGTTGGATACAACCGTG TCATGCACTTAAGTATGTAGGATGCTG	55	**M80358**	100-104	...
INRABERN185	CHI18	CAATCTTGCTCCCACTATGC CTCCTAAAACACTCCCACACTA	55	**X73937**	261-289	...
P19 (DYA)	Unknown	AACACCATCAAACAGTAAGAG CATAGTAACAGATCTTCCTACA	55	**AJ621046**	160-196	...
TCRVB6	BTA10	GAGTCCTCAGCAAGCAGGTC CCAGGAATTGGATCACACCT	55	**L18953**	217-255	...
SRCRSP7	CHI6	TCTCAGCACCTTAATTGCTCT GGTCAACACTCCAATGGTGAG	55	**L22199**	117-131	...
BM6444	BTA2	CTCTGGGTACAACACTGAGTCC TAGAGAGTTTCCCTGTCCATCC	65	**G18444**	118-200	...
DRBP1	BTA23	ATGGTGCAGCAGCAAGGTGAGCA GGGACTCAGTCTCTCTATCTCTTTG	58	**M55069**	195-229	...

[1] Only a portion of markers were assigned to multiplex groups. Multiplex groups and proposed dye labels:

1. SRCRSP5 (FAM) + SRCRSP8 (PET)
2. MAF065 (VIC) + SRCRSP23 (FAM) + INRA023 (FAM) +OarFCB20 (NED) + CSRD247 (PET) + McM527 (PET) + ILSTS087 (FAM) + INRA063 (VIC)
3. OarFCB48 (FAM) + SRCRSP9 (NED) + OarAE54 (VIC) + SPS113 (PET) + INRABERN172 (FAM) + ILSTS011 (FAM)

HORSE

Name(s)	Chromosome	Primer sequence (5’ -> 3’) Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Name(s)
HMS07	1	CAGGAAACTCATGTTGATACCATC TGTTGTTGAAACATACCTTGACTGT	60	X74636	165 – 183	1
HMS06	4	GAAGCTGCCAGTATTCAACCATTG CTCCATCTTGTGAAGTGTAACTCA	60	X74635	153 – 169	1
HTG07	4	CCTGAAGCAGAACATCCCTCCTTG ATAAAGTGTCTGGGCAGAGCTGCT	60	...	120 – 130	...
AHT05	8	ACGGACACATCCCTGCCTGC GCAGGCTAAGGGGGCTCAGC	60	...	130 – 146	1
HTG04	9	CTATCTCAGTCTTCATTGCAGGAC CTCCCTCCCTCCCTCTGTTCTC	55	...	127 – 141	1
HMS02	10	ACGGTGGCAACTGCCAAGGAAG CTTGCAGTCGAATGTGTATTAAATG	60	X74631	218 – 238	...
ASB02	15	CCTTCCGTAGTTTAAGCTTCTG CACAACTGAGTTCTCTGATAGG	55	X93516	222 – 254	2
HMS03	9	CCAACTCTTTGTCACATAACAAGA CCATCCTCACTTTTTCACTTTGTT	60	X74632	150 – 170	2
HTG06	15	CCTGCTTGGAGGCTGTGATAAGAT GTTCACTGAATGTCAAATTCTGCT	60	...	84 – 106	...
HTG10	21	CAATTCCCGCCCCACCCCCGGCA TTTTTATTCTGATCTGTCACATTT	55	AF169294	93 – 113	2

AHT04	24	AACCGCCTGAGCAAGGAAGT CCCAGAGAGTTTACCCT	60	**...**	148 – 164	**1C**
VHL20	30	CAAGTCCTCTTACTTGAAGACTAG AACTCAGGGAGAATCTTCCTCAG	60	**...**	86 – 106	**1**
ASB17	2	GAGGGCGGTACCTTTGTACC ACCAGTCAGGATCTCCACCG	60	X93531	91 – 109	**1**
ASB23	3	GAGGTTTGTAATTGGAATG GAGAAGTCATTTTTAACACCT	60	X93537	128 –154	**1**
LEX33	4	TTTAATCAAAGGATTCAGTTG TTTCTCTTCAGGTGTCCTC	60	AF075635	203 – 217	**2**
UCDEQ425	28	AGCTGCCTCGTTAATTCA CTCATGTCCGCTTGTCTC	60	U67406	237 – 247	**...**
LEX34	5	GCGGAGGTAAGAAGTGGTAG GGCCTAAGATGAGGGTGAA	55	AF075636	245 – 255	**3**
SGCV28	7	CTGTGGCAGCTGTCATCTTGG CCCAATTCCAGCCCAGCTTGC	62	U90604	151 – 163	**...**
COR058	12	GGGAAGGACGATGAGTGAC CACCAGGCTAAGTAGCCAAAG	58	AF108375	210 – 230	**...**
COR069	13	AGCCACCAGTCTGTTCTCTG AATGTCCTTTGGTGGATGAAC	58	AF142606	273 – 279	**3**
VHL209	14	TCTTACATCCTTCCATTACAACTA TGATACATATGTACGTGAAAGGAT	57	Y08451	84 – 96	**4**
COR007	17	GTGTTGGATGAAGCGAATGA GACTTGCCTGGCTTTGAGTC	58	AF083450	156 – 170	**...**

LEX54	18	TGCATGAGCCAATTCCTTAT TGGACAGATGACAGCAGTTC	55	AF075656	165 – 177	4
LEX73	19	CCCTAGAGCCATCTCTTTACA CAGATCCAGACTCAGGACAG	55	AF213359	234 – 264	...
COR022	22	AAGACGTGATGGGAAATCAA AGAAAGTTTTCAAATGTGCCA	58	AF101391	254 – 264	...
LEX63	23	CGGGGTGTGCATCTCTTAGG TGGCGAATGCTGAATCTGG	55	AF075663	241 – 249	
COR018	25	AGTCTGGCAATATTGAGGATGT AGCAGCTACCCTTTGAATACTG	58	AF083461	249 – 271	5
COR071	26	CTTGGGCTACAACAGGGAATA CTGCTATTTCAAACACTTGGA	58	AF142608	190 – 202	5
HMS45	27	TGTTACAGGTATTGGTAAACTGTGC GGAACAAGAAGAAATCACTAATGTC	60	U89813	185 – 197	...
COR082	29	GCTTTTGTTTCTCAATCCTAGC TGAAGTCAAATCCCTGCTTC	59	AF154935	192 – 226	...

[1] Only a portion of the markers have been assigned to multiplex groups. Five multiplex groups are proposed:

1. Multiplex Master Mix : HMS07 + HMS06 + AHT05 + HTG04 + AHT04 + VHL20 + ASB17 + ASB23
2. Multiplex Master Mix : ASB02 + HMS03 + HTG10 + LEX33
3. Multiplex Master Mix : LEX34 + COR069
4. Multiplex Master Mix : VHL209 + LEX54
5. Multiplex Master Mix : COR018 + COR071

DONKEY

Name(s)	Chromosome	Primer sequence (5’ -> 3’) Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Name(s)
HMS07	1	CAGGAAACTCATGTTGATACCATC TGTTGTTGAAACATACCTTGACTGT	58	X74636	165 – 183	1
ASB17	2	GAGGGCGGTACCTTTGTACC ACCAGTCAGGATCTCCACCG	58	X93531	91 – 109	1
ASB23	3	GAGGTTTGTAATTGGAATG GAGAAGTCATTTTTAACACCT	58	X93537	128 –154	1
HMS06	4	GAAGCTGCCAGTATTCAACCATTG CTCCATCTTGTGAAGTGTAACTCA	58	X74635	153 – 169	1
LEX34	5	GCGGAGGTAAGAAGTGGTAG GGCCTAAGATGAGGGTGAA	54	AF075636	245 – 255	3
HTG07	4	CCTGAAGCAGAACATCCCTCCTTG ATAAAGTGTCTGGGCAGAGCTGCT	58	AF142607	272 – 297	4
SGCV28	7	CTGTGGCAGCTGTCATCTTGG CCCAATTCCAGCCCAGCTTGC	60	U90604	151 – 163	...
AHT05	8	ACGGACACATCCCTGCCTGC GCAGGCTAAGGGGGCTCAGC	58	...	130 – 146	1
HMS03	9	CCAACTCTTTGTCACATAACAAGA CCATCCTCACTTTTTCACTTTGTT	58	X74632	150 – 170	2
HMS02	10	ACGGTGGCAACTGCCAAGGAAG CTTGCAGTCGAATGTGTATTAAATG	58	X74631	218 – 238	...

LEX68	11	AAATCCCGAGCTAAAATGTA TAGGAAGATAGGATCACAAGG	54	...	162 – 174	...
COR058	12	GGGAAGGACGATGAGTGAC CACCAGGCTAAGTAGCCAAAG	56	AF108375	210 – 230	4
COR069	13	AGCCACCAGTCTGTTCTCTG AATGTCCTTTGGTGGATGAAC	56	AF142606	273 – 279	3
VHL209	14	TCTTACATCCTTCCATTACAACTA TGATACATATGTACGTGAAAGGAT	56	Y08451	84 – 96	5
ASB02	15	CCTTCCGTAGTTTAAGCTTCTG CACAACTGAGTTCTCTGATAGG	54	X93516	222 – 254	2
HMS20	16	TGGGAGAGGTACCTGAAATGTAC GTTGCTATAAAAAATTGTCTCCCTAC	58	...	116 – 140	...
COR007	17	GTGTTGGATGAAGCGAATGA GACTTGCCTGGCTTTGAGTC	56	AF083450	156 – 170	...
LEX54	18	TGCATGAGCCAATTCCTTAT TGGACAGATGACAGCAGTTC	55	AF075656	165 – 177	5
LEX73	19	CCCTAGAGCCATCTCTTTACA CAGATCCAGACTCAGGACAG	54	AF213359	234 – 264	...
HTG06	15	CCTGCTTGGAGGCTGTGATAAGAT GTTCACTGAATGTCAAATTCTGCT	58	...	84 – 106	...
HTG10	21	CAATTCCCGCCCCACCCCCGGCA TTTTTATTCTGATCTGTCACATTT	54	AF169294	93 – 113	2
COR022	22	AAGACGTGATGGGAAATCAA AGAAAGTTTTCAAATGTGCCA	56	AF101391	254 – 264	...

LEX63	23	CGGGGTGTGCATCTCTTAGG TGGCGAATGCTGAATCTGG	54	AF075663	241 – 249	**...**
AHT04	24	AACCGCCTGAGCAAGGAAGT CCCAGAGAGTTTACCCT	58	**...**	148 – 164	**1**
COR018	25	AGTCTGGCAATATTGAGGATGT AGCAGCTACCCTTTGAATACTG	56	AF083461	249 – 271	**6**
COR071	26	CTTGGGCTACAACAGGGAATA CTGCTATTTCAAACACTTGGA	56	AF142608	190 – 202	**6**
HMS45	27	TGTTACAGGTATTGGTAAACTGTGC GGAACAAGAAGAAATCACTAATGTC	58	U89813	185 – 197	**...**
NVHEQ054	28	AGATGTCCACCTTCTCGCTG CGGGGCTTTTAGGAGGTAACTA	62	AJ245763	172 – 186	**...**
COR082	29	GCTTTTGTTTCTCAATCCTAGC TGAAGTCAAATCCCTGCTTC	58	AF154935	192 – 226	**...**
LEX33	4	TTTAATCAAAGGATTCAGTTG TTTCTCTTCAGGTGTCCTC	58	AF075635	203 – 217	**2**

[1] Only a portion of the markers have been assigned to multiplex groups. Six multiplex groups are proposed:

1. Multiplex Master Mix : HMS07 + ASB17 + ASB23 + HMS06 + AHT05 + AHT04
2. Multiplex Master Mix : HMS03 + ASB02 + HTG10 + LEX33
3. Multiplex Master Mix : LEX34 + COR069
4. Multiplex Master Mix : HTG07 + COR058
5. Multiplex Master Mix : VHL209 + LEX54
6. Multiplex Master Mix : COR018 + COR071

CAMELID

Name[1]	Primer sequence (5’ -> 3’) Forward Reverse	Annealing Temperature (°C)	Genebank Accession Number	Allele Lengths[2,3] (bp)
CMS9	TGCTTTAGACGACTTTTACTTTAC ATTTCACTTTCTTCATACTTGTGAT	55	**AF329160**	229–237A 227–247L 233–256B 231–243D
CMS13	TAGCCTGACTCTATCCATTTCTC ATTATTTGGAATTCAACTGTAAGG	55	**AF329158**	246–265A 242–261L 248–265B 238–254D
CMS15	AAATACTTAAAGGTTCCCAGA TTGTAAACTAAAGCCAGAAAG	55	**AF329151**	138–146A 140–146L 140–159B 121–144D
CMS17	TATAAAGGATCACTGCCTTC AAAATGAACCTCCATAAAGTTAG	55	**AF329147**	140–161A 135–147L 144–149B 149–167D
CMS18	GAACGACCCTTGAAGACGAA AGCAGCTGGTTTTAGGTCCA	60	**AF329148**	165–182A 165–188L 157–186B 157–163D
CMS25	GATCCTCCTGCGTTCTTATT CTAGCCTTTGATTGGAGCAT	58	**AF380345**	93–118A 93–95L 118–128B 93–102D
CMS32	ACGGACAAGAACTGCTCATA ACAACCAATAAATCCCCATT	55	**AF329146**	167–169A 167–169L 198–204B

				198–209D
CMS50	TTTATAGTCAGAGAGAGTGCTG TGTAGGGTTCATTGTAACA	55	**AF329149**	129–135A 129–140L 154–183B 170–190D
CMS121	CAAGAGAACTGGTGAGGATTTTC AGTTGATAAAAATACAGCTGGAAAG	60	**AF329159**	128–157A 128–151L 151–159B 147–166D
CVRL01	GAAGAGGTTGGGGCACTAC CAGGCAGATATCCATTGAA	55	**AF217601**	Polymorphic A 188-253B 196-253D
CVRL02	TGTCACAAATGGCAAGAT AGTGTACGTAGCAGCATTATTT	55	**AF217602**	Polymorphic A 206-216B 205-216D
CVRL05	CCTTGGACCTCCTTGCTCTG GCCACTGGTCCCTGTCATT	60	**AF217602**	Polymorphic A 148-174B 155-176D
CVRL06	TTTTAAAAATTCTGACCAGGAGTCTG CATAATAGCCAAAACATGGAAACAAC	60	**AF217606**	Polymorphic A 185-205B 196-203D
CVRL07	AATACCCTAGTTGAAGCTCTGTCCT GAGTGCCTTTATAAATATGGGTCTG	55	**AF217607**	Polymorphic A 255-263B 272-306D
CVRL07	GTGCAGCGTCCAAATAGTCA CCAGCATCGTCCAGTATTCA	50-58	**AF091125**	220-262A+L 212-242B 240-244D
VOLP03	AGACGGTTGGGAAGGTGGTA CGACAGCAAGGCACAGGA	55-60	**AF305228**	129-169A 145-206B 145-176D
VOLP08	CCATTCACCCCATCTCTC	55	**AF305230**	148-152A

	TCGCCAGTGACCTTATTTAGA			142-180B 144-150D
VOLP10	CTTTCTCCTTTCCTCCCTACT CGTCCACTTCCTTCATTTC	55	**AF305231**	231-235A 232-260B 250-268D
VOLP32	GTGATCGGAATGGCTTGAAA CAGCGAGCACCTGAAAGAA	55	**AF305234**	192-247A 256-262B 256-262D
VOLP67	TTAGAGGGTCTATCCAGTTTC TGGACCTAAAAGAGTGGAG	55	**AF305237**	158-170A 142-172B 150-203D
YWLL 08	ATCAAGTTTGAGGTGCTTTCC CCATGGCATTGTGTTGAAGAC	55-60	**...**	135-177A+L 154-180B 133-172D
YWLL 09	AAGTCTAGGAACCGGAATGC AGTCAATCTACACTCCTTGC	50-58	**...**	154-180A+L 158-177B 158-162D
YWLL 38	GGCCTAAATCCTACTAGAC CCTCTCACTCTTGTTCTCCTC	55-60	**...**	174-178A+L 180-192B 182-190D
YWLL 44	CTCAACAATGCTAGACCTTGG GAGAACACAGGCTGGTGAATA	55-60	**...**	86-120A+L 101-117B 90-114D
YWLL 59	TGTGCAGGAGTTAGGTGTA CCATGTCTCTGAAGCTCTGGA	50-58	**...**	96-136A+L 109-135B 109-111D

[1] Markers are not assigned to chromosomes but are all believed to be autosomal.

[2] A = alpaca (*Lama pacos*), L = llama (*Lama glama*), B = Bactrian camel (*Camelus bactrianus*), D = dromedary camel (*Camelus dromedarius*).

[3] No multiplexes developed.

PIG

Name	Chromosome	Primer sequence (5’ -> 3’) Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Multiplex[1] Group
S0026	16	AACCTTCCCTTCCCAATCAC CACAGACTGCTTTTTACTCC	55	L30152	156-178	1
S0155	1	TGTTCTCTGTTTCTCCTCTGTTTG AAAGTGGAAAGAGTCAATGGCTAT	55	...	116-158	1
S0005	5	TCCTTCCCTCCTGGTAACTA GCACTTCCTGATTCTGGGTA	55	...	134-168	1
Sw2410	8	ATTTGCCCCCAAGGTATTTC CAGGGTGTGGAGGGTAGAAG	50	AF207836	81-119	2
Sw830	10	AAGTACCATGGAGAGGGAAATG ACATGGTTCCAAAGACCTGTG	50	AF235378	149-173	2
S0355	15	TCTGGCTCCTACACTCCTTCTTGATG TTGGGTGGGTGCTGAAAAATAGGA	50	L29049	196-215	2
Sw24	17	CTTTGGGTGGAGTGTGTGC ATCCAAATGCTGCAAGCG	55	AF235245	99-135	3
Sw632	7	TGGGTTGAAAGATTTCCCAA GGAGTCAGTACTTTGGCTTGA	55	AF225099	115-138	3
Swr1941	13	AGAAAGCAATTTGATTTGCATAATC ACAAGGACCTACTGTATAGCACAGG	55	AF253904	215-255	3
Sw936	15	TCTGGAGCTAGCATAAGTGCC GTGCAAGTACACATGCAGGG	55	AF225107	134-158	4
S0218	x	GTGTAGGCTGGCGGTTGT CCCTGAAACCTAAAGCAAAG	55	L29048	234-256	4
S0228	6	GGCATAGGCTGGCAGCAACA AGCCCACCTCATCTTATCTACACT	55	L29195	93-112	4
Sw122	6	CAAAAAAGGCAAAAGATTGACA	55	AF235206	220-247	5

		TTGTCTTTTTATTTTGCTTTTGG				
Sw857	14	TGAGAGGTCAGTTACAGAAGACC GATCCTCCTCCAAATCCCAT	55	**AF225105**	165-187	**5**
S0097	4	GACCTATCTAATGTCATTATAGT TTCCTCCTAGAGTTGACAAACTT	55	**M95020**	135-155	**5**
sw240	2	AGAAATTAGTGCCTCAAATTGG AAACCATTAAGTCCCTAGCAAA	55	**AF235246**	164-186	**6**
IGF1	5	GCTTGGATGGACCATGTTG CATATTTTTCTGCATAACTTGAACCT	55	**...**	256-294	**6**
Sw2406	6	AATGTCACCTTTAAGACGTGGG AATGCGAAACTCCTGAATTAGC	55	**AF225140**	117-131	**6**
Sw72	3	ATCAGAACAGTGCGCCGT TTTGAAAATGGGGTGTTTCC	55	**AF235346**	172-218	**7**
S0226	2	GCACTTTTAACTTTCATGATACTCC GGTTAAACTTTTNCCCCAATACA	55	**L29230**	172-198	**7**
S0090	12	GCACTTTTAACTTTCATGATACTCC GGTTAAACTTTTNCCCCAATACA	55	**M95002**	98-122	**7**
Sw2008	11	CAGGCCAGAGTAGCGTGC CAGTCCTCCCAAAAATAACATG	55	**AF253773**	148-170	**8**
Sw1067	6	TGCTGGCCAGTGACTCTG CCGGGGGATTAAACAAAAAG	55	**AF235183**	126-160	**8**
S0101	7	GAATGCAAAGAGTTCAGTGTAGG GTCTCCCTCACACTTACCGCAG	55	**...**	200-210	**8**
Sw1828	1	AATGCATTGTCTTCATTCAACC TTAACCGGGGCACTTGTG	55	**AF253712**	100-104	**9**
S0143	12	ACTCACAGCTTGTCCTGGGTGT CAGTCAGCAGGCTGACAAAAAC	55	**...**	261-289	**9**
S0068	13	CCTTCAACCTTTGAGCAAGAAC AGTGGTCTCTCTCCCTCTTGCT	55	**...**	118-200	**9**

S0178	8	TAGCCTGGGAACCTCCACACGCTG GGCACCAGGAATCTGCAATCCAGT	60	**...**	160-196	**10**
Sw911	9	CTCAGTTCTTTGGGACTGAACC CATCTGTGGAAAAAAAAAGCC	60	**AF225106**	217-255	**10**
S0002	3	GAAGCCAAAGAGACAACTGC GTTCTTTACCCACTGAGCCA	60	**...**	195-229	**10**

[1] Multiplex indicated here should only be considered as propositions of sets of markers compatible in size (which could thus be labelled with the same dye) which generally could be amplified in similar conditions. Most of these markers have also been successfully amplified in very different conditions (annealing conditions differing of up to 10°C), it is strongly recommended, however, that all persons using this information determine the annealing temperature that is optimal in their own PCR conditions.

CHICKEN

Name	Chromosome	Primer sequence (5' -> 3') Forward Reverse	Annealing temperature (°C)	Genebank Accession Number	Allele range (bp)	Multiplex[1] group
ADL0268	1	CTCCACCCCTCTCAGAACTA CAACTTCCCATCTACCTACT	60	**G01688**	102-116	**1**
MCW0206	2	CTTGACAGTGATGCATTAAATG ACATCTAGAATTGACTGTTCAC	60	**AF030579**	221-249	**7**
LEI0166	3	CTCCTGCCCTTAGCTACGCA TATCCCCTGGCTGGGAGTTT	60	**X85531**	354-370	**3**
MCW0295	4	ATCACTACAGAACACCCTCTC TATGTATGCACGCAGATATCC	60	**G32051**	88-106	**2**
MCW0081	5	GTTGCTGAGAGCCTGGTGCAG CCTGTATGTGGAATTACTTCTC	60	**...**	112-135	**2**
MCW0014	6	TATTGGCTCTAGGAACTGTC GAAATGAAGGTAAGACTAGC	58	**...**	164-182	**4**
MCW0183	7	ATCCCAGTGTCGAGTATCCGA TGAGATTTACTGGAGCCTGCC	58	**G31974**	296-326	**4**
ADL0278	8	CCAGCAGTCTACCTTCCTAT TGTCATCCAAGAACAGTGTG	60	**G01698**	114-126	**1**
MCW0067	10	GCACTACTGTGTGCTGCAGTTT GAGATGTAGTTGCCACATTCCGAC	60	**G31945**	176-186	**6**
MCW0104	13	TAGCACAACTCAAGCTGTGAG AGACTTGCACAGCTGTGTACC	60	**...**	190-234	**5**

MCW0123	14	CCACTAGAAAAGAACATCCTC GGCTGATGTAAGAAGGGATGA	60	**...**	76-100	**5**
MCW0330	17	TGGACCTCATCAGTCTGACAG AATGTTCTCATAGAGTTCCTGC	60	**G32085**	256-300	**6**
MCW0165	23	CAGACATGCATGCCCAGATGA GATCCAGTCCTGCAGGCTGC	60	**...**	114-118	**5**
MCW0069	E60C04W23	GCACTCGAGAAAACTTCCTGCG ATTGCTTCAGCAAGCATGGGAGGA	60	**...**	158-176	**2**
MCW0248	1	GTTGTTCAAAAGAAGATGCATG TTGCATTAACTGGGCACTTTC	60	**G32016**	205-225	**1**
MCW0111	1	GCTCCATGTGAAGTGGTTTA ATGTCCACTTGTCAATGATG	60	**L48909**	96-120	**3**
MCW0020	1	TCTTCTTTGACATGAATTGGCA GCAAGGAAGATTTTGTACAAAATC	60	**...**	179-185	**5**
MCW0034	2	TGCACGCACTTACATACTTAGAGA TGTCCTTCCAATTACATTCATGGG	60	**...**	212-246	**2**
LEI0234	2	ATGCATCAGATTGGTATTCAA CGTGGCTGTGAACAAATATG	60	**Z94837**	216-364	**3**
MCW0103	3	AACTGCGTTGAGAGTGAATGC TTTCCTAACTGGATGCTTCTG	64	**G31956**	266-270	**7**
MCW0222	3	GCAGTTACATTGAAATGATTCC TTCTCAAAACACCTAGAAGAC	60	**G31997**	220-226	**2**
MCW0016	3	ATGGCGCAGAAGGCAAAGCGATAT TGGCTTCTGAAGCAGTTGCTATGG	60	**...**	162-206	**3**

MCW0037	3	ACCGGTGCCATCAATTACCTATTA GAAAGCTCACATGACACTGCGAAA	64	**...**	154-160	**3**
MCW0098	4	GGCTGCTTTGTGCTCTTCTCG CGATGGTCGTAATTCTCACGT	60	**...**	261-265	**6**
LEI0094	4	GATCTCACCAGTATGAGCTGC TCTCACACTGTAACACAGTGC	60	**X83246**	247-287	**1**
MCW0284	4	GCCTTAGGAAAAACTCCTAAGG CAGAGCTGGATTGGTGTCAAG	60	**G32043**	235-243	**...**
MCW0078	5	CCACACGGAGAGGAGAAGGTCT TAGCATATGAGTGTACTGAGCTTC	60	**...**	135-147	**6**
LEI0192	6	TGCCAGAGCTTCAGTCTGT GTCATTACTGTTATGTTTATTGC	60	**Z83797**	244-370	**...**
ADL0112	10	GGCTTAAGCTGACCCATTAT ATCTCAAATGTAATGCGTGC	58	**G01725**	120-134	**4**
MCW0216	13	GGGTTTTACAGGATGGGACG AGTTTCACTCCCAGGGCTCG	60	**AF030586**	139-149	**1**

[1] All but two of the markers have been assigned to multiplex groups. Seven multiplex groups are proposed, which have been tested in the Qiagen system:

1. Multiplex Master Mix Qiagen: ADL0268 + ADL0278 + MCW0248 + LEI0094 + MCW0216
2. Multiplex Master Mix Qiagen: MCW0295 + MCW0081 + MCW0069 + MCW0034 + MCW0222
3. Multiplex Master Mix Qiagen: LEI0166 + MCW0111 + LEI0234 + MCW0016 + MCW0037
4. Multiplex Master Mix Qiagen: MCW0014 + MCW0183 + ADL0112
5. Multiplex Master Mix Qiagen: MCW0104 + MCW0123 + MCW0165 + MCW0020
6. Multiplex Master Mix Qiagen: MCW0067 + MCW0330 + MCW0098 + MCW0078
7. Hot Star Taq Master Mix Qiagen : MCW0206 + MCW0103