FOOD STANDARDS
Australia New Zealand
Te Mana Kounga Kai – Ahitereiria me Aotearoa

20 December 2017
[35-17]

Supporting document 1

Safety Assessment Report (at Approval) – Application A1138

Food derived from Provitamin A Rice Line GR2E

Summary and conclusions

Background

A genetically modified (GM) rice line with OECD Unique Identifier IR-00GR2E-5 (herein referred to as GR2E) has been developed by the International Rice Research Institute. This rice line has been genetically modified to produce beta (β)-carotene (a form of provitamin A) in the endosperm of the rice grain. This trait has been achieved through expression of a phytoene synthase protein (PSY1) encoded by a gene (*Zmpsy1*) from *Zea mays* (corn) and a carotene desaturase protein (CRTI) encoded by a gene (*crtI*) from the bacterium *Pantoea ananatis*. These two proteins, normally absent in rice endosperm, supply the necessary intermediaries to support a functional β-carotene biosynthetic pathway.

GR2E also contains the bacterial phosphomannose isomerase (PMI) gene, which is derived from *Escherichia coli* strain K-12. Expression of the PMI protein in plants allows growth on mannose as a carbon source. This was used as a selectable marker to assist with identification of transformed rice cells in the early stages of selection.

In conducting a safety assessment of food derived from GR2E, a number of criteria have been addressed including: a characterisation of the transferred gene sequences, their origin, function and stability in the rice genome; the changes at the level of DNA and protein in the whole food; compositional analyses and evaluation of intended and unintended changes.

This safety assessment report addresses only food safety and nutritional issues of the GM food *per se*. It therefore does not address:

- environmental risks related to the environmental release of GM plants used in food production
- the safety of animal feed, or animals fed with feed, derived from GM plants
- the safety of food derived from the non-GM (conventional) plant.

History of use

Rice (*Oryza* spp.) is a significant caloric source for over half of the world's population and the cultivated species *Oryza sativa* is the most widely utilised, with thousands of cultivars being grown in over 100 countries. In terms of production, paddy rice is the world's second most dominant cereal crop behind corn.

Most rice worldwide is consumed in the form of cooked rice. However, there are now a variety of rice-based value added products available, such as breakfast cereals (puffed rice), pasta, rice flakes, and rice crackers. Rice flour is used to make products such as rice noodles, egg-roll wrappers, edible rice paper and rice flour cakes/dumplings. Liquid-based rice products include rice bran oil that is used in cooking, rice-based alcoholic beverages produced after fermentation, rice vinegar, rice 'milk' and rice syrup. Rice starch is used as a thickening agent in food preparation. Rice by-products such as bran and straw are used in livestock feed.

Molecular characterisation

GR2E was generated through *Agrobacterium*-mediated transformation with a single T-DNA containing three expression cassettes. Comprehensive molecular analyses indicate there is a single insertion site comprising a single, complete copy of each of the *Zmpsy1, SSU-crtI* and *pmi* genes with their regulatory elements. The introduced genes are stably inherited from one generation to the next.

There are no antibiotic resistance marker genes present in the line and no plasmid backbone has been incorporated into the transgenic locus.

Characterisation and safety assessment of new substances

Newly expressed proteins

GR2E expresses three novel proteins, *Zm*PSY1, CRTI and PMI.

Analyses of all three proteins in straw and various grain stages indicated that measurable but very low concentrations of both PSY1 and CRTI proteins were found in all GR2E grain developmental stages but not in GR2E stem/straw tissue. Levels of CRTI were lower than *Zm*PSY1; both proteins had the highest mean concentration in grain at dough stage, ranging between *ca*. 308–359 ng/g and *ca*. 54–68 ng/g for *Zm*PSY1 and CRTI, respectively. Concentrations of PMI protein were significantly higher than either *Zm*PSY1 or CRTI in samples from all grain growth stages and were highest in dough-stage grain, averaging *ca*. 2015 ng/g across the locations/growing seasons tested. PMI was also present in straw.

Very low yields of the plant-produced *Zm*PSY1 and CRTI proteins precluded their specific characterisation. However the weight of evidence, provided by a) translation of the known DNA sequences of the three genes introduced into GR2E, b) the fact that the proteins function as predicted in the plant c) detection by specific antibodies and d) the detailed characterisation of equivalent microbially-produced proteins, is sufficient to confirm the identity of the proteins expressed in GR2E.

For the *Zm*PSY1 and CRTI proteins, bioinformatic studies confirmed the lack of any significant amino acid sequence similarity to know protein toxins or allergens; digestibility studies suggest the proteins would be rapidly degraded in the gastro-intestinal tract following ingestion; and thermolability studies indicate both proteins are functionally inactivated following heating.

There are also no concerns regarding the potential toxicity or allergenicity of the PMI protein which has been previously assessed as safe by FSANZ. Updated bioinformatics studies submitted in this application confirm the lack of any significant amino acid sequence similarity of the PMI expressed in GR2E to known protein toxins or allergens.

Taken together, the evidence indicates that *Zm*PSY1, CRTI and PMI in the diet are unlikely to be toxic or allergenic in humans.

Compositional analyses

Detailed compositional analyses were done to establish the nutritional adequacy of grain from GR2E and to characterise any unintended compositional changes. Samples were analysed for proximates, fibre, sugars, fatty acids, amino acids, minerals, vitamins, and anti-nutrients (phytic acid and trypsin inhibitor). In addition to these analytes the Applicant provided data and analysis on five carotenoid categories.

As expected, there are significant increases in the carotenoid component of the grain, particularly the production of β-carotene. FSANZ has conducted a separate nutrition safety assessment to determine the potential risks associated with the intake of the contained carotenoids by the general population and whether this may cause any adverse health effects (see Supporting Document 2).

The levels of each analyte in the major analyses were compared to levels in: a) the non-GM parental line, Rc82 and b) levels recorded in the literature, noting that data are limited for raw rice. Of the 69 analytes considered in the major analyses, and with the exception of β-carotene which was expected to be higher in GR2E, only stearic acid in GR2E differed significantly from the control. All non-carotenoid analyte levels in GR2E were within, or similar to, the range of natural variability of those components in conventional rice varieties as reported in the limited literature. It can therefore be concluded that apart from the elevated levels of carotenoids, the grain in GR2E is compositionally equivalent to grain from conventional rice varieties.

Conclusion

No potential public health and safety concerns have been identified in the assessment of GR2E. On the basis of the data provided in the present Application, and other available information, food derived from GR2E is considered to be as safe for human consumption as food derived from conventional rice varieties.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AOAC	Association of Official Analytical Chemists
AP	alkaline phosphatase
BLAST	Basic Local Alignment Search Tool
BLASTP	Basic Local Alignment Search Tool - Protein
bp	base pairs
CRTI	carotene desaturase I
DIG	digoxigenin
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
dw	dry weight
ELISA	enzyme linked immunosorbent assay
FAD	flavin adenine dinucleotide
FAO	Food & Agricultural Organization of the United Nations
FARRP	Food Allergy Research and Resource Program
FASTA	Fast Alignment Search Tool – All
FPIES	food protein-induced enterocolitis syndrome
FSANZ	Food Standards Australia New Zealand
fw	fresh weight
g	gram
GGPP	geranylgeranyl diphosphate
gluA-2	glutelin
GM	genetically modified
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
IgE	Immunoglobulin E
IPP	isopentenyl diphosphate
IRRI	International Rice Research Institute
kb	kilo base
kDa	kilo Dalton
kg	kilogram
LAAO	L-amino acid oxidase
LB	Left Border of T-DNA (*Agrobacterium tumefaciens*)
LOQ	limit of quantification
MALDI-TOF MS/MS	matrix-assisted laser desorption/ionisation time of flight tandem mass spectrometry
mg	milligram
MT	Million tonnes
NCBI	National Centre for Biotechnology Information
ng	nannogram
NOS	nopaline synthase
OECD	Organisation for Economic Co-operation and Development
OGTR	Office of the Gene Technology Regulator
ORF	open reading frame
PMI	phosphomannose isomerase
PCR	polymerase chain reaction
PSY1	phytoene synthase

PVDF	polyvinylidene fluoride
RB	Right Border of T-DNA (*Agrobacterium tumefaciens*)
RUBISCO SSU	ribulose bisphosphate carboxylase small subunit
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGF	simulated gastric fluid
the Code	Australia New Zealand Food Standards Code
T-DNA	transfer DNA
Ti	tumour inducing
TIU	Trypsin inhibitor units
ubi	polyubiquitin
µg	microgram
U.S.	United States of America
UTR	untranslated region

Greek symbols used in the text

α	alpha
β	beta
ζ	zeta
χ	chi
µ	mu (micro)

1 Introduction

The International Rice Research Institute (IRRI) has submitted an application to FSANZ to vary Schedule 26 in the *Australia New Zealand Food Standards Code* (the Code) to include food from a new genetically modified (GM) rice (*Oryza sativa*) line, GR2E, with OECD Unique Identifier IR-00GR2E-5 (herein referred to as GR2E). This rice line has been genetically modified to produce beta (β)-carotene (a form of provitamin A) in the endosperm of the rice grain.

This trait has been achieved through expression of a phytoene synthase protein (PSY1) encoded by a gene (*Zmpsy1*) from *Zea mays* (corn) and a carotene desaturase protein (CRTI) encoded by a gene (*crtI*) from the bacterium *Pantoea ananatis*. These two proteins, normally absent in rice endosperm, supply the necessary intermediaries to support a functional β-carotene biosynthetic pathway. When ingested, β-carotene is an effective source of vitamin A. The name 'Golden Rice' has been used to describe a number of versions of rice containing these two proteins (not necessarily from the same genes as used in GR2E).

The Applicant states the intended purpose of GR2E is to complement existing vitamin A deficiency control efforts by supplying up to 30–50 percent of the estimated average requirement for vitamin A for preschool age children and pregnant or lactating mothers in high-risk countries, including Bangladesh, Indonesia, and the Philippines.

GR2E also contains the bacterial *phosphomannose isomerase* (*pmi)* gene, which is derived from *Escherichia coli* strain K-12. Expression of the PMI protein in cells allows growth on mannose as a carbon source. This was used as a selectable marker to assist with identification of transformed rice cells in the early stages of selection. The PMI protein has been previously assessed by FSANZ in four corn applications - A564 (FSANZ 2006), A580 (FSANZ 2008b), A1001 (FSANZ 2008a), A1060 (FSANZ 2012).

GR2E was developed from a specific rice cultivar (see Section 2.1 below). Countries that wish to adopt the Golden Rice technology are free to introduce the GR2E event into preferred varieties that suit the local environment and meet certain humanitarian use criteria[1]. Rice containing the GR2E event is not intended for commercialisation (either growing or intentional sale in the food supply) in Australia or New Zealand but it is possible it could inadvertently enter the food supply via exports from countries that may supply significant quantities of milled rice to Australia/New Zealand.

2 History of use

2.1 Host organism

Unless otherwise referenced, the following description of the host organism was adapted from documents published by the Australian Government Office of the Gene Technology Regulator (OGTR 2005), the International Life Sciences Institute (ILSI 2008), the Global Rice Science Partnership (GRiSP 2013) Muthayya et al. (2014) and OECD (2016). Statistical data are from the most recent data available from the Food and Agriculture Organization of the United Nations (FAOSTAT 2017).

Rice (*Oryza* spp.) is a significant caloric source for over half of the world's population and the cultivated species *Oryza sativa* is the most widely grown. The earliest evidence of cultivation of *O. sativa* was in Southeast Asia and dates back to between 8,000 and 15,000 years ago.

[1] Golden Rice Project: Golden Rice and Intellectual Property - http://www.goldenrice.org/Content1-Who/who4_IP.php

Thousands of cultivars are now grown in over 100 countries. There are three main ecological varieties – indica (distribution in tropical and sub-tropical Asia and accounting for *ca.* 80% of global rice production), javanica (grown in Indonesia and the Philippines) and japonica (temperate distribution – especially Japan, northern China, Europe and Australia).

The parent line that was transformed to give GR2E is Kaybonnet (Gravois et al. 1995). For technical reasons the cultivars donated for humanitarian use in the Golden Rice project were North American varieties, of which Kaybonnet was one. This long-grain cultivar was developed jointly by the Arkansas Agricultural Experiment Station and the United States Department of Agriculture Agricultural Research Service and was released in 1994 for its blast resistance, high yield potential and excellent milling yields. It is a tropical japonica cultivar with an introgression from indica (Jia et al. 2012).

Rice is grown mainly as an annual under a variety of water regimes that range from unsubmerged upland rice (about 10% of total cultivation), moderately submerged lowland rice (irrigated – 45% or rain-fed – 30%) and submerged (up to 6 m of water – 11% or floating – 4%). Worldwide, rice production was significantly increased as a result of the Green Revolution between the 1940s and the late 1960s which led to semi-dwarf, early maturing varieties that allowed up to three plantings a year.

Paddy rice (also known as 'rough' rice) is the name given to harvested and threshed rice grains that comprise an outer hull (also called husk) layer plus germ and bran layers (Figure 1). Removal of hulls ('dehulled' or 'unpolished' rice) through a preliminary milling gives brown rice while further milling also removes the germ and bran layers to give white rice. Since a high proportion of vitamins, minerals and dietary fibre are found in the germ and bran layers, increased milling depletes the nutritional value of the grain. On the other hand, since white rice does not have the oil-rich bran layer, it stores for much longer than brown rice; the main reason for milling of rice for use in tropical and sub-tropical climates is to prevent losses due to rancidity (Beyer et al. 2002). A process called 'parboiling', in which raw paddy rice grains are soaked in water and partially steamed before drying and milling, is sometimes employed to promote the migration of B vitamins into the grain and, hence, improve retention in white rice; parboiling also improves rice hardness and therefore reduces breakage loss during subsequent milling. In the United States of America (U.S.) most fully milled and polished white rice is 'enriched' with vitamins B1, B3, iron and folic acid. Milled rice in general comprises 'head-rice' (high value intact grains) and 'broken-rice' (low value broken grains).

In terms of production, paddy rice is the world's second most dominant cereal crop behind corn. In 2014 world production was nearly 741 MT with mainland China (206 MT), India (157 MT), Indonesia (70 MT), Bangladesh (52 MT) and Viet Nam (45 MT) being the top five producers. The Asian region accounted for 91% of production. The paddy rice global production translated to approximately 494 MT milled rice in 2014; the current (November 2016) forecast for milled rice production[2] in 2016/2017 is 498 MT. Most rice is eaten in the same country where it is produced and global trade accounts for less than 10% of total production. Australia, while being a modest producer (819,000 T in 2014), is a net exporter of milled rice, and has the world's third highest yields (behind China and Egypt). New Zealand produces negligible rice and is a net importer. The main counties exporting to Australia and New Zealand are India, Thailand, Viet Nam and Pakistan. No GM rice is currently grown in either Australia or New Zealand.

[2] FAO – Cereal Supply and Demand Brief - http://www.fao.org/worldfoodsituation/csdb/en/

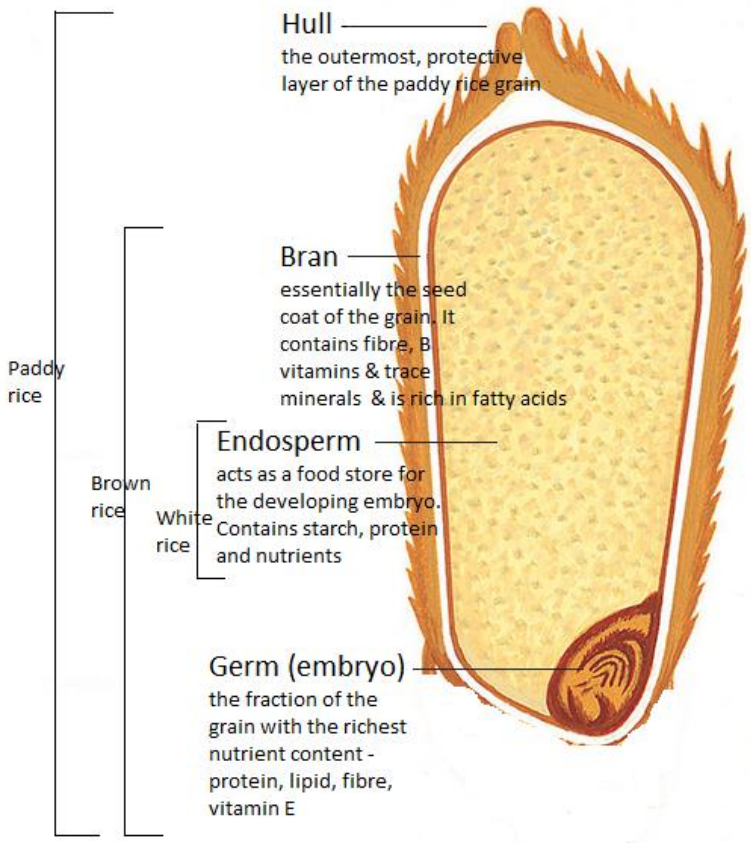

Figure 1: Stylised representation of a longitudinal section through a rice grain
Adapted from World Book Online, diagram by James Teason - www.worldbookonline.com/student/media?id=lr000932

Most rice worldwide is consumed in the form of cooked head- or broken-rice. However, there are now a variety of rice-based value added products[3] available, such as breakfast cereals (puffed rice), pasta, rice flakes, and rice crackers. Rice flour is used to make products such as rice noodles, egg-roll wrappers, edible rice paper and rice flour cakes/dumplings. Liquid-based rice products include rice bran oil that is used in cooking, rice-based alcoholic beverages produced after fermentation, rice vinegar, rice 'milk' and rice syrup. Rice starch is used as a thickening agent in food preparation. Rice by-products such as bran and straw are used in livestock feed.

Few food-induced allergic reactions to rice have been reported and rice is not regarded as a commonly allergenic food, although there is an increasing number of allergy cases being reported in Japan, where the frequency of IgE-mediated rice allergy is about 10% in atopic subjects (Suvarna 2008), and Korea (see e.g. Jeon et al. 2011). A number of proteins have been identified as possible allergens, including 14-16 kDa proteins (Izumi et al. 1999; Ito et al. 2005) listed as Oryza trypsin alpha-amylase inhibitors in the Food Allergy Research and Resource Program[4] database, a 33 kDa protein identified as a plant glyoxalase I (Usui et al. 2001) and possibly others (see references in Trcka et al. 2012). More recently (Jeon et al. 2011) clinical symptoms were linked to 4, 13 and 31 kDa proteins. Raw rice is more allergenic than cooked.

Rice is also reported to be one of the several triggers in infants and young children of food protein-induced enterocolitis syndrome (FPIES), an adverse reaction involving the immune system but not IgE antibodies (Mehr et al. 2009; ASCIA 2014).

[3] Ricepedia - http://ricepedia.org/rice-as-food/other-rice-products
[4] FARRP - http://www.allergenonline.org

2.2 Donor organisms

2.2.1 *Zea mays*

Corn, *Zea mays*, is the source of the *psy1* gene and the *UBI* promoter. In terms of production, corn is the world's dominant cereal crop (1,038 MT in 2014) ahead of rice (741 MT) and wheat (729 MT) and is grown (to some extent in over 200 countries (FAOSTAT 2017) and across a wide range of geographical conditions (OGTR 2008). Also known as maize, corn has been grown and consumed in Mexico and Central America for some 8000 years and in Europe for 500 years and can thus be said to have a long history of safe use as a human food. The majority of corn that is grown however is destined for use as animal feed.

2.2.2 *Pantoea ananatis*

This bacterium, formerly *Erwinia uredovora* (Mergaert et al. 1993), is the source of the *crtI* gene. It is a pathogen of a wide range of agricultural food crops (e.g. pineapple, corn, onion, honeydew melon, rice and banana) and forest tree species (Coutinho and Venter 2009). While it associates with plants as an epiphyte, endophyte, pathogen or symbiont, it can also occupy other ecological niches and has been found, for example, in rivers, soil samples and aviation fuel tanks (see references in Coutinho and Venter 2009) and even in refrigerated beef (Ercolini et al. 2006); limited reports of colonisation in humans have been made (e.g. De Baere et al. 2004). It is also considered to be a potentially useful bio-control agent for some plant pathogens - such as fungi (Enya et al. 2007) and other bacteria (cited in Wu et al. 2016) – and insect pests (Watanabe and Sato 1999). The genome has been sequenced (Wu et al. 2016). It is the nature of the pan-genome[5] of *P. ananatis* that confers such wide adaptability (DE Maayer et al. 2014).

2.2.3 *Escherichia coli*

The bacterium *Escherichia coli* is the source of the selectable marker gene, *pmi. E. coli* belongs to the Enterobacteriaceae, a relatively homogeneous group of rod-shaped, Gram-negative, facultative aerobic bacteria. Members of the genus *Escherichia* are ubiquitous in the environment and are normally found in the digestive tracts of vertebrates, including humans where they are the most abundant facultative aerobe (Donnenberg 2002). The vast majority of *E. coli* strains are harmless to humans, although some strains can cause diarrhoea and occasionally urinary tract infections.

Some strains of *E. coli*, such as the enterohaemorrhagic *E. coli* group (e.g. 0157:H7), are particularly virulent pathogenic strains responsible for causing serious food-borne illness. This particular group of pathogenic *E. coli* are distinct from the strains of *E. coli* (the K-12 strains) that are used routinely in laboratory manipulations; the *E. coli* used as a donor organism in this application is K-12. The K-12 strains of *E. coli* have a long history of safe use and are commonly used as protein production systems in many commercial applications, including for pharmaceutical products (Baeshen et al. 2015) and food ingredients (e.g. Schedule 18 of the Code permits the use of chymosin derived from *E. coli* K-12 strain as a food processing aid).

2.2.4 Other organisms

Genetic elements from two other organisms not mentioned above (*Pisum sativum* – pea; and *Agrobacterium tumefaciens*) have been used in the genetic modification of GR2E (refer to Table 1). These sequences are used to target, drive or terminate expression of the novel

[5] Defined as the global gene repertoire and consists of a core genome and an accessory genome (DE Maayer et al. 2014)

genetic material. None of the sources of these genetic elements is associated with toxic or allergenic responses in humans. The genetic elements derived from the plant pathogen *A. tumefaciens* are not pathogenic in themselves and do not cause pathogenic symptoms in GR2E.

3 Molecular characterisation

Molecular characterisation is necessary to provide an understanding of the genetic material introduced into the host genome and helps to frame the subsequent parts of the safety assessment. The molecular characterisation addresses three main aspects:

- the transformation method together with a detailed description of the DNA sequences introduced to the host genome
- a characterisation of the inserted DNA including any rearrangements that may have occurred as a consequence of the transformation
- the genetic stability of the inserted DNA and any accompanying expressed traits.

Studies submitted:

2016. Southern hybridization characterization of event IR-ØØGR2E-5 rice. Technical report, IR2015-07003 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines.
2015. Nucleotide sequence analysis of the inserted DNA and host genomic flanking regions in rice event IR-ØØGR2E-5. Technical report, IR2015-08001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines
2016. Stability of the elevated beta-carotene trait across multiple generations of rice event IR-ØØGR2E-5. Technical report, IR-2016-07001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines
2015. Segregation of the inserted DNA within multiple generations of rice event IR-ØØGR2E-5. Technical report, IR2015-08003 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

3.1 Method used in the genetic modification

The GR2E transformation procedure (Paine et al. 2005) was based on modifications of two previous protocols (Hiei et al. 1994; Zhang et al. 1997). Embryogenic cultures were established from mature embryos of cultivar Kaybonnet and embryogenic calli were incubated with a disarmed strain of *Agrobacterium tumefaciens* harbouring plasmid pSYN12424 (Figure 2). Following co-culture, the calli were then grown on medium containing ticarcillin, to inhibit the growth of excess *Agrobacterium*, and later transferred to mannose selection medium for five weeks in the dark. Proliferating colonies were transferred to regeneration medium. Shoots were grown in culture until large enough to be transferred to soil in a glasshouse facility.

Some 619 primary transformants were initially created. A series of quantitative polymerase chain reaction (PCR) analyses were used to retain events that were likely to contain a single copy of the T-DNA. Those events (103 in total) that could be grown on to produce at least 100 grains were then harvested and the grain was analysed for carotenoid content. Grains showing the highest carotenoid levels were selected for further assessment and development. GR2E was ultimately chosen as the lead event based on superior agronomic, phenotypic and molecular characteristics.

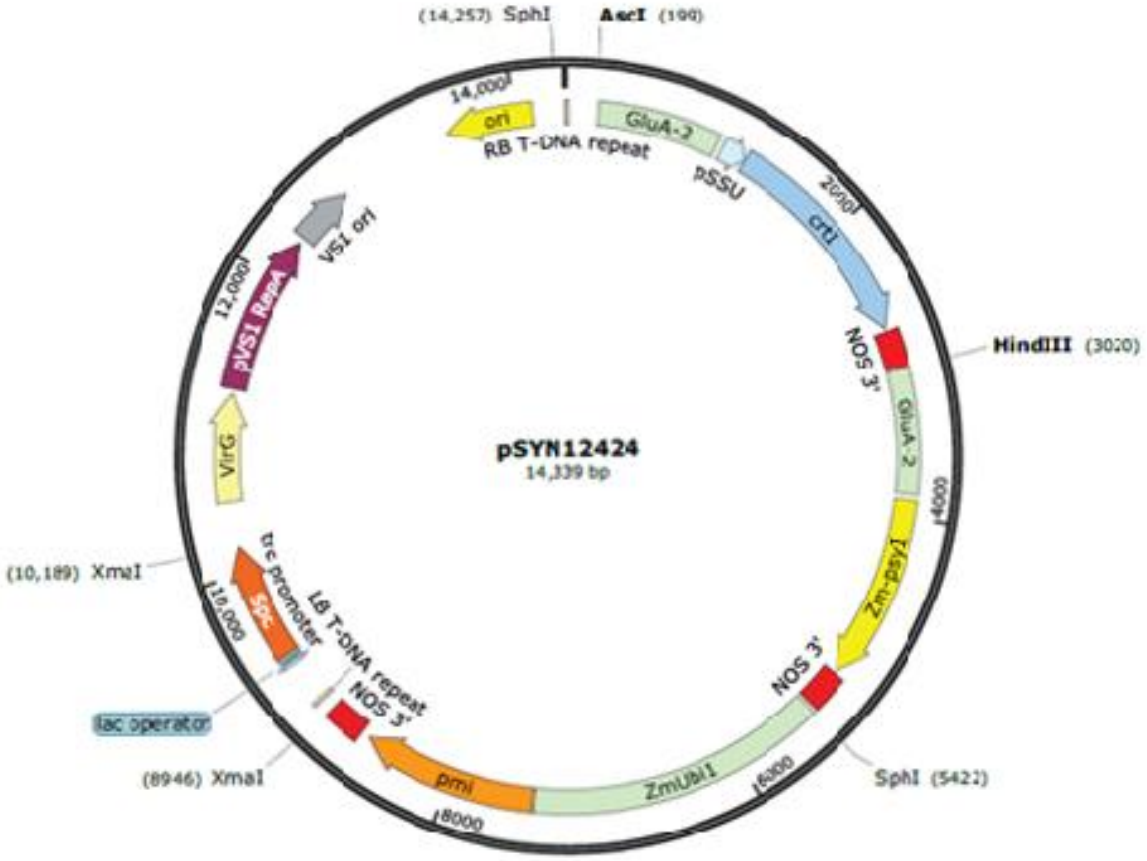

Figure 2: Genes and regulatory elements contained in plasmid pSYN12424

3.2 Function and regulation of introduced gene fragments

Information on the genetic elements in the T-DNA used for transformation is summarised in Table 1. The complete plasmid is 14,339 bp comprising 5,309 bp vector backbone and 9,030 bp T-DNA (Figures 2 & 3). The T-DNA contains three cassettes located between a 25 bp Left Border (LB) and 25 bp Right Border (RB) together with short non-coding regions of the *A. tumefaciens* Ti plasmid that help support transfer of the T-DNA into the host genome. Intervening sequences, where present, have assisted with the cloning of the various components of each cassette.

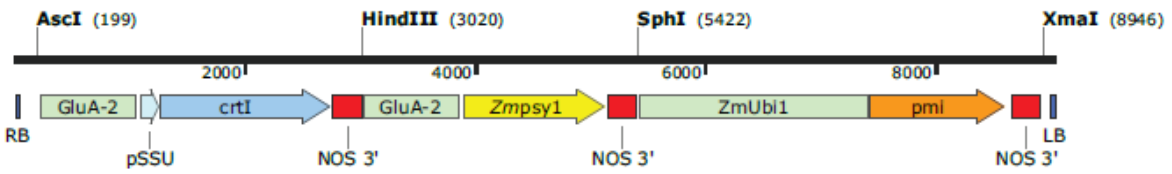

Figure 3: Design of pSYN12424 T-DNA region

Table 1: Description of the genetic elements contained in the T-DNA of pSYN12424

Genetic element	Relative bp location on plasmid	Size (bp)	Source	Orient.	Description &Function	Reference
Right Border	1 - 25	25	[1]			
Ti plasmid region	26 - 219	194	*Agrobacterium tumefaciens*			
***crtI* cassette**						
GluA-2	220 – 1047	828	*O.sativa*	clockwise	• Glutelin promoter (GenBank # D00584) • Targets expression of *crtI* to the endosperm • Directs transcription of the *crtI* gene	Takaiwa et al. (1987)
Intervening sequence	1048 – 1086	39				
pSSU	1087 - 1257	171	*Pisum sativum* (pea)	clockwise	• RUBISCO SSU transit peptide (GenBank # X00806) • Directs transport of CRTI to the plastids where the GGPP precursor is formed (see Figure 7)	Coruzzi et al. (1984); Beyer et al (2002)
crtI	1258 - 2736	1479	*P. ananatis*	clockwise	• Coding sequence of the *carotene desaturase* gene (GenBank # D90087)	Misawa et al (1990)
Intervening sequence	2737 - 2756	20				
NOS 3'	2757 - 3009	253	*Agrobacterium tumefaciens*	clockwise	• Terminator region from the *nopaline synthase gene* (GenBank # V00087) • Directs polyadenylation of the *crtI* gene	Depicker et al (1982)
Intervening sequence	3010 - 3025	16			Sequence used for DNA cloning	
***psy1* cassette**						
GluA-2	3026 - 3853	828	*O.sativa*	clockwise	• Glutelin promoter (GenBank # D00584) • Targets expression of *psyI* to the endosperm • Directs transcription of the *psyI* gene	Takaiwa et al. (1987)
Intervening sequence	3854 - 3892	39				
Zmpsy1	3893 - 5125	1233	*Zea mays* (corn)	clockwise	Coding sequence of the phytoene synthase gene (GenBank #U32636)	Buckner et al (1996)
Intervening sequence	5126 - 5145	20				
NOS 3'	5146 - 5398	253	*Agrobacterium tumefaciens*	clockwise	• Terminator region from the *nopaline synthase gene* (GenBank # V00087) • Directs polyadenylation of the *psy1* gene	Depicker et al (1982)
Intervening sequence	5399 - 5423	25				

Genetic element	Relative bp location on plasmid	Size (bp)	Source	Orient.	Description &Function	Reference
***pmi* cassette**						
ZmUBI	5424 - 7416	1993	*Zea mays* (corn)	clockwise	• Modified promoter region and first intron from polyubiquitin gene (GenBank # S94464) • Directs transcription of the *pmi* gene	Christensen et al (1992)
Intervening sequence	7417 - 7428	12				
pmi	7429 - 8604	1176	*Escherichia coli* (strain K12)	clockwise	• Coding sequence of the *phosphomannose isomerase* gene (GenBank # M15380)	Miles & Guest (1984)
Intervening sequence	8605 - 8664	60				
NOS 3'	8665 - 8917	253	*Agrobacterium tumefaciens*	clockwise	• Terminator region from the *nopaline synthase gene* (GenBank # V00087) • Directs polyadenylation of the *pmi* gene	Depicker et al (1982)
Ti plasmid region	8918 - 9005	88				
Left Border	9006 - 9030	25				

3.2.1 *crtI* cassette

Transcription of the *crtI* gene is controlled by the endosperm-specific rice *GluA-2 (glutelin)* promoter. Glutelin is the major protein in rice seed, accounting for approximately 80% of total endosperm protein (Takaiwa et al. 1987). Thus, the promoter ensures strong transcription in developing endosperm tissue. To further enhance effectiveness of the *crtI* gene, it is fused to the transit peptide sequence of the pea Rubisco small subunit (RUBISCO SSU) to ensure targeting to the endosperm plastids (Beyer et al. 2002). The plastids naturally contain the lycopene precursor geranylgeranyl diphosphate (GGPP) that is then used as a substrate to produce phytoene (see discussion and Figure 7 in Section 4.1); the CRTI protein catalyses the conversion of 15-*cis*-phytoene to all-*trans*-lycopene. Transcription termination of the *crtI* gene is provided by the polyadenylation signal and 3' untranslated region (UTR) from the *nopaline synthase* (*NOS)* gene of *A. tumefaciens* Ti plasmid pTiT37.

3.2.2 *Zmpsy1* cassette 2

The expression of the corn *psy1* gene (*Zmpsy1*) produces PSY1 that converts GGPP to 15-*cis*-phytoene in the carotenoid biosynthetic pathway (Section 4.1). A number of sources (daffodil, corn, tomato, capsicum, carrot, *Arabidopsis* and rice) of the *psy1* gene were tested during the development of the Golden Rice project (Paine et al. 2005) with *Zmpsy1* proving to be the most efficacious. The gene naturally incorporates a coding sequence for an N-terminal signal peptide that ensures targeting of the product to the plastids. Expression of the gene is under the control of the same promoter and terminator as the *crtI* gene.

3.2.3 *pmi* cassette

The *pmi* gene from *E. coli* encodes the PMI protein that catalyses the interconversion of mannose-6-phosphate and fructose-6-phosphate. The gene is not normally present in rice and its expression allows cells to survive on medium containing only mannose as a carbon

source. Thus, it is used to select putative transformants growing on mannose-containing medium (Negrotto et al. 2000). Gene expression is controlled by the constitutive polyubiquitin (*ubi*) promoter from corn. The *NOS* UTR terminates transcription.

3.3 Breeding of GR2E

Event GR2E in japonica Kaybonnet background (inbred, direct line of descent from the original T_0 transformant) was crossed into three different indica rice backgrounds, PSB Rc82 (Rc82), BRRI *dhan* 29 (BR29), and IR64 (Figure 4). During the breeding and development of GR2E rice, seed was produced for multiple purposes, including product evaluation, research testing, and regulatory testing. Due to seed limitations, seed was sourced from multiple generations of GR2E rice in different genetic backgrounds, depending on the study, and where applicable studies utilized the appropriate near-isogenic conventional comparator (Table 2).

Table 2: GR2E generations used for various analyses

Analysis	GR2E generation(s) used	Control(s) used	Reference material
Molecular characterisation (Sections 3.4.1.1 – 3.4.1.4)	T_n, BC_3F_5*^{1*2*3}, BC_4F_3*^{1*2*3}, BC_5F_3*^{1*2*3}	• Non-GM Kaybonnet • Non-GM Kaybonnet spiked with pSYN1242	Plasmid pSYN12424
Genetic stability (Section 3.5.1)	T_n, BC_3F_5*^{1*2*3}, BC_4F_3*^{1*2*3}, BC_5F_3*^{1*2*33}	• Non-GM Kaybonnet • Non-GM Kaybonnet spiked with pSYN1242	Plasmid pSYN12424
Phenotypic stability – carotenoid levels (Section 3.5.2.1)	T_n, BC_3F_5*^{1*2*3}, BC_4F_3*^{1*2*3}, BC_5F_3*^{1*2*3}	• Non-GM Kaybonnet • Non-GM Rc82	N/A
Mendelian inheritance (Section 3.5.2.2)	BC_4F_2*^{1*2*3}, BC_5F_1*^{1*2*3}, BC_5F_2*^{1*2*3}	N/A	N/A
Protein expression levels in plant parts (Section 4.1.3.2)	BC_5F_3*^{1}, BC_5F_4*^{1}	Non-GM Rc82	N/A
Compositional analyses (Section 5)	BC_5F_3*^{1}, BC_5F_4*^{1}	Non-GM Rc82	N/A

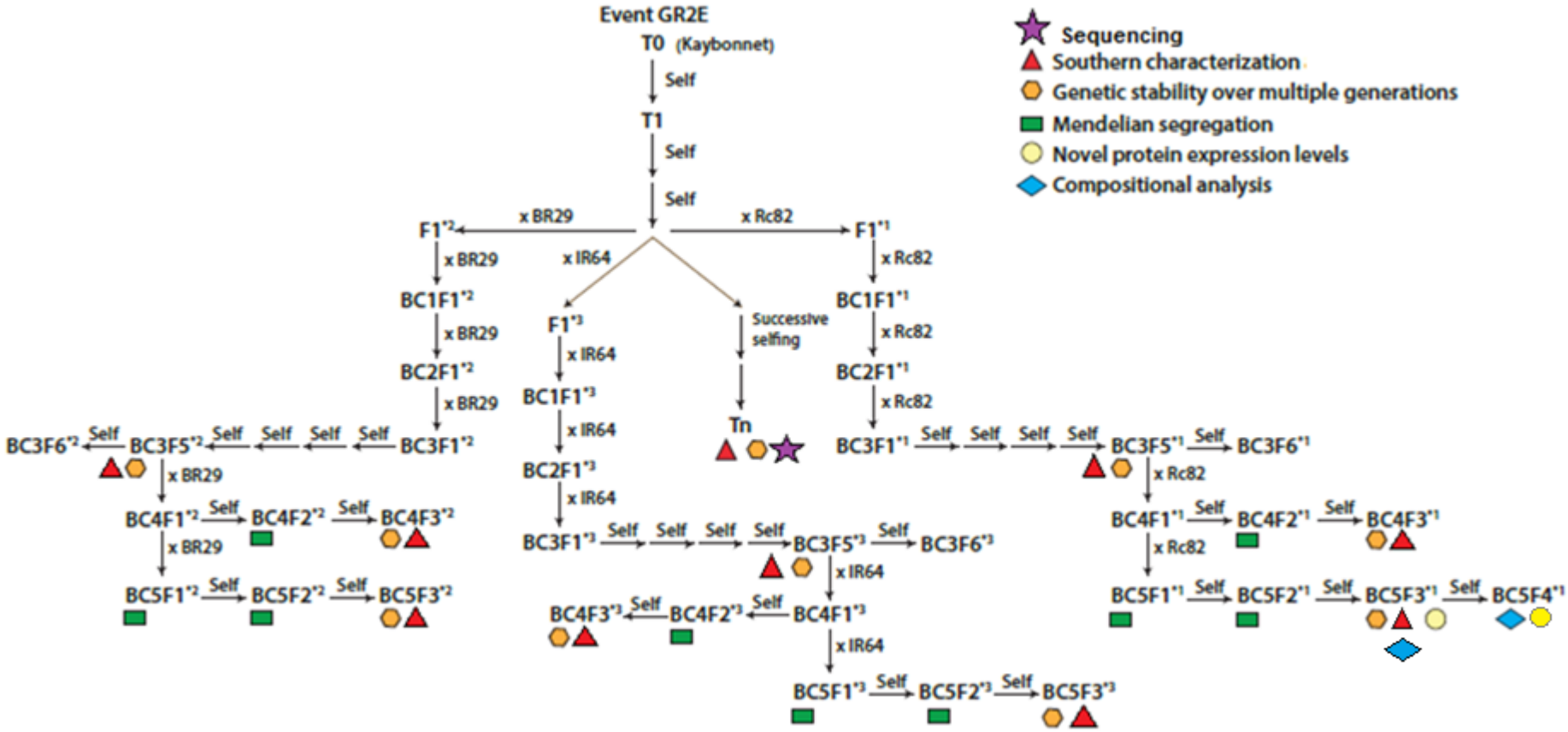

Figure 4: Breeding diagram for GR2E rice

3.4 Characterisation of the genetic modification in the plant.

A range of analyses were undertaken to characterise the genetic modification in GR2E. These analyses focussed on the nature of the insertion of the introduced genetic elements and whether any unintended re-arrangements may have occurred as a consequence of the transformation procedure.

3.4.1 Southern blot analysis: insert copy number, insert integrity and plasmid backbone

Genomic DNA isolation was performed as described by Murray & Thompson (1980) from pooled leaf samples from three glasshouse-grown plants containing the GR2E event from each of the four germplasm backgrounds (Figure 4 and Table 2), and from the unmodified parent (Kaybonnet negative control). Multiplex PCR zygosity testing was used to confirm the identity of the samples.

The DNA was then used to determine the number of insertion sites, and the integrity of the inserted T-DNA and to test for the presence or absence of plasmid vector backbone sequences by Southern blotting. For each characterisation, the DNA was first digested with appropriate restriction enzyme(s) before electrophoretic separation and visualisation following labelling with appropriate digoxigenin (DIG)-labelled probes. As a positive control, Kaybonnet DNA was spiked with the reference plasmid (pSYN12424) in an amount equivalent to 1.0 or 0.2 gene copies per rice genome.

3.4.1.1 Number of insertion sites

DNA was digested with two restriction enzymes (*Hin*dIII or *Sph*I) with unique recognition sites within the T-DNA (Figure 3). Following Southern blotting, three probes were used to cover each of the *Zmpsy1*, *SSU-crtI* and *pmi* genes.

In analysing the results, the insertion of a single copy of the T-DNA should have resulted in:

- a single detectable hybridization fragment of >3020 bp following *Hin*dIII digestion and hybridization with the *SSU-crtI* gene probe, and >6010 bp following hybridization with either the *Zmpsy1* or *pmi* probes.

- a single detectable hybridization fragment of >5422 bp following *Sph*I digestion and hybridization with the *SSU-crtI* or *psy1* probes, and >3608 bp following hybridization with the *pmi* probe.

Other considerations are:

- The presence of additional bands in GR2E but not in the negative control would indicate the presence of a second insert.

- As the native Kaybonnet genome contains a rice *psy1* gene, it would be expected there would be hybridisation between sequences in this and the *Zmpsy1* probe which share some 83% sequence identity.

For genomic DNA from event GR2E in all four genetic backgrounds, and for Kaybonnet spiked with pSYN1242, single hybridising fragments of the predicted sizes were detected for each restriction enzyme digest. For all samples, including the negative control, a weak hybridisation fragment was noted for the *Zmpsy1* probe and was attributed to the presence of the endogenous rice *psy1* gene. No other fragments were detected in the negative control.

The results are consistent with the presence of a single insertion site in event GR2E.

3.4.1.2 Insert integrity

The DNA was digested with a combination of the two restriction enzymes *AscI* plus *XmaI* and probed with the three probes covering each of the *Zmpsy1*, *SSU-crtI* and *pmi* genes. Since the restriction enzyme sites are located at the two ends of the T-DNA (Figure 3), an intact copy of the T-DNA would result in a single hybridisation fragment with all three probes. The expected size single fragment was obtained for DNA from all four backgrounds containing event GR2E, thereby confirming the presence of an intact T-DNA insert.

3.4.1.3 Plasmid backbone

Again, the DNA was digested with a combination of the two restriction enzymes *AscI* plus *XmaI* but was probed with five probes which together covered the entire plasmid backbone region. Only the Kaybonnet positive control spiked with pSYN1242 showed a hybridisation fragment. No hybridisation fragments were detected in any other samples. This indicated that no backbone sequences are present in event GR2E.

3.4.2 Insert organisation and sequence

Genomic DNA was extracted from leaf tissue obtained from a single GR2E rice progeny plant (generation T_n) derived from the original transformation event in Kaybonnet germplasm. The nucleotide sequence of pSYN12424 T-DNA, together with preliminary sequence information from the 5' and 3' flanking genomic DNA, was used to design seven sets of oligonucleotide primers that were used to amplify the insert and flanking regions as seven individual overlapping fragments. These fragments were then cloned into a vector and transformed into an *E. coli* strain. Three colonies from the bacterium were randomly selected, and the plasmid DNA isolated before dye-terminator nucleotide sequencing (ABI Big Dye® 3.1 terminator chemistry). The sequences from each clone were aligned (AlignX™) to obtain the final consensus sequence for each amplification fragment.

In total, 12,772 bp of GR2E sequence were confirmed, comprising 1,988 bp of 5' flanking region, 1,788 bp of 3' flanking region and 8,996 bp of inserted T-DNA. The inserted T-DNA was found to have a 23 bp deletion at the RB, and an 11 bp deletion at the LB. This truncation of the border sequences is not uncommon for *Agrobacterium*-mediated transformation events (Tzfira et al. 2004; Kim et al. 2007). All remaining sequence was intact and identical to that of the T-DNA region of plasmid pSYN12424.

3.4.3 Insertion site

Basic local alignment search tool (BLAST) searches using the 5' and 3' flanking region sequences as queries against the *O. sativa* (japonica cultivar-group, Nipponbare) genome (MSU Rice Genome Annotation Project Release 7)[6] identified that the site of insertion of the T-DNA was located on chromosome 3 within the intergenic region between LOC_Os03g43980 (3' proximal) and LOC_Os03g43990 (5' proximal) (Figure 5). The locations of the LB and RB flanking sequences correspond to positions 24,698,762–24,700,549 and 24,700,565–24,702,552 of the host genome, respectively. The insertion of the pSYN12424 T-DNA resulted in the deletion of 15 bp of host genomic DNA.

[6] Rice Genome Annotation Project - http://rice.plantbiology.msu.edu/

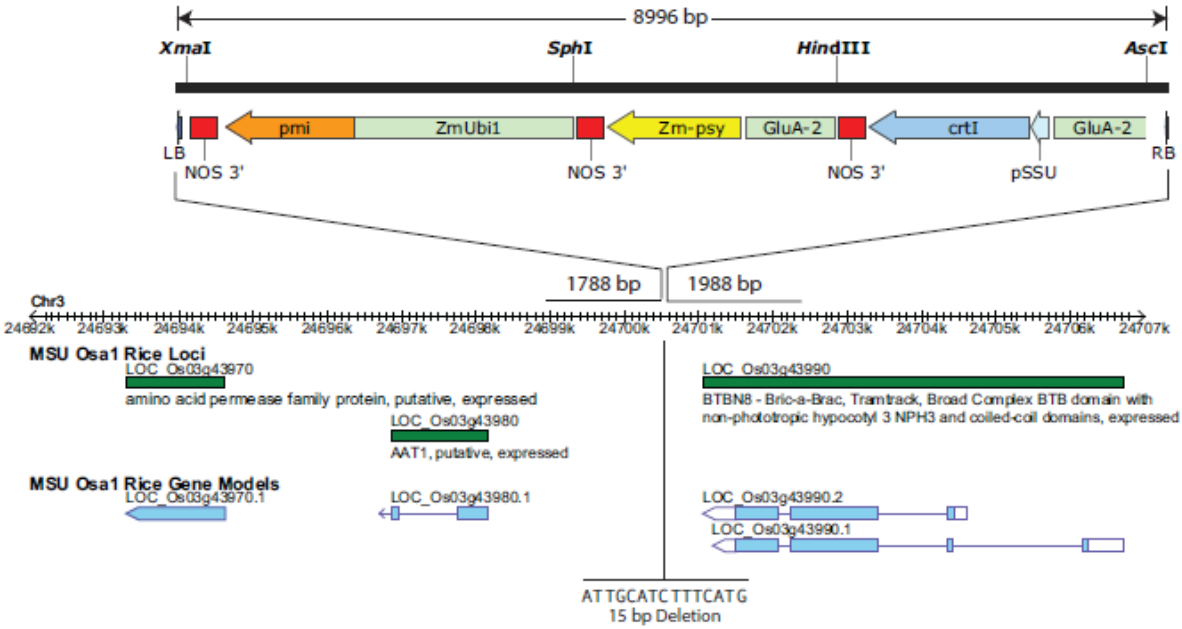

Figure 5: Representation of the map position of the T-DNA insert in the Kaybonnet genome

3.4.4 Open reading frame (ORF) analysis

Sequences spanning the 5′ and 3′ junctions of the insert in GR2E were translated from start codon to stop codon (TGA, TAG, TAA) in all six reading frames. Only those encoding sequences of 30 or more amino acids were considered; ORFs shorter than 29 amino acids were not evaluated since a minimum 35% identity requires at least a match of 29 amino acids over 80 amino acids. The 35% identity is a recommended criterion for indicating potential allergenicity (Codex 2009). Two ORFs were identified, one in the reverse orientation that spanned the 5' T-DNA insert–genomic DNA border (207 bp; 68 amino acids), and one in the forward orientation that spanned the 3' T-DNA insert–genomic DNA border (240 bp; 79 amino acids).These putative polypeptides were analysed using a bioinformatics strategy to determine similarity to known protein toxins or allergens (refer to Section 4.1.6)

3.5 Stability of the genetic change in GR2E

The concept of stability encompasses both the genetic and phenotypic stability of the introduced trait over a number of generations. Genetic stability refers to maintenance of the modification (as produced in the initial transformation event) over successive generations. It is best assessed by molecular techniques, such as Southern blot analysis. Phenotypic stability refers to the expressed trait remaining unchanged over successive generations. It is often quantified by a trait inheritance analysis to determine Mendelian heritability via assay techniques (chemical, molecular, visual).

3.5.1 Genetic stability

The genetic stability of event GR2E was evaluated by Southern blot analysis of genomic leaf DNA from verified plants from generations in the four background germplasms (see Figure 3). A negative control of DNA from non-GM Kaybonnet and a positive control of Kaybonnet spiked with pSYN12424 were also included. Restriction enzyme digestions with *Hin*dIII, *Sph*I, and *Asc*I+*Xma*I were separated by agarose gel electrophoresis and blots were probed with DIG-labelled DNA probes specific for the *Zmpsy1*, *pSSU-crtI*, or *pmi* genes, respectively. The same single hybridising fragments were detected across all GR2E samples thereby

confirming stable integration and inheritance of the inserted DNA in GR2E. No bands were obtained for the negative control; the expected band was obtained for the positive control.

3.5.2 Phenotypic stability

Phenotypic stability was assessed through two separate approaches.

Inheritance of elevated carotenoid levels

Dehulled and polished grain (harvested at physiological maturity) from each of the verified plant backgrounds/generations described in Section 3.4.1 (see also Figure 4 and Table 2) was ground and acetone-extracted for analysis of total carotenoids using UV/VIS spectral analysis from 800 – 350 nm. Due to limitations in grain availability the plant sample size for each analysis ranged from 1 – 5. Two non-GM controls were included – Kaybonnet and Rc82.

Although there was variation between the results from different genetic backgrounds (Table 3) there was consistency within the different generations from each background thereby demonstrating stable inheritance within each background.

Table 3: Concentrations of total carotenoids in different generations and germplasm backgrounds of GR2E

	Mean total carotenoids (µg/g fw)[1]			
Generation/background	Kaybonnet (GR2E)	Rc82 (GR2E)	IR64 (GR2E)	BR29(GR2E)
T_n	30.50 (1)			
BC_3F_5		12.84 ± 3.32 (2)	20.11 ± 8.71 (5)	No grain available
BC_4F_3		9.09 ± 2.26 (5)	19.40 ± 2.00 (5)	24.50 ± 2.85 (2)
BC_5F_3		14.12 (1)	13.23 ± 1.16 (2)	29.33 ± 3.14 (4)
Non-GM Kaybonnet	0.87 (1)			
Non-GM RC82	1.29 (1)			

[1]The number of plants sampled to provide each mean result is indicated in brackets

Mendelian segregation of inserted DNA

Leaf tissue samples were collected from plants derived from seed of the BC_4F_2, BC_5F_1, and BC_5F_2 generations of three genetic backgrounds (Rc82, BR29 and IR64) (Figure 4). Genomic DNA was isolated and then analysed by multiplex PCR amplification (using three primers – one forward and two reverse) to determine zygosity of the insert. Two primers were complementary to sequences in the 5' and 3' flanking host genomic regions close to the Right and Left borders, and the third primer was complementary to *pmi* gene sequences within the inserted T-DNA (Figure 6).

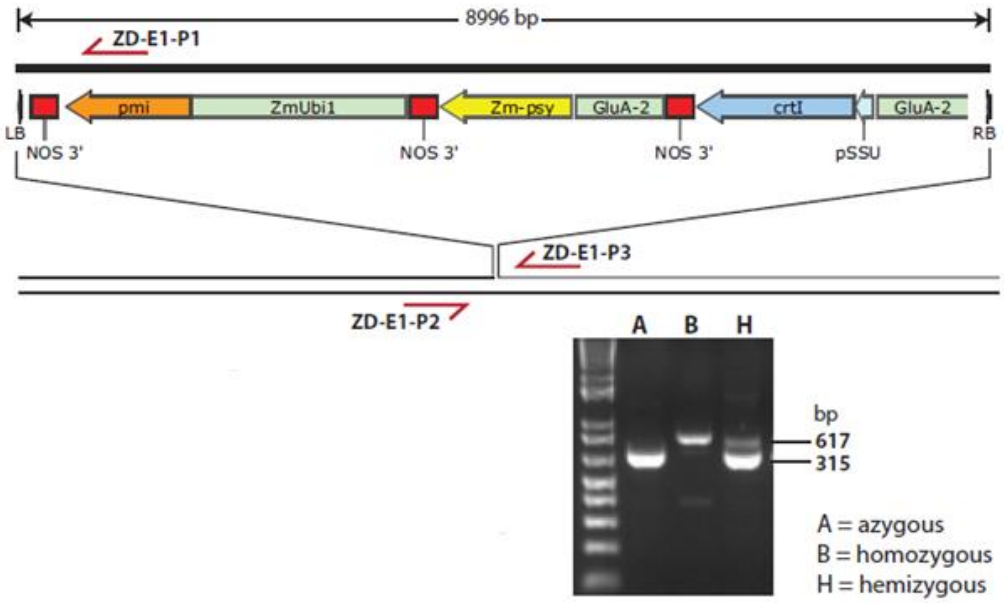

Figure 6: Representation of location of primers for multiplex PCR analysis and the three outcomes that could be obtained

In this analysis, since the Southern blots (Section 3.4.1) have indicated an intact insert has gone in to the GR2E event, if the insert is absent (azygous or null segregant) the product resulting from primers P2 and P3 will be very small (315 bp). If the insert is present the design of primers P2 and P1 is such that a larger product (617 bp) will be amplified either in the homozygous state (2 copies/chromosome – a single heavy band of 617 bp) or the hemizygous state (1 copy per chromosome – a single fainter band at 617 bp plus a single band at 315 bp). In scoring the results it is only absence or presence of the insert that is relevant.

The results of presence/absence of the insert are shown in Table 4. The chi-square (X^2) critical value at significance level $\alpha = 0.05$ is 3.84 i.e. if the X^2 value is < 3.84 the observed ratio is not significantly different from the expected ratio. The X^2 values for all generations were less than 3.84 thereby indicating that the insert in each genetic background was inherited according to Mendelian principles.

Table 4: Segregation of the T-DNA insert over three generations and backgrounds

Generation/Background	Total plants	Ratio presence:absence		X^2	Probability[1]
		Observed	Expected		
BC_4F_2*[1]	64	44:20	48:16	1.333	0.2482 (NS)
BC_4F_2*[2]	51	38:13	38.25:12.75	0.007	0.9356 (NS)
BC_4F_2*[3]	63	43:20	47.25:15.75	1.529	0.2162 (NS)
BC_5F_1*[1]	59	32:27	29.5:29.5	0.424	0.5151 (NS)
BC_5F_1*[1]	49	25:24	24.5:24.5	0.020	0.8864 (NS)
BC_5F_1*[1]	127	68:59	63.5:63.5	0.638	0.4245 (NS)
BC_5F_2*[1]	100	74:26	75:25	0.053	0.8174 (NS)
BC_5F_2*[1]	99	72:27	74.25:24.75	0.273	0.6015 (NS)
BC_5F_2*[1]	100	70:30	75:25	1.333	0.2482 (NS)

[1]NS = not significant - $P<0.05$ ($X^2<3.84$)

3.6 Antibiotic resistance marker genes

No antibiotic marker genes are present in GR2E. The insert sequence analysis (Section 3.4.1.3) showed no plasmid backbone sequences had been integrated into the GR2E genome during transformation.

3.7 Conclusion

GR2E was generated through *Agrobacterium*-mediated transformation with a single T-DNA containing three expression cassettes. Comprehensive molecular analyses indicate there is a single insertion site comprising a single, complete copy of each of the *Zmpsy1, SSU-crtI* and *pmi* genes with their regulatory elements. The introduced genes are stably inherited from one generation to the next.

There are no antibiotic resistance marker genes present in the line and no plasmid backbone has been incorporated into the transgenic locus.

4 Characterisation and safety assessment of new substances

The main purpose of the characterisation is to describe the nature of any new substances and their phenotypic and biochemical effects on the organism in which they are expressed, particularly in the parts of the organism consumed as food. Typically, the main focus of the characterisation is on newly expressed (or potentially expressed) proteins, but other (non-protein) substances may need to be considered.

4.1 Newly expressed proteins

In considering the safety of newly expressed proteins it is important to note that a large and diverse range of proteins are ingested as part of the normal human diet without any adverse effects, although a small number have the potential to impair health, e.g., because they are allergens or anti-nutrients (Delaney et al. 2008). As proteins perform a wide variety of functions, different possible effects have to be considered during the safety assessment including potential toxic, anti-nutritional and allergenic effects. To effectively identify any potential hazards requires knowledge of the characteristics, concentration and localisation of all newly expressed proteins in the organism as well as a detailed understanding of their biochemical function and phenotypic effects. It is also important to determine if the newly expressed proteins are expressed as expected, including whether any post-translational modifications have occurred.

Two types of proteins were considered:

- The proteins expected to be directly produced as a result of the translation of the introduced genes. A number of different analyses were done to characterise these proteins and determine their *in planta* expression.
- Those that may be potentially translated as a result of the creation of ORFs during the transformation process (see Section 3.4.4).

4.1.1 The *Zm*PSY1 and CRTI proteins

Carotenoids in general play an essential role in plants such as protecting the photosynthetic apparatus and in hormone signalling (Ruiz-Sola and Rodriguez-Concepcion 2012), but they also provide nutritional benefit to humans and animals; with only a few exceptions, animals

cannot synthesise carotenoids and therefore need to obtain them from the diet. Certain carotenoids, most importantly β-carotene, are cleaved to vitamin A within the body and are referred to as provitamin A (Yeum and Russell 2002). Vitamin A deficiency can result in permanent blindness and other disorders (McLaren and Kraemer 2012).

Rice leaves produce and accumulate β-carotene but milled rice (the usual form that is consumed in Asia) is characterised by the absence of provitamin A. The plastids within rice endosperm naturally contain the lycopene precursor geranylgeranyl diphosphate (GGPP) a key isoprenoid that is the precursor for entry into biosynthetic pathways for a wide range of compounds, not only carotenoids (e.g. gibberellins, chlorophylls, isoprenoid quinones), considered necessary for plant growth (Okada et al. 2000). Rice endosperm was thought to not naturally contain the four necessary enzymes (phytoene synthase, phytoene desaturase, ζ-carotene desaturase and lycopene β-cyclase) to facilitate the provitamin A biosynthetic pathway (Figure 7). The formation of phytoene is the first critical step in the pathway.

The Golden Rice project aimed to engineer the pathway by introducing three enzymes – phytoene synthase, lycopene β-cyclase, and a carotene desaturase (from a non-photosynthetic bacterial source that could substitute for both phytoene desaturase and ζ-carotene desaturase by introducing the required double bonds in one step). During project development, it was discovered that addition of lycopene β-cyclase was unnecessary for completion of the pathway to β-carotene, suggesting this enzyme is intrinsically expressed in sufficient quantity in non-GM rice (Schaub et al. 2005). Therefore in GR2E only two genes *psy1* from corn and *crtI* from *Pantoea ananatis* are necessary to drive the production of β-carotene in plastids of the endosperm.

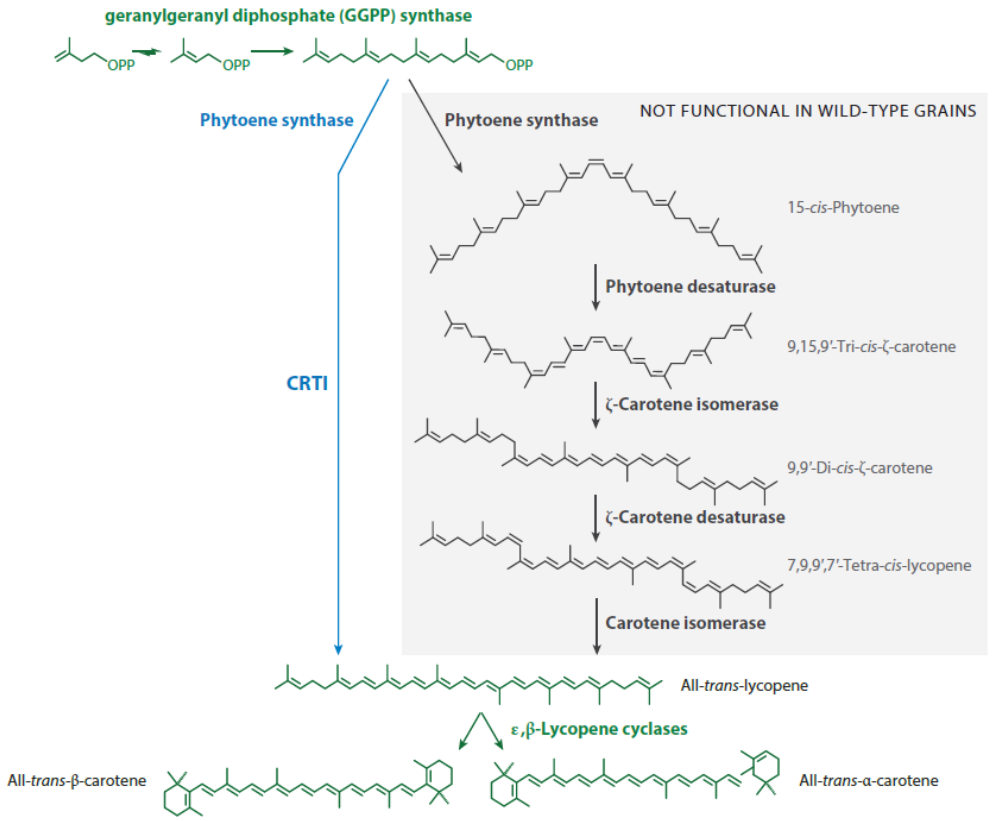

Figure 7: the β-carotene biosynthetic pathway
Steps in green denote activities naturally present in non-GM rice endosperm. Steps in grey are not functional in non-GM rice endosperm. Proteins in blue are expressed by the two exogenous genes present in GR2E.

Specifically, phytoene synthase catalyses the production of the 40-carbon phytoene from condensation of two GGPP molecules. Phytoene is then desaturated to create the chain of double bonds that forms the backbone of plant carotenoids. This desaturation proceeds via a series of four desaturation and two isomerisation reactions that transforms the phytoene into all-*trans* lycopene; in GR2E the single CRTI enzyme catalyses this entire process (Ruiz-Sola and Rodriguez-Concepcion 2012).

A translation of the DNA sequence of the *Zmpsy1* gene in GR2E (see Section 3.4.2) yielded a protein comprising 410 amino acids (Figure 8), with a calculated molecular weight of 46.5 kDa. Using the ChloroP Transit Peptide Predictor[7] (Emanuelsson et al. 1999), the *Zm*PSY1 protein expressed in MS11 is predicted to have a 62-residue N-terminal transit peptide that is processed to a 39.8 kDa (348 residues) mature plastid protein.

```
          10         20         30         40         50         60
MAIILVRAASPGLSAADSISHQGTLQCSTLLKTKRPAARRWMPCSLLGLHPWEAGRPSPA

          70         80         90        100        110        120
VYSSLPVNPAGEAVVSSEQKVYDVVLKQAALLKRQLRTPVLDARPQDMDMPRNGLKEAYD

         130        140        150        160        170        180
RCGEICEEYAKTFYLGTMLMTEERRRAIWAIYVWCRRTDELVDGPNANYITPTALDRWEK

         190        200        210        220        230        240
RLEDLFTGRPYDMLDAALSDTISRFPIDIQPFRDMIEGMRSDLRKTRYNNFDELYMYCYY

         250        260        270        280        290        300
VAGTVGLMSVPVMGIATESKATTESVYSAALALGIANQLTNILRDVGEDARRGRIYLPQD

         310        320        330        340        350        360
ELAQAGLSDEDIFKGVVTNRWRNFMKRQIKRARMFFEEAERGVTELSQASRWPVWASLLL

         370        380        390        400        410
YRQILDEIEANDYNNFTKRAYVGKGKKLLALPVAYGKSLLLPCSLRNGQT
```

Figure 8: Deduced amino acid sequence for the ZmPSY1 protein expressed in GR2E
The transit peptide that is cleaved off during processing is underlined in red.

The CRTI protein encoded in GR2E belongs to the class of enzymes known as flavoenzymes. A translation of the DNA sequence of the *crtI* gene in GR2E (see Section 3.4.2) yielded a protein comprising 492 amino acids (Figure 9), with a calculated molecular weight of 55 kDa

```
          10         20         30         40         50         60
MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYEDQGFTFDAGPTVITDP

          70         80         90        100        110        120
SAIEELFALAGKQLKEYVELLPVTPFYRLCWESGKVFNYDNDQTRLEAQIQQFNPRDVEG

         130        140        150        160        170        180
YRQFLDYSRAVFKEGYLKLGTVPFLSFRDMLRAAPQLAKLQAWRSVYSKVASYIEDEHLR

         190        200        210        220        230        240
QAFSFHSLLVGGNPFATSSIYTLIHALEREWGVWFPRGGTGALVQGMIKLFQDLGGEVVL

         250        260        270        280        290        300
NARVSHMETTGNKIEAVHLEDGRRFLTQAVASNADVVHTYRDLLSQHPAAVKQSNKLQTK

         310        320        330        340        350        360
RMSNSLFVLYFGLNHHHDQLAHHTVCFGPRYRELIDEIFNHDGLAEDFSLYLHAPCVTDS

         370        380        390        400        410        420
SLAPEGCGSYYVLAPVPHLGTANLDWTVEGPKLRDRIFAYLEQHYMPGLRSQLVTHRMFT

         430        440        450        460        470        480
PFDFRDQLNAYHGSAFSVEPVLTQSAWFRPHNRDKTITNLYLVGAGTHPGAGIPGVIGSA

         490
KATAGLMLEDLI
```

Figure 9: Deduced amino acid sequence of the CRTI protein expressed in GR2E

[7] ChloroP - http://www.cbs.dtu.dk/services/ChloroP/

4.1.2 PMI protein

Plants transformed with the *pmi* gene can utilise mannose as a source of carbon. The expression of the PMI enzyme in plant cells is therefore useful as a selectable marker to assist with identification of transformed cells. Phosphomannose isomerase catalyses the interconversion of mannose 6-phosphate and fructose 6-phosphate and is present in some plants. Plants lacking this enzyme are unable to survive on culture media containing mannose.

Mannose, a hexose sugar, is taken up by plants and converted to mannose-6-phosphate by hexokinase. This product cannot be further utilised in many plants as they lack the PMI enzyme. The accumulation of mannose-6-phosphate inhibits phosphoglucose isomerase, causing a block in glycolysis. It also depletes cells of orthophosphate required for the production of ATP. Therefore, while mannose has no direct toxicity on plant cells, it causes growth inhibition (Negrotto et al. 2000).

The enzyme is required in many eukaryotic systems including humans (Proudfoot et al. 1994a; Proudfoot et al. 1994b). It has been assessed by FSANZ previously as a novel protein in corn lines MIR604 (FSANZ 2006), MIR162 (FSANZ 2008a), 3272 (FSANZ 2008b) and 5307 (FSANZ 2012). A translation of the DNA sequence of the *pmi* gene in GR2E (see Section 3.4.2) yielded a protein whose sequence is identical to that expressed in three of the previously assessed lines - MIR162, line 3272 and line 5307[8] (Figure 10).

```
  1   MQKLINSVQN YAWGSKTALT ELYGMENPSS QPMAELWMGA HPKSSSRVQN
 51   AAGDIVSLRD VIESDKSTLL GEAVAKRFGE LPFLFKVLCA AQPLSIQVHP
101   NKHNSEIGFA KENAAGIPMD AAERNYKDPN HKPELVFALT PFLAMNAFRE
151   FSEIVSLLQP VAGAHPAIAH FLQQPDAERL SELFASLLNM QGEEKSRALA
201   ILKSALDSQQ GEPWQTIRLI SEFYPEDSGL FSPLLLNVVK LNPGEAMFLF
251   AETPHAYLQG VALEVMANSD NVLRAGLTPK YIDIPELVAN VKFEAKPANQ
301   LLTQPVKQGA ELDFPIPVDD FAFSLHDLSD KETTISQQSA AILFCVEGDA
351   TLWKGSQQLQ LKPGESAFIA ANESPVTVKG HGRLARVYNK L
```

Figure 10: Deduced amino acid sequence of the PMI protein expressed in GR2E

4.1.3 Protein expression in the tissues of GR2E

Studies submitted:

2016. Western immunoblot analysis of *Zm*PSY1, CRTI, and PMI expression in different plant tissues from rice event IR-ØØGR2E-5. Technical report, IR2016-04003 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Concentrations of phytoene synthase (*Zm*PSY1), phytoene desaturase (CRTI), and phosphomannose isomerase (PMI) in grain and straw harvested from event IR-ØØGR2E-5 rice grown in the Philippines in 2015. Technical report, IR2015-08004 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Concentrations of phytoene synthase (*Zm*PSY1), phytoene desaturase (CRTI), and phosphomannose isomerase (PMI) in grain and straw harvested from event IR-ØØGR2E-5 rice grown in the Philippines in 2016. Technical report, IR2016-04004 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

[8] As a consequence of the genetic modification resulting in event MIR604, two nucleotide sequence changes occurred within the *pmi* gene resulting in two amino acid changes; a valine to alanine substitution at position 61 and a glutamine to histidine substitution at position 210 (FSANZ 2006). Because of these changes, the amino acid sequence of the PMI protein in GR2E rice was not identical to the corresponding sequence of the PMI protein expressed in MIR604 maize.

Tissue specificity of protein expression

Western immunoblot analysis was used to confirm the tissue specificity of *Zm*PSY1, CRTI, and PMI expression in GR2E rice. Samples of grain (at three stages – milk, dough and mature[9]), bran, and hulls were obtained from a confined field trial grown during 2015 while stem, leaf, and root tissue were obtained from shade-house grown plants. The negative control substance consisted of dough-stage grain obtained from non-transgenic unmodified Kaybonnet rice grown under the same conditions as the test plants while the positive control was unmodified Kaybonnet spiked with purified samples of *Zm*Psy1, CRTI and PMI obtained from a recombinant *Escherichia coli* system (refer to Section 4.1.4.1).

Samples were extracted and the total proteins were separated by sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE). Following membrane transfer, the blots were incubated as follows:

- for *Zm*PSY1 – mouse monoclonal primary anti-*Zm*PSY1 antibody followed by alkaline phosphatase (AP)-conjugated goat anti-mouse IgG secondary antibody.
- for CRTI - mouse monoclonal primary anti-CRTI antibody followed by alkaline AP-conjugated goat anti-mouse IgG secondary antibody.
- for PMI – horseradish peroxidase (HRP)-conjugated rabbit anti-PMI antiserum (Romer Labs) followed by chemiluminescent substrate development using luminol substrate (GE Healthcare).

There was no detection of any of the proteins in the Kaybonnet negative control samples. Expression of *Zm*PSY1 and CRTI was detected only in milk, dough, and mature stage grain from GR2E plants and not in samples of bran, hulls, leaf, stem, or root tissue. In comparison, PMI expression was detected in all GR2E rice tissue types tested. These results were consistent with the use of the endosperm-specific promoter for the *Zmpsy1* and *crtI* genes, and the constitutive *Ubi* promoter for the *pmi* gene.

Protein expression in the tissues of GR2E

Verified (multiplex PCR) plants of GR2E in the Rc82 background (generation BC_5F_3*[1] and BC_5F_4*[1] – see Figure 4) together with a near-isogenic non-transgenic control, Rc82, were grown at four locations in the Philippines[10] on the largest island of Luzon (Figure 11) during the rainy season in 2015 and the dry season in 2016. These plantings overlapped with those used for the rainy and dry season compositional analyses described in Section 5.2. Grain was collected for analysis at three stages – milk, dough and mature; analysis of levels in straw at maturity was also done.

[9] The rice grain filling stages are separated into the milk, dough and physiological maturity stages. The milk stage is observed when a milky white substance begins to accumulate, usually seven to 10 days after the panicle has emerged from the stem. The dough stage occurs about a week later as the milky substance begins to change and become the texture of bread dough. When rice grains first become firm, they are at the physiological maturity stage (Dunand and Saichuk 2014)

[10] Batac City (Ilocos Norte), Los Baños (Laguna), Muñoz (Nueva Ecija), San Mateo (Isabela)

Figure 11: GR2E growing locations for protein expression and compositional analyses

Three blocks (replicates) of each entry (sample type) were established at each test site in a randomized complete block design. Each entry was planted in 10-row plots of 5 meters in length (20 cm × 20 cm spacing) for a total plot area of 10 m^2. There were 25 plants per row, totalling 250 plants per entry per plot. For each tissue type, a grain sample was collected for each of the three replicated blocks. Samples of rice straw were collected at harvest; each individual sample was a composite of material obtained from at least five representative plants within each block.

Levels of each protein were determined for each sample type using a validated enzyme linked immunosorbent assay (ELISA) employing a commercially prepared kit as follows:

- ZmPSY1 - Quantiplate™ Kit for PSY (EnviroLogix)
- CRTI - Quantiplate™ Kit for CRT1 (EnviroLogix)
- PMI - AgraQuant™ PMI Plate (Romer Labs)

A standard curve was prepared for each ELISA, and sample values were interpolated from this. Sample concentrations (ng/ml) were converted to amounts per weight of tissue homogenized for each sample type and analyte, and expressed as ng protein per gram fresh weight tissue (Tables 5 and 6

Table 5: *Zm*Psy1, CRTI and PMI protein content in various tissues of GR2E (averaged across four sites) grown in the 2015 rainy season

Sample type	*Zm*PSY1 (ng/g fw)		CRTI (ng/g fw)		PMI (ng/g fw)	
	Mean	Range	Mean	Range	Mean	Range
Grain (milk stage)	156.6	126.4 – 180.0	43.6	23.3 – 52.2	1548.1	1096.5 – 1797.5
Grain (dough stage)	338.6	326.3 – 358.9	61.8	55.3 – 67.6	1897.9	1564.5 – 2196.5
Grain (mature stage)	226.6	196.7 – 244.8	26.1	23.0 – 29.7	1369.0	915 – 1890.5
Straw	<LOQ[1]	<LOQ	<LOQ	<LOQ	466.7	320 – 617.5

[1] LOQ = Limit of quantification, which for the *Zm*PSY1 ELISA in straw (stem) samples was 66.4 ng/g FWT and for CRTI in straw samples was 6.2 ng/g fw.

Table 6: ZmPsy1, CRTI and PMI protein content in various tissues of GR2E (averaged across four sites) grown in the 2016 dry season

Sample type	*Zm*PSY1 (ng/g fw)		CRTI (ng/g fw)		PMI (ng/g fw)	
	Mean	Range	Mean	Range	Mean	Range
Grain (milk stage)	161.7	138.5 – 184.6	36.6	31.4 – 40.0	1828.4	1352 – 2083
Grain (dough stage)	328.9	308.2 – 356.4	60.3	53.8 – 68.0	2130.7	1912 – 2397
Grain (mature stage)	220.2	195.7 – 239.6	22.8	16.6 – 27.4	1195	765 – 1780
Straw	<LOQ[1]	<LOQ	<LOQ	<LOQ	498.0	339 – 796

[1] LOQ = Limit of quantification, which for the *Zm*PSY1 ELISA in straw (stem) samples was 55.4 ng/g FWT and for CRTI in straw samples was 1.94 ng/g fw.

Measurable but low concentrations of both PSY1 and CRTI proteins were found in all GR2E grain developmental stages but not in GR2E stem/straw tissue. Levels of CRTI were lower than *Zm*PSY1; both proteins had the highest mean concentration in grain at dough stage, ranging between *ca*. 308–359 ng/g and *ca*. 54–68 ng/g for *Zm*PSY1 and CRTI, respectively, across both growing seasons. Concentrations of PMI protein were significantly higher than either *Zm*PSY1 or CRTI in samples from all grain growth stages and were highest in dough-stage grain, averaging *ca*. 2015 ng/g across the four locations over both growing seasons.

Concentrations of *Zm*PSY1, CRTI, and PMI were below the limit of quantification in all tissue samples collected from control non-transformed Rc82 rice.

4.1.4 Characterisation of the proteins produced in GR2E

In many cases, insufficient amounts of newly-expressed proteins for safety evaluations (e.g. pepsin digestibility and heat stability studies, acute oral toxicity testing and use as a calibration standard for quantitative ELISA) are obtained from the plant. A standard practice in such cases is to produce the proteins for these evaluations in a microbial system and directly confirm their equivalence to the plant-produced proteins in simultaneous analyses. Hence, individual studies will concurrently determine both the structural and functional characteristics of the plant-purified protein as well as the similarity, to the plant protein, of the equivalent microbially-produced protein.

Studies submitted

2016. Characterization of CRTI protein (Lot Number M20454-02) derived from a microbial expression system. Technical report, IR2016-02004 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines
2016. Characterization of CRTI protein (Lot Number M20603) derived from a microbial expression system. Technical report, IR2016-03002 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines
2016. Characterization of ZmPSY1 protein (Lot Number M20452-05) derived from a microbial expression system. Technical report, IR2016-02005 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

ZmPSY1 and CRTI

Both the *Zm*PSY1 and CRTI proteins are peripheral membrane associated proteins, and as such, are difficult to work with. This, combined with the very low expression levels in GR2E rice grains made them intractable to standard methods or purification (Bushey et al. 2014) meaning it was not possible to obtain sufficient quantities for characterisation.

Therefore, a weight of evidence approach had to be used to assess the equivalence of the plant proteins with the microbially-produced proteins and hence to indirectly characterise the plant-produced proteins. It is noted, however, that the sequence analysis of the GR2E insert confirmed the expected sequences of both the *Zmpsy1* and *crtI* genes. From these, the expected protein sequences were translated (see Section 4.1.1). It is also noted that, since β-carotene was produced in GR2E (see Section 5.3.5) the expected phenotype was achieved and hence that functional *Zm*PSY1 and CRTI proteins had been expressed.

The two proteins were produced as fusion proteins[11] in *Escherichia coli.* The confirmed coding sequences of the *Zmpsy1* and *crtI* genes were separately transformed into *E. coli.* The expressed proteins were then purified by gel filtration chromatography and pooled fractions were concentrated by diafiltration.

The *Zm*PSY1 fusion protein purified from recombinant *E. coli* (Figure 12) differed from the ZmPSY expressed in GR2E (Figure 8). The fusion protein comprised a 386 residue protein with a 25 residue leader sequence containing an N-terminal $(His)_6$-tag sequence and a Factor Xa cleavage site IEGR. This molecular weight is slightly higher than that of the predicted mature plant protein (39.8 kDa – see Section 4.1.1).

[11] Fusion proteins had to be used because of the difficulty of expressing membrane proteins in either prokaryotic or eukaryotic systems. The His tag assists in purification using affinity chromatographic methods. Since it is difficult to remove the tag following purification it is essential the tag does not affect functional activity if the proteins are used in safety studies.

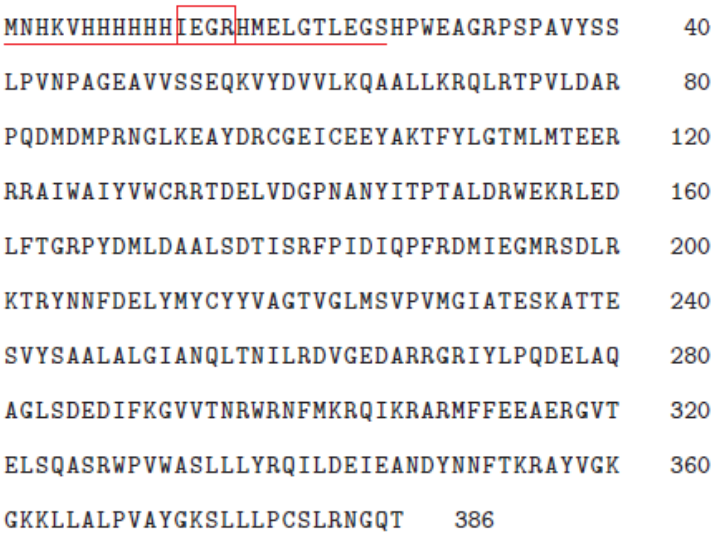

Figure 12: Deduced amino acid sequence of the microbially-produced ZmPSY1 protein
The N-terminal leader sequence containing the $(His)_6$-tag and the factor Xa cleavage site (IEGR) is underlined.

The CRTI fusion protein purified from recombinant *E. coli* (Figure 13) differed from the CRTI expressed in GR2E (Figure 9). The fusion protein comprised a full-length version of the CRTI protein with a GGS linker and C-terminal $(His)_6$-tag . Thus, the microbially-produced protein contains 501 amino acids and has a calculated molecular weight of 56.0 kDa.

Figure 13: Deduced amino acid sequence of the microbially-produced CRTI protein
The C-terminal $(His)_6$-tag and linker is underlined.

The only analyses in which it was possible to directly confirm the equivalence of the microbially-produced proteins was through SDS-PAGE and western blot as described in Section 4.1.3.1. It is noted in the western blots for all three proteins that the positive control (non-GM Kaybonnet spiked with the *E. coli*-produced protein) appeared to have slightly greater mobility than the plant proteins (from all three grain stages). In the case of the ZmPSY1 protein, this is most likely because the microbial fusion protein has a slightly higher

molecular weight than the plant protein. However, for all three proteins, it is likely the mobility of the plant proteins was slightly impeded by the sample matrix. Nevertheless, the apparent differences are within the margin of error of SDS-PAGE gels and, importantly, the proteins from plant and microbial sources are detected by the same specific antibodies thereby confirming their equivalence.

A number of techniques were used to analyse the microbially-produced proteins; for CRTI two separate lot numbers were analysed:

- SDS-PAGE and western blot analysis.
  Samples of each purified protein were analysed by SDS-PAGE followed by staining with SYPRO Orange[12]. Since both proteins contain a $(His)_6$ tag, western immunoblot analysis utilised a rabbit anti$(His)_6$ primary antibody followed by HRP-conjugated goat anti-rabbit secondary antibody.
- Reverse phase high performance liquid chromatography (HPLC) analysis.
  This was used to confirm the purity of the *E. coli*-derived *Zm*PSY1 and CRTI. The proteins were analysed on a Zorbax 300SB-C3 column, eluted and monitored at 214 nm.
- Relative amino acid analysis.
  Following hydrolysis, each protein was analysed by ion exchange chromatography for the relative presence (percentage of total amino acids) of the 20 most common amino acids (excluding tryptophan and cysteine that are not amenable to analysis under the conditions). The measured values were then compared to the predicted values based on the theoretical amino acid content of each protein.
- Matrix-assisted laser desorption/ionisation time of flight tandem mass spectrometry (MALDI-TOF MS/MS) mapping.
  For ZmPSY1 and CRTI, both the carbamidomethylated and trypsin-digested proteins were analysed. The resulting peptides were applied to an anchorchip target for analysis by MALDI-TOF. Mass spectrometry was then performed on a number of peptides for fragmentation analysis and the spectra were used for database searching (UniProt[13] and National Center for Biotechnology Information[14] - NCBI) against the predicted sequences for each protein using Mascot[15] software.
- N-terminal amino acid sequencing.
  Samples of the two purified proteins were immobilised on polyvinylidene fluoride (PVDF) membranes and subjected to N-terminal amino acid sequencing by Edman degradation chemistry (amino acid residues are cleaved off one at a time and identified by high performance liquid chromatography - HPLC).
- Enzymatic activity.
  Measurement of the activity of *Zm*PSY1 was based on the ability of the enzyme to catalyse (at 20^{o} C for 30 min) the formation of 15-*cis*-phytoene (phytoene) from GGPP in an appropriate reaction mixture containing the GGPP precursors - DMAPP and IPP - and the enzyme necessary to drive the formation of GGPP from these precursors - GGPP synthase (purified from *Arabidopsis thaliana*) (Figure 14). The phytoene was analysed following chloroform extraction, elution and detection using HPLC and a photodiodearray detector (287 nm).

  The activity of CRTI was measured by mixing the enzyme with a preparation of phytoene-containing liposomes and spectrophotometrically measuring the production of all-*trans*-lycopene (refer to Figure 7) after 30 min at 37^{o} C.

[12] SYPRO Orange is a sensitive, ready-to-use fluorescent stain for proteins in 1D gels. It is an alternative to the more traditional Coomassie Blue staining (Simpson 2010)
[13] http://www.uniprot.org/
[14] https://www.ncbi.nlm.nih.gov/
[15] http://www.matrixscience.com/

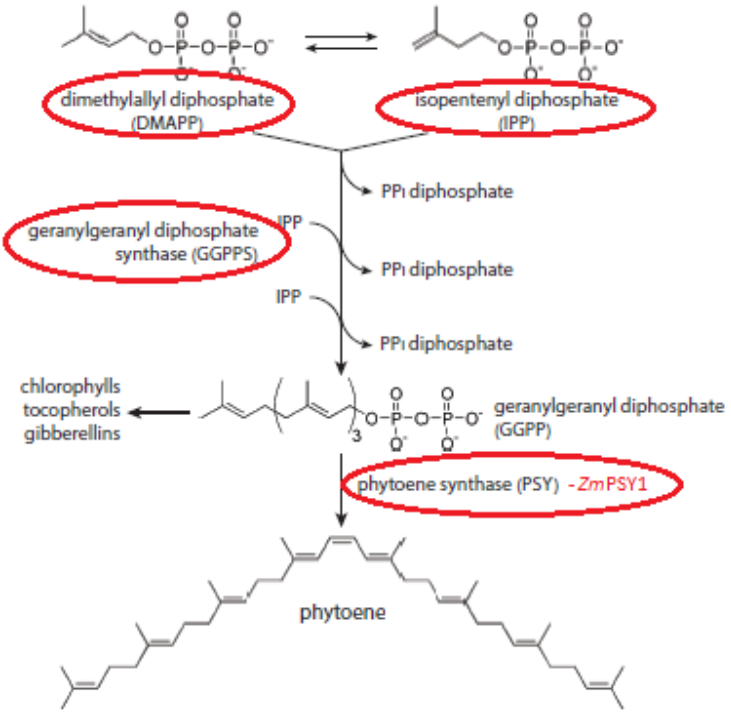

Figure 14: The phytoene biosynthetic pathway
Products circled in red are those added to the reaction mixture used for measuring the enzymatic activity of *Zm*PSY1

SDS-PAGE and western blot

Both SDS-PAGE and western blotting indicated that the apparent molecular weight of the microbially-produced *Zm*PSY1 protein was *ca*. 42 kDa; based on densitometric analysis, the protein was > 95% pure. Similarly, SDS-PAGE and western blotting confirmed that the apparent molecular weight of the microbially-produced CRTI protein was *ca*. 50 kDa (both batches); based on densitometric analysis the protein was >90% pure. These apparent molecular weights are consistent with the predicted molecular weights of the two proteins.

Reverse phase HPLC

This analysis showed the *Zm*PSY1 protein to be 93.7% pure, and the CRTI protein to be 91.6% pure and 84.9% pure in the two batches analysed.

Relative amino acid analysis

Overall, there was good agreement between the theoretical and measured amino acid compositions of *Zm*PSY1 and CRTI (both batches).

MALDI-TOF MS/MS mapping

A protein identification made by peptide mass fingerprinting is considered to be reliable if the measured coverage of the sequence is 15% or higher with a minimum of five unique peptide matches with those expected from a specified enzyme digestion of the theoretical protein (Jensen et al. 1997).

Analysis of the trypsin-digested *Zm*PSY1 protein identified 13 peptide fragments covering 48% of the amino acid sequence. MS/MS partial sequencing of 7 peptide fragments was

used to unambiguously identify the microbial-expressed *Zm*PSY1 as corresponding to *Zea mays* phytoene synthase (GenBank Accession no. AAB60314)).
Analysis of the first batch of trypsin-digested CRTI protein identified 16 peptide fragments covering 39% of the amino acid sequence. MS/MS partial sequencing of 11 peptide fragments was used to unambiguously identify the microbial-expressed CRTI as corresponding to CRTI_PANAN (UniProt entry P21685). Analysis of the second batch of trypsin-digested CRTI protein identified 28 peptide fragments covering 55% of the amino acid sequence. MS/MS partial sequencing of 11 peptide fragments was used to unambiguously identify the microbial-expressed CRTI as corresponding to CRTI_PANAN (UniProt entry P21685).

N-terminal sequencing

For *Zm*PSY1, the expected sequence was MNHKVHHHHHH (see Section 4.1.1) but each cycle of residue analysis by HPLC yielded multiple residue possibilities; this was likely due to some limited N-terminal proteolysis during production or purification of the protein. This precluded a definitive N-terminal sequence being obtained.

For both batches of CRTI, the predominant amino terminal sequence in the CRTI protein samples was MKPTTVIGAG, which corresponds to that of the CRTI N-terminus (see Section 4.1.1).

Enzymatic activity

E.coli-expressed *Zm*PSY1 catalysed the production of 15-*cis*-phytoene from DMAPP and IPP, in the presence of active *A. thaliana* GGPP synthase, at the rate of *ca*. 28.4 pmol µg−1 min−1 under the assay conditions used.

E.coli-expressed CRTI catalysed the conversion of liposome-incorporated phytoene to all-*trans*-lycopene at the rate of *ca*. 5.4 pmol (one batch) and *ca*. 2.02 pmol (second batch) all-*trans*-lycopene μg^{-1} min^{-1} under the assay conditions used in this study.

These results show that the two *E. coli*-expressed proteins both possess the expected activity.

PMI

Since the DNA sequence of the *pmi* gene is the same as that used for the genetic modification of lines previously assessed by FSANZ it was not necessary for characterisation of the PMI protein, as expressed in GR2E, to be done, other than to confirm that the amino acid sequence was the same as that in previously assessed lines expressing the same protein (see Section 4.1.2).

4.1.5 Safety of the introduced proteins

Data were provided to assess the potential toxicity and allergenicity of the proteins expressed in GR2E.

Potential toxicity of ZmPSY1 and CRTI

While the vast majority of proteins ingested as part of the diet are not typically associated with toxic effects, a small number may be harmful to health. Therefore, if a GM food differs from its conventional counterpart by the presence of one or more novel proteins, these proteins should be assessed for their potential toxicity. The main purpose of an assessment

of potential toxicity is to establish, using a weight of evidence approach, that the novel proteins will behave like any other dietary protein.

The assessment focuses on:

- whether the novel proteins have a prior history of safe human consumption, or are sufficiently similar to proteins that have been safely consumed in food;
- amino acid sequence similarity with known protein toxins and anti-nutrients;
- structural properties of the novel proteins including whether they are resistant to heat or processing.

An oral toxicity study is only deemed necessary if the results of biochemical, bioinformatic, digestibility or stability studies indicate further investigation of potential toxicity is warranted.

Studies submitted

2016. Amino acid sequence similarity search between Zea mays phytoene synthase (ZmPSY1) and known and putative protein toxins. Technical report, IR2016-01005 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Amino acid sequence similarity search between Pantoea ananatis phytoene desaturase (CRTI) and known and putative protein toxins. Technical report, IR2016-01004 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Characterization of the in vitro pepsin digestibility of phytoene synthase protein (ZmPSY1). Technical report, IR2016-01002 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Characterization of the in vitro pepsin digestibility of phytoene desaturase protein (CRTI) using western blot analysis. Technical report, IR2016-07003 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Characterization of the heat stability of phytoene synthase protein (ZmPSY1). Technical report, IR2015-12002 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Characterization of the heat stability of phytoene desaturase protein (CRTI). Technical report, IR2015-12001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

2016. CRTI: Acute oral toxicity study in mice. Technical report, IR2016-04001

History of human consumption

Homologues of the PSY1 protein are present in a range of fruits and vegetables containing the carotenoid biosynthetic pathway (e.g. sweet potato, carrot, dark green leafy vegetables such as spinach and kale, pumpkin, rock melon, capsicum, apricots). The *Zm*PSY1 protein is itself naturally present in corn kernels. *Pantoea* sp. containing the CRTI protein has been detected on meat (Ercolini et al. 2006), and fresh fruit and vegetables (Leff and Fierer 2013) but is not likely to have been widely consumed.

Amino acid sequence similarity to known protein toxins

Bioinformatic analyses are useful for assessing whether introduced proteins share any amino acid sequence similarity with known protein toxins.

Similarity searches were done for the *Zm*PSY1 (full 410-amino acid sequence as shown in Figure 8) and CRTI (492 amino acids as shown in Figure 9) proteins using the Fast Alignment Search Tool – All (FASTA) algorithm (Pearson and Lipman 1988) version 36.3.8c and BLOSUM50[16] scoring matrix against known protein sequences present in a toxin database created from UniProtKB[17] (UniProt Knowledgebase). At the time of analysis, this

[16] The BLOSUM series of matrices tabulate the frequency with which different substitutions occur in conserved blocks of protein sequences and are effective in identifying distant relationships. BLOSOM50 is the default for the FASTA programmes (Pearson 2013).

[17] http://www.uniprot.org/

database contained a total of 24,098 sequences. An *E*-score[18] acceptance criteria of 1×10^{-5} was used to identify sequences from the toxin database with potential for significant sequence similarity to the *Zm*PSY1 and CRTI proteins.

The toxin database searches using the *Zm*PSY1 query sequence did not return any entries with an *E*-score less than 1×10^{-5} i.e. *Zm*PSY1 does not have similarity with known toxins that are harmful to human health.

The search using the CRTI query sequence returned three protein accessions from the toxin database with an *E*-score less than 1×10^{-5}. These three alignments were to the N-terminal region of L-amino acid oxidase enzymes (LAAOs) from three species of venomous snakes. Specifically the sequence similarity was limited to a portion of the flavin adenine dinucleotide (FAD) binding domain. In assessing the potential toxicity of CRTI, a number of points are relevant:

- Snake venom LAAOs are not toxic via the oral exposure route
- LAAOs are widely distributed in many species that are not considered to be sources of toxins
- A bioinformatics search using the endogenous rice phytoene desaturase (see discussion in Section 3.4.1.1 and Figure 6) also returned matches to LAAOs that aligned with sequences in the N-terminal region.
- The N-terminal region showing sequence similarity is a common feature of other flavoenzymes such as oxidoreductases, protoporphyrinogen oxidases and other phytoene desaturases which, in themselves, are not toxins.
- The FAD-binding domain is only one of the three domains required for a functional catalytic site of an LAAO (Mitra and Bhattacharyya 2013).

The conclusion therefore is that the sequence homology flagged by the bioinformatics search does not lead to a toxicity concern for the CRTI protein.

Heat stability

The thermolability of a protein provides an indication of the stability of the protein under cooking/processing conditions.

For both fusion proteins, heat stability was determined by measuring enzymatic activity (as described in Section 4.1.4.1) over a temperature range between 30^{o} C – 65^{o} C imposed for 15 minutes prior to addition of the enzyme to the reaction mixture. The purified enzyme preparations used in the analysis were both derived from the microbially-expressed proteins characterised in Section 4.1.4.1.

For *Zm*PSY1, activity at 30^{o} C was *ca*. 28.4 pmol µg−1 min−1 under the assay conditions used. There was 50% loss of activity following pre-incubation at *ca*. 42°C for 15 minutes and complete loss of activity following pre-incubation at 50°C for 15 minutes.

For CRTI, activity at 30^{o} C was *ca*. 5.4 pmol µg−1 min−1 under the assay conditions used. There was 50% loss of activity when pre-incubated at *ca*. 51°C for 15 minutes and complete inactivation following pre-incubation at 55°C for 15 minutes.

[18] Comparisons between highly homologous proteins yield E-values approaching zero, indicating the very low probability that such matches would occur by chance. A larger E-value indicates a lower degree of similarity. Typically, alignments between two sequences will need to have an *E*-score of $1e^{-5}$ (1×10^{-5}) or smaller to be considered to have significant homology.

The temperatures required to completely inactivate the *Zm*PSY1 and CRTI enzymes are significantly lower than temperatures normally employed during cooking or processing of rice.

Acute oral toxicity

An acute oral toxicity study may provide additional reassurance of safety if the results of the biochemical, bioinformatic, digestibility or stability studies indicate a reason to further investigate the potential toxicity *in vivo*.

Since no potential hazards of *Zm*PSY1 were identified through the other studies described in this Section, further hazard characterisation of this protein by animal toxicity testing was not necessary.

Although no toxicity concerns were raised by the studies conducted for CRTI, the fact the protein is from a non-food source led the Applicant to submit an acute oral toxicity study in mice using the microbially-derived fusion CRTI. The results indicated there was no evidence of toxicity resulting from oral administration of the protein at a dose of 100 mg/kg of body weight. This result provides additional assurance of safety of the CRTI protein.

Potential allergenicity of ZmPSY1 and CRTI

The potential allergenicity of the two novel proteins was evaluated using an integrated, step-wise, case-by-case approach relying on various criteria used in combination. This is because no single criterion is sufficiently predictive of either allergenicity or non-allergenicity (see e.g. Thomas et al. 2009). The assessment focuses on:

- the source of the novel protein;
- any significant amino acid sequence similarity between the novel protein and known allergens; any potential for glycosylation
- the structural properties of the novel protein, including susceptibility to digestion, heat stability and/or enzymatic treatment; and
- specific serum screening if the novel protein is derived from a source known to be allergenic or has amino acid sequence similarity to a known allergen.

Applying this approach systematically provides reasonable evidence about the potential of a novel protein to act as an allergen.

Studies submitted

2016. Amino acid sequence similarity search between Zea mays phytoene synthase (ZmPSY1) and known and putative protein allergens. Technical report, IR2016-02002 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Amino acid sequence similarity search between Pantoea ananatis phytoene desaturase (CRTI) and known and putative protein allergens. Technical report, IR2016-02001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Characterization of the in vitro pepsin digestibility of phytoene synthase protein (ZmPSY1). Technical report, IR2016-01002 (unpublished), International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Characterization of the in vitro pepsin digestibility of phytoene desaturase protein (CRTI) using western blot analysis. Technical report, IR2016-07003 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

Source of the protein

The *Zm*PSY1 protein is derived from *Zea mays*. Few food-induced allergic reactions to corn have been reported and corn tends not to be regarded as a commonly allergenic food (OECD 2002). Nevertheless, there are allergenicity concerns associated with certain proteins expressed in the pollen of corn, e.g. Zm13 protein (Heiss et al. 1996), and a number of allergens have also been isolated from the seed, although most have not been clinically evaluated for their allergenic potential. The most significant of these is Zm14 (Pastorello et al. 2000), a lipid transfer protein (α-amylase inhibitor) that can maintain its IgE binding capacity even after heat treatment (Pastorello et al. 2003). Zm14 has been associated with anaphylactic reaction in susceptible individuals. Pastorello et al (2009) also nominate endochitinases and α-zein precursors as possible corn allergens. The PSY1 proteins, from corn or elsewhere, have not themselves been implicated in any food-related allergic reactions.

The CRTI protein is derived from the bacterium *Pantoea ananatis.* No endogenous proteins from this species are listed in the Food Allergy Research and Resource Program (FARRP) Allergen Protein Database[19] (accessed 11 May 2017). As indicated in Section 4.1.5.1, *Pantoea* sp. containing the CRTI protein has been detected on meat (Ercolini et al. 2006), and fresh fruit and vegetables (Leff and Fierer 2013).

Similarity to known allergens

Bioinformatic analysis provides part of a weight of evidence approach for assessing potential allergenicity of novel proteins introduced to GM plants (Thomas et al. 2005; Goodman 2006). It is a method for comparing the amino acid sequence of the introduced protein with sequences of known allergens in order to indicate potential cross-reactivity between allergenic proteins and the introduced protein. As with the bioinformatic analysis that looked at similarities of the novel proteins with known protein toxins (refer to Section 4.1.5.1), the generation of an *E* value provides an important indicator of significance of matches (Pearson 2000; Baxevanis 2005).

To evaluate the similarity to known allergens of the ZmPSY1 and CRTI proteins an epitope search was carried out to identify any short sequences of amino acids that might represent an isolated shared allergenic epitope. The only detectable CRTI protein in plant tissues was the mature form following processing to remove the N-terminal transit peptide (RUBISCO-SSU) sequence. However, to account for the possibility that extremely low concentrations of SSU-CRTI protein could be present, the allergen database search was also conducted using the full-length sequence of the fused product.

The *Zm*PSY1 and CRTI/SSU-CRTI sequences were compared with all known putative allergen and celiac sequences residing in a reference allergen database, (FARRP version 16, released on 27 January 2016 – containing 1,956 non-redundant entries) using the FASTA (version 35) algorithm and BLOSUM50 scoring matrix (refer to Section 4.1.5.1).

No biologically relevant alignment for either protein met or exceeded the Codex Alimentarius (Codex 2009) FASTA alignment threshold for potential allergenicity (35% identity over 80 amino acids). For CRTI, a search was also done to compare every possible peptide of eight contiguous amino acids (Metcalfe et al. 1996) with the sequences in the FARRP database. No alignments were found. It was concluded that neither *Zm*PSY1 nor CRTI/SSU-CRTI contain any cross-reactive IgE binding epitopes with known allergens.

[19] AllergenOnline - http://www.allergenonline.org/

Glycosylation

The potential for glycosylation was investigated for each protein. Glycosylation essentially involves the covalent attachment of a carbohydrate (glycan) to the target protein. The carbohydrate component may represent from <1% to >80% of the total molecular weight of glycoprotein. N-glycosylation can affect protein stability and/or activity and has also been implicated in contributing to possible allergenicity (Huby et al. 2000) since it may affect the susceptibility of a protein to processing and proteolysis and may introduce glycan peptides which are known to be highly cross-reactive epitopes (Altmann 2007).

Glycosylation that occurs on side chains of asparagine residues is termed N-glycosylation. It is commonly associated with an asparagine-X-serine/threonine sequence (N-X~(P)-[S/T), where X~(P) indicates any amino acid except proline (Orlando and Yang 1998). Although rare, the sequence asparagine-X-Cysteine (N-X-C) can also be an N-glycosylation site (Miletich and Broze Jr. 1990).

Sufficient amounts of either protein could not be obtained for glycostaining. Therefore only potential N-glycosylation sites could be identified through an *in silico* search of the plant-expressed protein sequence through GlycoEP[20] (Chauhan et al. 2013). The results of the search indicate that *Zm*PSY1 has one potential glycosylation site and CRTI has no potential sites. While a negative result provides some surety the protein, as expressed in the plant, is unlikely to be glycosylated, a positive result does not, conversely, suggest a protein is likely to be glycosylated. In particular it is noted that a linear sequence search does not account for the fact that glycan attachment occurs before, and influences, protein folding and it is the folding itself, which imparts biophysical characteristics to a protein (Shental-Bechor and Levy 2008; Chuang et al. 2012). Also, while N-glycosylation is the most common stable modification to impact the physicochemical properties of a protein, its occurrence is highly variable between species, environments and cells and its value as a safety assessment tool has been questioned (Bushey et al. 2014).

Heat stability – see Section 4.1.5.1

In vitro *digestibility*

Typically, food proteins that are allergenic tend to be stable to enzymes such as pepsin and the acidic conditions of the digestive system, exposing them to the intestinal mucosa and leading to an allergic response (Astwood and Fuchs 1996; Metcalfe et al. 1996; Kimber et al. 1999). Therefore some correlation exists between resistance to digestion by pepsin and potential allergenicity although it does not necessarily follow that resistance to digestion, or slow and limited digestibility, is always an indicator of an allergenic protein (Thomas et al. 2004; Herman et al. 2007). As a consequence, one of the criteria for assessing potential allergenicity is to examine the stability of novel proteins in conditions mimicking human digestion. Proteins that are rapidly degraded in such conditions are considered less likely to be involved in eliciting an allergic response.

A pepsin digestibility assay (Thomas et al. 2004) was conducted on the *Zm*PSY1 and CRTI fusion proteins (purified from a bacterial system) using simulated gastric fluid (SGF) (U.S.Pharmacopeia 2000). Both proteins were incubated in SGF at 37º for 0, 0.5, 1, 2, 5, 10, 20, 30 and 60 minutes and then inactivated by addition of buffer containing β-mercaptoethanol. The samples were then run on SDS-PAGE. Proteins were visualised by Colloidal Blue staining of the resulting gels. Western blotting of the SDS gels was also performed using an appropriate rabbit primary antibody and a HRP-conjugated goat anti-rabbit IgG secondary antibody.

[20] GlycoEP - http://www.imtech.res.in/raghava/glycoep/index.html

Following exposure to SGF containing pepsin for 30 seconds, the earliest time point sampled during the digestion, no intact *Zm*PSY1 protein (*ca*. 42 kDa)[21] or CRTI (*ca*. 50 kDa) was evident as assessed by either SDS-PAGE or western immunoblot analysis. For *Zm*PSY1, faint, low molecular weight degradation products were visible in samples incubated for up to two minutes but not at later time points, and these were not detected in the western blot.

These results support the conclusion that *Zm*PSY1 and CRTI proteins will be readily digested in a typical mammalian gastric environment.

Safety of the PMI protein

The PMI protein has been previously assessed by FSANZ in corn lines MIR604 – Application A564 (FSANZ 2006), MIR162 – Application A1001 (FSANZ 2008a), 3272 – Application A580 (FSANZ 2008b) and 5307 – Application A1060 (FSANZ 2012). In all of these applications, studies on potential toxicity and allergenicity were submitted, the most recent of which were for A1060. Since no concerns were raised and since the sequence of the protein expressed in GR2E is identical to the PMI sequence expressed in three of the corn lines (see Section 4.1.2), no further safety evaluation is required other than the examination of updated bioinformatics searches.

Bioinformatics Studies submitted

2016. Phosphomannose isomerase: Assessment of amino acid sequence similarity to known or putative allergens. Report Number SSB-124-16 (unpublished) Syngenta Crop Protection LLC, North Carolina, USA

2016. Phosphomannose isomerase: Assessment of amino acid sequence similarity to known or putative toxins. Report Number SSB-130-16 A1 (unpublished) Syngenta Crop Protection LLC, North Carolina, USA

The PMI sequence (Figure 10) was compared with all known putative allergen and celiac sequences residing in a reference allergen database, (FARRP version 16, released on 27 January 2016 – containing 1,956 non-redundant entries) using the FASTA algorithm and BLOSUM50 scoring matrix (refer to Section 4.1.5.1). A second search for matches of eight or more contiguous amino acids was conducted to screen for short, local regions of amino acid identity that might indicate the presence of common immunoglobulin E epitopes

No significant sequence similarity was observed between the PMI amino acid sequence and any entry in the Database. The FASTA search returned 15 alignments with *E*-values less than 10, none of which exceeded the minimum significance criteria of 35.0 % shared identity over a minimum of 80 amino acids. There was one sequence identity match of eight contiguous identical amino acids between PMI and a known allergen, α-parvalbumin from an unidentified *Rana* (frog) species. This alignment has been previously reported in assessments of the sequence similarity of PMI to known and putative allergens (FSANZ 2012). Further investigation using serum IgE screening demonstrated no cross-reactivity between PMI and the α-parvalbumin protein using serum from the single individual known to have demonstrated IgE-mediated allergy to this specific α-parvalbumin from the *Rana* species. The results indicated that the allergic patient's serum IgE does not recognize any portion of PMI as an allergenic epitope.

Using the BLASTP algorithm, the PMI sequence was also used to search the NCBI Entrez®

[21] the band, close to 42 kDa, seen in all SDS-PAGE samples containing SGF was attributable to pepsin (mw – 35 kDa).

Protein Database[22] (containing 83,737,374 sequences – accessed on 24 March 2016) and an in-house toxin database (created as a sub-set of the NCBI database using filtering of keywords) for similarities to known or putative toxins. Of a total of 1000 sequences identified in the NCBI database as having significant amino acid similarity to PMI, 999 alignments were to proteins from 115 different species. These alignments were to PMI or related proteins. The remaining sequence was identified as originating from a plant transformation vector. No alignments were obtained from the search of the in-house, toxin-specific database.

These results support the conclusion that PMI shares no biologically relevant amino acid sequence similarity to either known or putative protein allergens or known or putative protein toxins.

4.1.6 Bioinformatic analyses of potential ORFs created by the transformation procedure

Study submitted:

2015. Nucleotide sequence analysis of the inserted DNA and host genomic flanking regions in rice event IR-ØØGR2E-5. Technical report, IR2015-08001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

Bioinformatics analyses were performed to assess the similarity to known allergens and toxins of the putative polypeptides encoded by the two identified ORFs in the flanking regions of the GR2E insert (see Section 3.4.4).

The bioinformatics analyses were carried out by comparing the sequences of the two ORFs with sequences present in the same databases as described for the toxin and allergen searches in Sections 4.1.5.1 and 4.1.5.2. respectively, and using the same FASTA36 algorithm and BLOSOM50 scoring matrix.

For the toxin search, no entries were returned with an *E*-score less than 1×10^{-5} i.e. neither ORF has similarity with known toxins that are harmful to human health.

For the allergen search, no identity matches of greater than 35% over 80 residues were observed for either ORF, and no eight contiguous identical amino acid matches were observed for either ORF i.e. neither ORF contains any cross-reactive IgE binding epitopes with known allergens.

It is concluded that in the event an unexpected translation product were to be derived from the two ORFs detected in the GR2E junction regions, these putative polypeptides are not expected to possess functional cross-reactivity with known allergenic proteins or be toxic.

4.1.7 Conclusion

GR2E expresses three novel proteins, *Zm*PSY1, CRTI and PMI.

Analyses of all three proteins in straw and various grain stages indicated that measurable but very low concentrations of both PSY1 and CRTI proteins were found in all GR2E grain developmental stages but not in GR2E stem/straw tissue. Levels of CRTI were lower than *Zm*PSY1; both proteins had the highest mean concentration in grain at dough stage, ranging between *ca*. 308–359 ng/g and *ca*. 54–68 ng/g for *Zm*PSY1 and CRTI, respectively. Concentrations of PMI protein were significantly higher than either *Zm*PSY1 or CRTI in samples from all grain growth stages and were highest in dough-stage grain, averaging *ca*. 2015 ng/g across the locations/growing seasons tested. PMI was also present in straw.

[22] https://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Entrez/

Very low yields of the plant-produced *Zm*PSY1 and CRTI proteins precluded their specific characterisation. However the weight of evidence, provided by a) translation of the known DNA sequences of the three genes introduced into GR2E, b) the fact that the proteins function as predicted in the plant c) detection by specific antibodies and d) the detailed characterisation of equivalent microbially-produced proteins is sufficient to confirm the identity of the proteins expressed in GR2E.

For the *Zm*PSY1 and CRTI proteins, bioinformatic studies confirmed the lack of any significant amino acid sequence similarity to known protein toxins or allergens; digestibility studies suggest the proteins would be rapidly degraded in the gastro-intestinal tract following ingestion; and thermolability studies indicate both proteins are functionally inactivated following heating. There are no toxicity or allergenicity concerns.

There are also no concerns regarding the potential toxicity or allergenicity of the PMI protein which has been previously assessed as safe by FSANZ. Updated bioinformatics studies submitted in this application confirm the lack of any significant amino acid sequence similarity of the PMI expressed in GR2E to known protein toxins or allergens.

Taken together, the evidence indicates that *Zm*PSY1, CRTI and PMI proteins in the diet are unlikely to be toxic or allergenic in humans.

4.2 β-carotene

β-carotene is not normally produced in the rice grain and is therefore a new substance. A separate consideration of the nutritional safety of this (and other) carotenoid(s) in the context of the Australian/New Zealand diet is provided in Supporting Document 2.

5 Compositional analyses

The main purpose of compositional analyses is to determine if, as a result of the genetic modification, any unexpected changes have occurred to the food. These changes could take the form of alterations in the composition of the plant and its tissues and thus its nutritional adequacy. Compositional analyses can also be important for evaluating the intended effect where there has been a deliberate change to the composition of the food.

The classic approach to the compositional analyses of GM food is a targeted one. Rather than analysing every possible constituent, which would be impractical, the aim is to analyse only those constituents most relevant to the safety of the food or that may have an impact on the whole diet. Important analytes therefore include the key nutrients, toxicants and anti-nutrients for the food in question. The key nutrients and anti-nutrients are those components in a particular food that may have a substantial impact in the overall diet. They may be major constituents (fats, proteins, carbohydrates or enzyme inhibitors such as anti-nutrients) or minor constituents (minerals, vitamins). Key toxicants are those toxicologically significant compounds known to be inherently present in an organism, such as compounds whose toxic potency and level may be significant to health (e.g. glycoalkaloids in potatoes).

5.1 Key components

There are a number of components that are considered important for compositional analysis in rice matrices for food use (OECD 2016). Depending on variety, climate, soil type and farming practice, grain composition may vary widely. As a minimum, the key constituents of rice grain suggested for consideration for a comparative study include proximates, total dietary fibre, vitamins (thiamine, riboflavin, niacin, pantothenic acid, pyridoxine and α-

tocopherol), amino acids and fatty acids.

While rice is considered to be a safe source of food, there are two compounds in particular which are not favourable. Phytic acid is the main storage form of phosphorus in plant tissues; it is not in a bioavailable form for monogastric animals that lack the digestive enzyme phytase. It binds strongly to a) important minerals such as calcium, magnesium, iron, and zinc, thus reducing their absorption of these minerals and b) proteins, thus reducing their digestibility and amino acid bioavailability. Trypsin inhibitors are proteins known to inhibit biologically active trypsin, interfere with digestion and ultimately absorption of food material. They are typical anti-nutritional components in soybeans, cereals, and potatoes.

Analyses for key components were done on paddy rice grain and straw. Rice is largely cultivated for the production of grain for human food, and the straw is only infrequently used for livestock feed. As there are no human food products derived from straw, only the results of the compositional analyses for grain is presented in this report. However, it is noted that the compositional analyses for straw focussed only on proximates and minerals (calcium and phosphorus) and there were no statistically significant differences in proximates and fibre, or calcium and phosphorus between samples of straw obtained from GR2E and the Rc82 control.

5.2 Study design and conduct for key components

Studies submitted:

2016. Nutrient composition of event IR-ØØGR2E-5 and non-transgenic control rice grown during the rainy season in 2015 in the Philippines. Technical report, IR2015-07001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Nutrient composition of event IR-ØØGR2E-5 and non-transgenic control rice grown during the dry season in 2016 in the Philippines. Technical report, IR2016-05001 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

2016. Concentrations of beta-carotene and other carotenoids in grain samples from rice event IR-ØØGR2E-5. Technical report, IR2016-07004 (unpublished) International Rice Research Institute, Los Banos, Laguna, Philippines

Verified (multiplex PCR) plants of GR2E in the Rc82 background (generation BC_5F_3*[1] and BC_5F_4*[1] – see Figure 4) together with a near-isogenic non-transgenic control, Rc82, were grown at four locations in the Philippines[23] on the largest island of Luzon (Figure 11) during: the rainy season in 2015 and the dry season in 2016. Planting and crop maintenance were done according to local agronomic practices at each site. These plantings overlapped with those used for the rainy and dry season protein expression analyses described in Section 4.1.3.2.

Paddy rice grain[24] (approximately 12% moisture content – see Table 7) and straw samples were collected from matured rice plants, the stage when typical grain harvest of paddy rice would occur. Grain was collected from at least 150 plants per plot, excluding the outer rows, and pooled into a single sample per plot. Straw was collected from at least eight plants per plot, randomly selected and excluding the outer rows, and pooled into a single sample per plot.

Methods of composition analysis were based on internationally recognised procedures (e.g. those of the Association of Official Analytical Chemists), methods specified by the manufacturer of the equipment used for analysis, or other published methods.

[23] Batac City (Ilocos Norte), Los Baños (Laguna), Muñoz (Nueva Ecija), San Mateo (Isabela)

[24] Paddy rice (also known as 'rough' rice) is the name given to harvested and threshed rice grains that comprise an outer hull (also called husk) layer plus germ and bran layers.

5.3 Analyses of key components in grain

Grain samples were analysed for proximates, fibre, sugars, fatty acids, amino acids, minerals, vitamins, and anti-nutrients (phytic acid and trypsin inhibitor). In total, 69 analytes were assessed. In addition to these analytes the Applicant provided data and analysis on five carotenoid categories.
For each analyte a least squares mean (LSM) averaged over all sites and both growing seasons, was generated. Statistical analyses were performed using R (R Development Core Team 2011) and included the "lmerTest" (Bates et al. 2016) package. For each analyte, a linear mixed model Analysis of Variance was used by combining data from the two test years and four locations. The “lmer” procedure from the “lmerTest” package was used to fit the linear mixed model and to generate estimates of variance components and p-values. In assessing the significance of any difference between the mean analyte value for GR2E and the control, a p-value of 0.05 was used. This means that approximately 5% of statistically significant differences are expected to occur due to chance alone.

To assess whether any statistically significant differences are likely to be biologically meaningful. the results for GR2E and the Rc82 control have been compared to a combined literature range for each analyte, compiled from published literature for commercially available rice. It is noted, however, that information in the published literature for paddy rice is very limited and is unlikely to provide a broad reflection of the natural diversity that occurs within rice. Therefore, even if means fall outside the published range, this is not necessarily a concern.

5.3.1 Proximates and fibre

The analyses in this category also included measurement of the levels of amylose and starch as these are nutritionally significant carbohydrate components of the rice grain. The results are given in Table 7 and show there was no significant difference between the means for GR2E and those for the control for any analyte measured.

Table 7: Mean percentage (range) of proximates and fibre in paddy rice grain of GR2E and the Rc82 control collected over two growing seasons and four locations

Analyte[1]	GR2E	Control	p-value[3]	Combined literature range[4]
Crude Protein (%)	8.10 (6.07 – 11.2)	8.26 (6.03 – 11.4)	0.5445	5.9 – 11.8
Crude Fat (%)	1.42 (0.838 – 2.16)	1.34 (0.555 – 1.98)	0.7113	1.7 – 3.47
Ash (%)	5.89 (4.95 – 7.17)	6.02 (5 – 7.06)	0.3739	3.61 – 8.6
Carbohydrate (%)[2]	84.58 (81 – 86.9)	84.38 (81.1 – 86.4)	0.5595	79.98 -85.53
Amylose (%)	12.87 (7.31 – 18.6)	12.81 (6.76 – 18.6)	0.9548	N/A
Starch (%)	59.52 (32.8 – 71.5)	61.05 (28.1 – 73.9)	0.6888	61.9 – 67.2
Crude Fibre (%)	11.96 (10.1 – 14.6)	11.1 (10.1 – 12.3)	0.2133	8.6 – 18.13
Acid detergent fibre (%)	18.53 (15.7 – 21.7)	17.65 (15.6 – 18.8)	0.3522	10.8 – 18.2

Analyte[1]	GR2E	Control	p-value[3]	Combined literature range[4]
Neutral detergent fibre (%)	22.1 (17.5 – 35.5)	20.6 (16.2 – 32.8)	0.4774	15.0 -32.2
Total dietary fibre (%)	16.96 (12.8 – 20.3)	16.9 (11.4 – 21.4)	0.9549	16.73 – 22.97
Moisture (% fw)	12.26 (11.1 – 13.8)	12.32 (10.9 – 13.6)	0.8023	7.6 – 28.35

[1] Analyte means determined on a dry weight (dw) basis (except for moisture)
[2] Carbohydrate determined by calculation
[3] p-values indicating significant differences are bolded and underlined.
[4] Combined literature range compiled from ILSI (2014) and/or Heuze et al (2015)

5.3.2 Fatty acids

Three fatty acids (palmitic, oleic and linoleic acids) account for approximately 93% of the total fatty acids in rice grain (ILSI 2014). The levels of 21 fatty acids were measured, and of these the following 10 were below the level of quantification (LOQ) and were therefore excluded from analysis: : caprylic (8:0), capric (10:0), lauric (12:0), pentadecanoic (15:0), heptadecanoic (17:0), eicosadienoic (20:2), eicosatrienoic (20:3), arachidonic (20:4), erucic (22:1) and nervonic (24:1).

Results for the remaining 11 fatty acids are given in Table 8. Stearic acid was significantly higher (6.5% higher) in grain of GR2E than in the control; the mean level of stearic acid in both GR2E and the control was slightly higher than the range reported in the ILSI (2014) Crop Composition Database. For all other fatty acids, there was no significant difference between levels in grain of GR2E and the control.

Table 8: Mean percentage (range) composition, relative to total fat of major fatty acids in paddy rice grain from GR2E and the Rc82 control collected over two growing seasons and four locations

Fatty acid	GR2E (% Total)	Control (%Total)	p-value[1]	Combined literature range[3] (% Total)
Myristic (14:0)	0.438 (0.346 – 0.518)	0.393 (0.251 – 0.492)	0.1983	1.12 – 1.21
Palmitic (16:0)	19.54 (18.8 – 20.4)	18.45 (15.8 – 19.1)	0.2216	14.9 – 16.94
Palmitoleic (16:1)	0.193 (0.16 – 0.218)	0.191 (0.153 – 0.217)	0.3614	0.26 – 0.93
Stearic (18:0)	2.25 (1.95 – 2.78)[2]	2.11 (1.71 – 2.68)	**0.0487**	1.68 – 2.09
Oleic (18:1)	39.7 (38.4 – 41.3)	40.23 (38.4 – 45.9)	0.4866	37.49 – 40.49
Linoleic (18:2)	33.45 (32.4 – 34.4)	34.03 (30.1 – 35.4)	0.3944	37.51 – 38.49
α-Linolenic (18:3)	1.63 (1.34 – 2.31)	1.64 (1.27 – 2.7)	0.8871	1.12 – 1.21
Arachidic (20:0)	0.862 (0.739 – 0.989)	0.887 (0.743 – 1.02)	0.1967	0.66 – 0.79
Eicosenoic (20:1)	0.476 (0.405 – 0.538)	0.515 (0.442 – 0.773)	0.2857	0.54 – 0.58

Fatty acid	GR2E (% Total)	Control (%Total)	p-value[1]	Combined literature range[3] (% Total)
Behenic (22:0)	0.508 (0.407 – 0.585)	0.543 (0.478 – 0.61)	0.142	0.48 – 0.82
Lignoceric (24:0)	0.931 (0.687 – 1.25)	1.00 (0.805 – 1.19)	0.1159	1.06 – 1.34

[1] p-values indicating significant differences are bolded and underlined.
[2] Orange shading represents GR2E mean with a significantly higher value than that for the Rc82 control.
[3] Literature range compiled from ILSI (2014)

5.3.3 Amino acids

Levels of 18 amino acids were measured. Since asparagine and glutamine are converted to aspartate and glutamate respectively during the analysis, levels for aspartate include both aspartate and asparagine, while glutamate levels include both glutamate and glutamine. Results of the analysis are given in Table 9. There were no statistically significant differences in the concentrations of any amino acids between grain of GR2E and Rc82.

Table 9: Mean (range) amino acid composition (g/100 g dw) in paddy rice grain from GR2E and the Rc82 control collected over two growing seasons and four locations

Amino acid	GR2E (mg/100 g dw)	Control (mg/100 g dw)	p-value[1]	Combined literature range[2] (mg/100 g dw)
Alanine	455.4 (329 - 625)	459.8 (331 - 628)	0.7404	380 - 570
Arginine	563.8 (409 - 737)	564 (408 - 782)	0.988	530 - 850
Aspartate	708.4 (493 - 1010)	709.5 (497 - 994)	0.9600	610 - 940
Cysteine	156.3 (117 - 214)	155 (113 - 198)	0.7671	100 - 260
Glutamate	1354 (942 - 1980)	1360 (890 - 1990)	0.8975	1100 - 1740
Glycine	388.7 (293 - 495)	392.5 (292 - 511)	0.7161	340 - 480
Histidine	211.8 (162 - 276)	214.5 (157 - 281)	0.6593	140 - 250
Isoleucine	329.2 (241 - 446)	331.9 (237 - 464)	0.7871	270 - 430
Leucine	643.6 (464 - 886)	650.5 (462 - 921)	0.743	550 - 780
Lysine	298.6 (216 - 443)	294.4 (211 - 434)	0.7111	280 - 420
Methionine	169.2 (124 - 228)	166.2 (127 - 215)	0.6019	140 – 310
Phenylalanine	439.8 (324 – 593)	443.5 (317 – 622)	0.7992	280 - 520
Proline	376.4 (276 – 510)	380.9 (278 – 521)	0.6777	290 - 540

Amino acid	GR2E (mg/100 g dw)	Control (mg/100 g dw)	p-value[1]	Combined literature range[2] (mg/100 g dw)
Serine	400.6 (296 – 540)	400.5 (231 – 556)	0.9975	360 - 510
Threonine	308.8 (235 – 400)	308.2 (217 – 409)	0.9416	270 - 410
Tryptophan	73.36 (44.2 – 107)	74.93 (50.9 – 101)	0.7954	90 - 180
Tyrosine	213.5 (158 -282)	206.7 (133 – 291)	0.5314	130 - 480
Valine	468.2 (341 – 627)	474 (338 – 653)	0.6987	390 - 630

[1] p-values indicating significant differences are bolded and underlined.
[2] Literature range compiled from Juliano & Bechtel (1985) and ILSI (2014)

5.3.4 Minerals

Levels of nine minerals were measured. The means for these are given in Table 10. There were no statistically significant differences between grain from GR2E and Rc82 in the concentrations of any measured analytes

Table 10: Mean (range) mineral composition (g/100 g dw) in paddy rice grain from GR2E and the Rc82 control collected over two growing seasons and four locations

Mineral	GR2E (mg/100 g dw)	Control (mg/100 g dw)	p-value[1]	Combined literature range[2] (mg/100 g dw)
Calcium	22.5 (14.2 - 35)	21.39 (15.3 – 29.8)	0.5535	10 - 150
Copper	0.391 (0.183 – 0.68)	0.365 (0.215 – 0.506)	0.5441	0.2 – 1.3
Iron	3.96 (2.37 – 10.6)	4.57 (2.58 – 9.08)	0.5136	1.6 – 5.3
Magnesium	130.5 (87.5 - 185)	133.4 (102 - 157)	0.6312	30 - 170
Manganese	6.61 (4.33 – 8.39)	6.47 (4.85 – 7.78)	0.6741	2.0 – 11.7
Phosphorus	327.1 (211 - 461)	329.4 (241 - 383)	0.8698	190 - 470
Potassium	345.9 (236 - 597)	338.9 (222 - 472)	0.6656	170 - 350
Sodium	1.50 (0.56 – 3.81)	1.30 (0.543 – 3.07)	0.4668	0 - 100
Zinc	2.31 (1.63 – 3.21)	2.19 (1.73 – 2.78)	0.5689	0.2 – 3.6

[1] p-values indicating significant differences are bolded and underlined.
[2] Literature range compiled from Juliano & Bechtel (1985), ILSI (2014) and Heuze et al (2015)

5.3.5 Vitamins

Levels of six B vitamins plus β-carotene and α-tocopherol were measured (Table 11). The values for vitamin B2 (riboflavin) were below the LOQ (0.9 mg/mg dw) and were excluded

from analysis. With the exception of β-carotene, which was intended to be elevated in GR2E rice, there were no statistically significant differences noted in the concentrations of any of the other measured vitamins between GR2E and Rc82 rice grains. A p-value could not be obtained for β-carotene because > 80% of the values obtained for grain from Rc82 were below the LOQ (0.05 mg/kg dw).

Table 11: Mean (range) vitamin composition (g/kg dw) in paddy rice grain from GR2E and the Rc82 control collected over two growing seasons and four locations

Vitamin	GR2E (mg/kg dw)	Control (mg/kg dw)	p-value[1]	Combined literature range[2] (mg/kg dw)
B1 (thiamine)	3.13 (2.33 – 3.77)	3.08 (2.35 – 3.81)	0.6344	3.0 – 6.25
B3 (niacin)	36.9 (23.4 – 58.3)	32.63 (20.2 – 48.8)	0.4171	34 - 65
B5 (pantothenic acid)	9.15 (7.31 – 11.9)	9.13 (7.22 – 11.4)	0.9291	8 - 14
B6 (pyridoxine)	2.72 (2.22 – 3.3)	2.75 (2.1 – 5.42)	0.9185	5 - 8
B9 (folic acid)	0.912 (0.562 – 2.56)	0.883 (0.393 – 1.54)	0.881	0.2 – 0.5
all-*trans*-β-carotene	1.26 (0.504 – 2.35)[3]	<LOQ (<LOQ – 0.07)	Not determined	Not reported
α-tocopherol	2.98 (2.47 – 3.87)	2.75 (2.1 – 3.5)	0.3358	10 - 23

[1] p-values indicating significant differences are bolded and underlined.
[2] Literature range compiled from Juliano & Bechtel (1985) and ILSI (2014)
[3] Orange shading represents GR2E mean with a considerably higher value than that for the Rc82 control

A separate analysis of various carotenoids (including all-*trans*-β-carotene) in milled and polished (white) rice prepared from the grains of GR2E and Rc82, grown under the same conditions as described for the other compositional analyses in this Section, was undertaken by the Applicant. It is noted that for this analysis a number of parameters were different from those used for the all-*trans*-β-carotene analysis the results of which are given in Table 11 and therefore it would not be unexpected to obtain different results. These differences included a) the sample matrix (i.e. paddy versus milled rice - typically *ca.* 6.5 g of milled rice was produced from 10 g of paddy rice) and b) different extraction efficiencies between the Association of Official Analytical Chemists method used for β-carotene (AOAC 2000) and the procedure used by IRRI for carotenoids which, in particular, involved a pre-incubation of the ground grains in water for 10 min at 60^{o} C in order to swell the starch and potentially improve extraction.

In addition to all-*trans*-β-carotene, the carotenoids chosen for analysis were:

- all-*trans*-α-carotene. Lycopene cyclases are normally present in the endosperm of non-GM rice grains (see Figure 6) and drive the formation of either all-*trans*-α-carotene or all-*trans*-β-carotene.
- 9'-*cis*-β-carotene. Naturally produced β-carotene contains a mixture of *cis* and *trans* isomers. In some foods such as raw carrot, the trans isomer makes up as much as 98% of the β-carotene.
- β-cryptoxanthin. This is the next step from β-carotene in the biosynthesis of xanthophylls and is produced from β-carotene by the action of the enzyme carotenoid β-hydroxylase. Like β-carotene it is a pre-cursor of vitamin A, and β-cryptoxanthin -rich

foods are regarded as equivalent to β-carotene-rich foods as sources of vitamin A (Burri 2015).
- Total carotenoids.

The results from the IRRI analysis are given in Table 12. For the same reason as given for the β-carotene result in Table 11, no statistical analysis could be performed because levels of all the carotenoids in the control were below the LOQ. However, the levels of all carotenoids were considerably higher in GR2E than in the control. The results also show in GR2E:

- The pathway of carotene synthesis is pushed towards production of β-carotene which comprises some 60% of the total carotenoids in the endosperm.
- The predominant β-carotene isomer is in the *trans* form.
- Little β-carotene is converted to β-cryptoxanthin.
- The mean level of β-carotene reported in this analysis is *ca.* 3x higher than that reported in Table 11. Possible reasons for this have been discussed above.

Table 12: Mean (range) carotenoid levels in milled rice grain collected from GR2E and the Rc82 control over two growing seasons and four locations

Carotenoid	GR2E (µg/g dw)	Control (µg/g dw)
all-*trans*-α-carotene	0.713 (0.35 – 1.32)[1]	<LOQ
all-*trans*-β-carotene	3.57 (1.96 – 7.31)[1]	<LOQ
9'-*cis*-β-carotene	0.762 (0.5 – 1.32)[1]	<LOQ
β-cryptoxanthin	0.312 (0.232 – 0.464)[1]	<LOQ
Total carotenoids	5.88 (3.5 – 10.9)[1]	<LOQ

[1] Orange shading represents GR2E mean with a considerably higher value than that for the Rc82 control

5.3.6 Anti-nutrients

There were no statistically significant differences in the concentrations of phytic acid or in the levels of trypsin inhibitor between grain samples of GR2E and Rc82 (Table 13).

Table 13: Mean (range) anti-nutrient levels in paddy rice grain from GR2E and the Rc82 control collected over two growing seasons and four locations

Anti-nutrient	GR2E (mg/kg dw)	Control (mg/kg dw)	p-value[1]	Combined literature range[2] (mg/kg dw)
Phytic acid (%dw)	0.861 (0.582 – 1.1)	0.881 (0.605 – 1.23)	0.6218	30.61 – 0.84
Trypsin inhibitor (TIU/mg)	0.924 (0.28 – 1.71)	0.996 (0.032 – 4.17)	0.8282	Not reported

[1] p-values indicating significant differences are bolded and underlined.
[2] Literature range compiled from ILSI (2014)

5.4 Conclusions of the compositional analyses

Detailed compositional analyses were done to establish the nutritional adequacy of grain from GR2E and to characterise any unintended compositional changes. Samples were analysed for proximates, fibre, sugars, fatty acids, amino acids, minerals, vitamins, and anti-nutrients (phytic acid and trypsin inhibitor). In addition to these analytes the Applicant provided data and analysis on five carotenoid categories.

The levels of each analyte in the major analyses were compared to levels in: a) the non-GM parental line, Rc82 and b) levels recorded in the literature, noting that data are limited for uncooked rice. Of the 69 analytes considered in the major analyses, and with the exception of β-carotene which was expected to be higher in GR2E, only stearic acid in GR2E differed significantly from the control. All non-carotenoid analyte levels in GR2E were within, or similar to, the range of natural variability of those components in conventional rice varieties as reported in the limited literature. It can therefore be concluded that apart from the elevated levels of carotenoids, the grain in GR2E is compositionally equivalent to grain from conventional rice varieties.

6 Nutritional impact

In assessing the safety of a GM food, a key factor is the need to establish that the food is nutritionally adequate and will support typical growth and well-being. In most cases, this can be achieved through an understanding of the genetic modification and its consequences, together with an extensive compositional analysis of the food.

If the compositional analysis indicates biologically significant changes in the levels of certain nutrients in the GM food, additional nutritional assessment may be required to determine if the change is likely to be associated with any adverse health consequences. Evidence indicates that feeding studies using target livestock species will add little to the safety assessment (see e.g. OECD 2003; Bartholomaeus et al. 2013; Herman and Ekmay 2014).

In the case of GR2E, there are significant changes in the carotenoid component of the grain, particularly the production of β-carotene. However, β-carotene is a typical constituent of the human diet being found in a range of fresh fruits and vegetables (e.g. sweet potato, carrot, dark green leafy vegetables such as spinach and kale, pumpkin, rock melon, capsicum, apricots).

FSANZ has conducted a separate nutrition safety assessment of GR2E to determine the potential risks associated with the intake of the contained carotenoids by the general population and whether this may cause any adverse health effects (see Supporting Document 2).

7 References[25]

Altmann F (2007) The role of protein glycosylation in allergy. International Archives of Allergy and Immunology 142:99–115.
DOI: 10.1159/000096114.

AOAC (2000) AOAC Method 941.15. Carotene in fresh plant materials and silage. In: Horowitz W (ed) Official methods of analysis of AOAC International. 17 ed, Association of Official Analytical Chemists, Gaithersburg, MD,

[25] All website references were current as of 7 July 2017

ASCIA (2014) Food protein-induced enterocolitis syndrome (FPIES)., Australasian Society of Clinical Immunology and Allergy.
http://www.allergy.org.au/patients/food-other-adverse-reactions/food-protein-induced-enterocolitis-syndrome-fpies

Astwood JD, Fuchs RL (1996) Allergenicity of foods derived from transgenic plants. Monographs in Allergy 32 (Highlights in Food Allergy):105–120

Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MMM, Ramadan HAI, Saini KS, Baeshen NA, Redwan EM (2015) Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. Journal of Microbiology and Biotechnology 27(7):953–962

Bartholomaeus A, Parrott W, Bondy G, Walker K (2013) The use of whole food animal studies in the safety assessment of genetically modified crops: Limitations and recommendations. Critical Reviews in Toxicology 43(S2):1–24

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Grothendieck G, Green P (2016) lme4: Linear mixed-effects models using Eigen and S4. R package ver. 1.0-6. http://cran.r-project.org/package=lme4.

Baxevanis AD (2005) Assessing Pairwise Sequence Similarity: BLAST and FASTA. Ch 11 In: Baxevanis AD, Ouellette BFF (eds) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley & Sons, Inc., p. 295–324

Beyer P, Al-Babili S, Ye X, Lucca P, Schaub P, Welsch R, Potrykus I (2002) Golden Rice: introducing the ß-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. The Journal of Nutrition 132:506S–510S

Buckner B, San Miguel P, Janick-Buckner D, Bennetzen JL (1996) The *y1* gene of maize codes for phytoene synthase. Genetics 143:479–488

Burri BJ (2015) Beta-cryptoxanthin as a source of vitamin A. Journal of the Science of Food and Agriculture 95(9):1788–1794.
DOI: 10.1002/jsfa.6942.

Bushey DF, Bannon GA, Delaney BF, Graser G, Hefford M, Jiang X, Lee TC, Madduri KM, Pariza M, Privalle LS, Ranjan R, Saab-Rincon G, Schafer BW, Thelen JJ, Zhang JXQ, Harper MS (2014) Characteristics and safety assessment of intractable proteins in genetically modified crops. Regulatory Toxicology and Pharmacology 69:154–170

Chauhan JS, Rao A, Raghava GPS (2013) *In silico* platform for prediction of N-, O-, and C- glycosites in eukaryotic protein sequences. PLoS ONE (open access) 8(6):e67008.
doi:10.1371/journal.pone.0067008.

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Molecular Biology 18(4):675–689

Chuang G-Y, Boyington JC, Joyce MG, Zhu J, Nabel GJ, Kwong PD, Georgiev I (2012) Computational prediction of N-linked glycosylation incorporating structural proteins and patterns. Bioinformatics 28(17):2249–2255.
doi:10.1093/bioinformatics/bts426.

Codex (2009) Foods derived from modern biotechnology, 2nd edition. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome.
http://www.codexalimentarius.org/standards/thematic-publications/

Coruzzi G, Broglie R, Edwards C, Chua N-H (1984) Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. The EMBO

Journal 3(8):1671–1679

Coutinho T, Venter SN (2009) *Pantoea ananatis*: an unconventional plant pathogen. Molecular Plant Pathology 10(3):325–335.
DOI: 10.1111/J.1364-3703.2009.00542.X.

De Baere T, verhelst R, Labit C, Verschraegen G, Wauters G, Claeys G, Vaneechoutte M (2004) Bacteremic infection with *Pantoea ananatis*. Journal of Clinical Microbiology 42(9):4393–4395.
DOI: 10.1128/JCM.42.9.4393-4395.2004.

DE Maayer P, Chan WY, Rubagotti E, Venter SN, Toth IK, Birch PRJ, Coutinho TA (2014) Analysis of the Pantoea ananatis pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. BMC Genomics 15(404)

Delaney B, Astwood JD, Cunny H, Eichen Conn R, Herouet-Guicheney C, MacIntosh S, Meyer LS, Privalle LS, Gao Y, Mattsson J, Levine M, ILSI (2008) Evaluation of protein safety in the context of agricultural biotechnology. Food and Chemical Toxicology 46:S71–S97

Depicker A, Stachel S, Dhaese P, Zambryski P, Goodman HM (1982) Nopaline synthase: transcript mapping and DNA sequence. Journal of Molecular and Applied Genetics 1(6):561–573

Donnenberg MSe (2002) *Escherichia coli*: virulence mechanisms of a versatile pathogen. Academic Press, San Diego

Dunand R, Saichuk J (2014) Rice growth and development. Ch 4 In: Saichuk J (ed) Louisiana rice production handbook (Pub. 2321 (3M) 5/14 Rev.). 4 ed, Louisiana State University Ag Center p. 41–53 http://www.lsuagcenter.com/topics/crops/rice/variety_trials_recommendations/rice-production-handbook

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8:978–984

Enya J, Shinohara H, Yoshida S, Tsukiboshi T, Negishi H, Suyama K, Tsushima S (2007) Culturable Leaf-Associated Bacteria on Tomato Plants and Their Potential as Biological Control Agents. Microbial Ecology 53(4):524–536

Ercolini D, Russo F, Torrieri E, Masi P, Villani F (2006) Changes in the spoilage-related microbiota of beef during refrigerated storage under different packaging conditions. Applied and Environmental Microbiology 72(7):4663–4671.
doi:10.1128/AEM.00468-06.

FAOSTAT (2017) Online database of the Food and Agriculture Organization of the United Nations. http://faostat3.fao.org/browse/T/TP/E

FSANZ (2006) Application A564 - Food derived from insect-protected corn line MIR604. Report prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/pages/applicationa564foodd2972.aspx

FSANZ (2008a) Application A1001 - Food derived from insect-protected corn line MIR162. Report prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/pages/applicationa1001food3769.aspx

FSANZ (2008b) Application A580 - Food derived from amylase-modified corn line 3272. Report prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/pages/applicationa580foodd3243.aspx

FSANZ (2012) Application A1060 - Food derived from insect-protected corn line 5307. Report prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/pages/applicationa1060food5180.aspx

Goodman RE (2006) Practical and predictive bioinformatics methods for the identification of potentially cross-reactive protein matches. Molecular Nutrition and Food Research 50:655–660

Gravois KA, Moldenhauer KAK, Lee FN, Norman RJ, Helms RS, Bernhardt JL, Wells BR, Dilday RH, Rohman PC, Blocker MM (1995) Registration of 'Kaybonnet' rice. Crop Science 35(2):587. https://dl.sciencesocieties.org/publications/cs/abstracts/35/2/CS0350020587?access=0&view=pdf.

GRiSP (2013) Rice almanac (4th edition). Report prepared by the Global Rice Science Partnership (GRiSP) and published by the International Rice Research Institute (IRRI), Los Baños (Philippines). http://ricepedia.org/

Heiss S, Flicker S, Hamilton DA, Kraft D, Mascarenhas JP, Valenta R (1996) Expression of Zm13, a pollen specific maize protein in *Escherichia coli* reveals IgE-binding capacity and allergenic potential. FEBS Letters 381:217–221

Herman RA, Ekmay R (2014) Do whole-food animal feeding studies have any value in the safety assessment of GM crops? Regulatory Toxicology and Pharmacology 68:171–174

Herman RA, Woolhiser MM, Ladics GS, Korjagin VA, Schafer BW, Storer NP, Green SB, Kan L (2007) Stability of a set of allergens and non-allergens in simulated gastric fluid. International Journal of Food Sciences and Nutrition 58:125–141

Heuzé V, Tran G, Hassoun P (2015) Rouch rice (paddy rice). Feedipedia: an open access information system on animal feed resources. A joint online project of INRA, CIRAD, AFZ and FAO. http://www.feedipedia.org/node/226

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. The Plant Journal 6(2):271–282

Huby RDJ, Dearman RJ, Kimber I (2000) Why are some proteins allergens? Toxicological Sciences 55:235–246.
https://doi.org/10.1093/toxsci/55.2.235.

ILSI (2008) Chapter 5: Golden Rice 2. Comprehensive Reviews in Food Science and Food Safety(7):92–98

ILSI (2014) International Life Sciences Institute Crop Composition Database Version 4.0. https://www.cropcomposition.org/query/index.html

Ito M, Kato T, Matsuda T (2005) Rice allergenic proteins, 14-16 kDa albumin and alpha-globulin, remain insoluble in rice grains recovered from rice miso (rice-containing fermented soybean paste). Bioscience, Biotechnology and Biochemistry 69(6):1137–1144

Izumi H, Sugiyama M, Matsuda T, Nakamura R (1999) Structural characterization of the 16-kDa allergen, RA17, in rice seeds. Prediction of the secondary structure and identification of intramolecular disulfide bridges. Bioscience, Biotechnology and Biochemistry 63(12):2059–2063

Jensen ON, Podtelejnikov AV, Mann M (1997) Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. Annals of Chemistry 69(23):4741–4750

Jeon YH, Oh SJ, Yang HJ, Lee SY, Pyun BY (2011) Identification of major rice allergen and their clinical significance in children. Korean Journal of Paediatrics 54(10):414–421

Jia Y, Jiah MH, Wang X, Liu G (2012) *Indica* and *Japonica* crosses resulting in linkage block and recombination suppression on rice chromosome 12. PLoS ONE (open access) 7(8):e43066. doi:10.1371/journal.pone.0043066.

Juliano B, Bechtel D (1985) The rice grain and its gross composition. Ch 4 In: Juliano B (ed) Rice:

Chemistry and technology. American Association of Cereal Chemists, St Paul, MN, p. 17–57

Kim S-I, Veena, Gelvin SB (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. The Plant Journal 51:779–791

Kimber I, Kerkvliet NI, Taylor SL, Astwood JD, Sarlo K, Dearman RJ (1999) Toxicology of protein allergenicity: prediction and characterization. Toxicological Sciences 48(2):157–162

Leff JW, Fierer N (2013) Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS ONE (open access) 8(3):e593 10.
doi:10.1371/journal.pone.0059310.

McLaren DS, Kraemer K (2012) Manual on vitamin A deficiency disorders (VADD). World Review of Nutrition & Dietetics, No.103 ed, Karger, Basel

Mehr SS, Kakakios AM, Kemp AS (2009) Rice: a common and severe cause of food protein-induced enterocolitis syndrome. Archives of Diseases in Childhood 94:220–223.
doi:10.1136/adc.2008.145144.

Mergaert J, Verdonck L, Kersters K (1993) Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. International Journal of Systematic Bacteriology 43(1):162–173

Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Critical Reviews in Food Science and Nutrition 36 Suppl:S165–S186

Miles JS, Guest JR (1984) Nucleotide sequence and transcriptional start point of the phosphomannose isomerase gene (manA) of *Escherichia coli*. Gene 32(1-2):41–48

Miletich JP, Broze Jr. GJ (1990) b Protein C is not glycosylated at asparagine 329. The Journal of Biological Chemistry 265:11397–11404

Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K (1990) Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. Journal of Bacteriology 172(12):6704–6712

Mitra J, Bhattacharyya D (2013) Irreversible inactivation of snake venom L-amino acid oxidase by covalent modification during catalysis of L-propargylglycine. FEBS Open Bio 3:135–143

Murray MG, Thompson WF (1980) Rapid isolation of higher weight DNA. Nucleic Acids Res 8(19):4321–4325.
DOI: 10.1093/nar/8.19.4321.

Muthayya S, Sugimoto JD, Montgomery S, Maberly GF (2014) An overview of global rice production, supply, trade, and consumption. Annals of the New York Academy of Sciences 1324:7–14.
doi: 10.1111/nyas.12540.

Negrotto D, Jolley M, Beer S, Wenck AR, Hansen G (2000) The use of phosphomannose-isomerase as a selectable marker to recover transgenic maize plants (*Zea mays* L.) via *Agrobacterium* transformation. Plant Cell Reports 19:798–803

OECD (2002) Consensus document on compositional considerations for new varieties of maize (*Zea mays*): Key food and feed nutrients, antinutrients and secondard plant metabolites. Organisation for Economic Co-operation and Development, Paris.
http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=env/jm/mono(2002)25&doclanguage=en

OECD (2003) Considerations for the safety assessment of animal feedstuffs derived from genetically

modified plants. ENV/JM/MONO(2003)10. Organisation for Economic Co-operation and Development, Paris.
http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=ENV/JM/MONO(2003)10&doclanguage=en

OECD (2016) Revised consensus document on compositional considerations fro new varieties of rice (*Oryza sativa*): key food and feed nutrients, anti-nutrients and other constituents. ENV/JM/MONO(2016)38. Organisation for Economic Co-operation and Development, Paris.
http://www.oecd.org/chemicalsafety/biotrack/consensus-document-for-work-on-safety-novel-and-foods-feeds-plants.htm

OGTR (2005) The biology and ecology of rice (*Oryza sativa* L.) in Australia. Report prepared by the Australian Government Office of the Gene Technology Regulator.
http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/biology-documents-1

OGTR (2008) The biology of *Zea mays* L. ssp *mays* (maize or corn). Prepared by the Office of the Gene Technology Regulator, Australia.
http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/riskassessments-1

Okada K, Saito T, Nakagawa T, Kawamukai M, Kumiya Y (2000) Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in Arabidopsis. Plant Physiol 122(4):1045–1056

Orlando R, Yang Y (1998) Analysis of Glycoproteins. Ch 9 In: Larsen BS, McEwen CN (eds) Mass Spectrometry of Biological Materials. 2nd ed, Marcel Dekker, p. 216–246

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nature Biotechnology 23:482–487

Pastorello EA, Farioli L, Pravettoni V, Scibilia J, Conti A, Fortunato D, Borgonovo L, Bonomi S, Primavesi L, Ballmer-Weber B (2009) Maize food allergy: lipid-transfer proteins, endochitinases, and alpha-zein precursors are relevant maize allergens in double-blind placebo-controlled maize-challenge-positive patients. Analytical and Bioanalytical Chemistry 395(1):93–102. DOI: 10.1007/s00216-009-2945-z.

Pastorello EA, Farioli L, Pravettoni V, Ispano M, Scibola E, Trambaioli C, Giuffrida MG, Ansaloni R, Godovac-Zimmermann J, Conti A, Fortunato D, Ortolani C (2000) The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. Journal of Allergy and Clinical Immunology 106(4):744–751

Pastorello EA, Pompei C, Pravettoni V, Farioli L, Calamari AM, Scibilia J, Robino AM, Conti A, Iametti S, Fortunato D, Bonomi S, Ortolani C (2003) Lipid-transfer protein is the major maize allergen maintaining IgE-binding activity after cooking at 100 degrees C, as demonstrated in anaphylactic patients and patients with positive double-blind, placebo-controlled food challenge results. Journal of Allergy and Clinical Immunology 112(4):775–783

Pearson WR (2000) Flexible Sequence Similarity Searching with the FASTA3 Program Package. Ch 10 In: Misener S, Krawetz SA (eds) Methods in Molecular Biology, Volume 132: Bioinformatics Methods and Protocols. Human Press Inc., Totowa, NJ, p. 185–219

Pearson WR (2013) Selecting the right similarity-scoring matrix. Current Protocols in Bioinformatics 43:3.5.1–3.5.9

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences 85(8):2444–2448

Proudfoot AE, Payton MA, Wells TN (1994a) Purification and characterization of fungal and mammalian phosphomannose isomerases. Journal of Protein Chemistry 13(7):619–627

Proudfoot AEI, Turcatti G, Wells TNC, Payton MA, Smith DJ (1994b) Purification cDNA cloning and heterologous expression of human phosphomannose isomerase. The FEBS Journal 219(1-2):415–423.
DOI: 10.1111/j.1432-1033.1994.tb19954.x.

R Development Core Team (2011) R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna.
http://www.R-project.org/.

Ruiz-Sola MA, Rodriguez-Concepcion M (2012) Carotenoid biosynthesis in Arabidopsis: a colorful pathway. The Arabidopsis Book 10:e0158.
doi: 10.1199/tab.0158.

Schaub P, Al-Babili S, Drake R, Beyer P (2005) Why is golden rice golden (yellow) instead of red? Plant Physiol 138:441–450

Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proceedings of the National Academy of Sciences 105(24):8256–8261

Simpson RJ (2010) SYPRO Orange fluorescent staining of protein gels. Cold Spring Harbor Protocols 2010(4):pdb.prot5414.
https://www.ncbi.nlm.nih.gov/pubmed/20360368. doi: 10.1101/pdb.prot5414.

Suvarna BS (2008) Rice, the allergen. Nepal Journal of Science and Technology 9:195–199

Takaiwa F, Ebinuma H, Kikuchi S, Oono K (1987) Nucleotide sequence of a rice glutelin gene. FEBS Letters 22(1):43–47

Thomas K, Bannon G, Hefle S, Herouet C, Holsapple M, Ladics G, MacIntosh S, Privalle L (2005) In silico methods for evaluating human allergenicity to novel proteins: International Bioinformatics Workshop Meeting Report February 23 - 24, 2005. Toxicological Sciences 88:307–310

Thomas K, MacIntosh S, Bannon G, Herouet-Guicheney C, Holsapple M, Ladics G, McClain S, Vieths S, Woolhiser M, Privalle L (2009) Scientific advancement of novel protein allergenicity evaluation: An overview of work from the HESI Protein Allergenicity Technical Committee (2000 - 2008). Food and Chemical Toxicology 47:1041–1050

Thomas K, Aalbers M, Bannon GA, Bartels M, Dearman RJ, Esdaile DJ, Fu T-J, Glatt CM, Hadfield N, Hatzos C, Hefle SL, Heylings JR, Goodman RE, Henry B, Herouet C, Holsapple M, Ladics GS, Landry TD, MacIntosh SC, Rice EA, Privalle LS, Steiner HY, Teshima R, Van Ree R, Woolhiser M, Zawodny J (2004) A multi-laboratory evaluation of a common *in vitro* pepsin digestion assay protocol used in assessing the safety of novel proteins. Regulatory Toxicology and Pharmacology 39:87–98

Trcka J, Schäd SG, Scheurer S, Conti A, Vieths S, Gross G, Trautmann A (2012) Rice-Induced anaphylaxis: IgE-mediated allergy against a 56-kDa glycoprotein. International Archives of Allergy and Immunology 158:9–17

Tzfira T, Li J, Lacroix B, Citosky V (2004) *Agrobacterium* T-DNA integration: molecules and models. Trends in Genetics 20(8):375–383

U.S.Pharmacopeia (2000) The United States Pharmacopeia 24. The National Formulary 19th ed. United States Pharmacopeia Convention, Inc, Rockville, MD

Usui Y, Nakase M, Hotta H, Urisu A, Aoki N, Kitajima K, Matsuda T (2001) A 33-kDa Allergen from rice (*Oryza sativa* L. *Japonica*). The Journal of Biological Chemistry 276(14):11376–11381

Watanabe K, Sato M (1999) Gut colonization by an ice nucleation active bacterium *(Erwinia) Pantoea ananas* reduces the cold hardiness of mulberry pyralid larvae. Cryobiology 38(4):281–289.
DOI: 10.1006/cryo.1999.2169.

Wu L, Liu R, Niu Y, Lin H, Ye W, Guo L, Hu X (2016) Whole genome sequence of *Pantoea ananatis* R100, an antagonistic bacterium isolated from rice seed. Journal of Biotechnology 225:1–2. http://dx.doi.org/10.1016/j.jbiotec.2016.03.007

Yeum K-J, Russell RM (2002) Carotenoid bioavailability and bioconversion. Annual Review of Nutrition 22:483–504.
DOI: 10.1146/annurev.nutr.22.010402.102834.

Zhang J, Xu RJ, Elliott MC, Chen DF (1997) *Agrobacterium*-mediated transformation of elite indica and japonica rice cultivars. Molecular Biotechnology 8(3):223–231.
DOI:10.1007/BF02760776.