SCIENTIFIC REPORTS
nature research

OPEN

Molecular characterization and safety assessment of biofortified provitamin A rice

Norman Oliva[1], Maria Florida Cueto-Reaño[1], Kurniawan R. Trijatmiko[1], Mercy Samia[1], Ralf Welsch[2], Patrick Schaub[2], Peter Beyer[2], Donald Mackenzie[3*], Raul Boncodin[1], Russell Reinke[1], Inez Slamet-Loedin[1*] & B. P. Mallikarjuna Swamy[1*]

Part of the studies involved in safety assessment of genetically engineered crops includes characterizing the organization, integrity, and stability of the inserted DNA and evaluating the potential allergenicity and toxicity of newly-expressed proteins. Molecular characterization of the introduced DNA in provitamin A biofortified rice event GR2E confirmed insertion of a single copy of the transfer-DNA in the genome and its inheritance as a single locus. Nucleotide sequencing of the inserted DNA confirmed it was introduced without modifications. The phytoene synthase, and carotene desaturase proteins did not display sequence similarity with allergens or toxins. Both proteins were rapidly digested in simulated gastric fluid and their enzymatic activity was inhibited upon heat treatment. Acute oral toxicity testing of the protein in mice demonstrated lack of adverse effects. These evidences substantiated the lack of any identifiable hazards for both proteins and in combination with other existing comparative analyses provided assurance that food derived from this rice is safe. This conclusion is in line with those of the regulatory agencies of US Food and Drug Administration, Health Canada and Food Standard Australia and New Zealand.

Vitamin A deficiency (VAD) persists as a major public health nutrition problem, especially in sub-Saharan Africa and South Asia. Severe deficiency can lead to disorders such as xerophthalmia, the leading cause of preventable childhood blindness, anemia, and weakened host resistance to infection, which can increase the severity of infectious diseases and risk of death[1,2]. VAD increases vulnerability to a range of illnesses including diarrhea, measles, and respiratory infections, which are the leading causes of mortality among children in low and middle income countries, where risk of infection and risk of mortality can be compounded by coexisting under-nutrition. The prevalence and severity of xerophthalmia, anemia and the less-measurable "vicious cycle" between VAD and infection in vulnerable groups, notably young children and pregnant or lactating mothers, represent the most compelling consequences of VAD and underscore its significance as a public health problem around the world. In Bangladesh, the prevalence of VAD among preschool and school age children is a serious public health problem, with respective rates found to be 20.5 percent and 20.9 percent, respectively[3].

In Southeast Asian countries, the average per capita milled rice consumption is 130 kg/year or 360 g/day, providing nearly two-thirds of the necessary caloric intake. The high consumption of rice may reflect a lack of dietary diversity, which, when combined with the poor micronutrient content of polished rice in general and lack of any provitamin A, specifically, is a risk factor for VAD[4]. In addition to existing efforts, a complementary intervention for reducing VAD in the highest-risk countries is the biofortification of staple food crops with provitamin A. This approach has been successfully used with crops such as maize, cassava, and sweet potato where existing germplasm with elevated levels of provitamin A expression has been introduced into breeding programs[5,6].

Golden Rice event 2E (GR2E) was developed through the use of recombinant-DNA techniques to express elevated levels of provitamin A (mainly ß-carotene) in the endosperm, which is converted in the body to vitamin A. The use of recombinant DNA technologies was required because provitamin A formation in the grain endosperm was not found in rice germplasm collections. GR2E rice provides a complementary effort to reduce VAD by

[1]Strategic Innovation Platform, International Rice Research Institute (IRRI), Metro Manila, Philippines. [2]Faculty of Biology, Cell Biology, University of Freiburg, Freiburg, Germany. [3]Institute for International Crop Improvement, Donald Danforth Plant Science Center, Saint Louis, Missouri, USA. *email: DMacKenzie@danforthcenter.org; I.Slamet-Loedin@irri.org; m.swamy@irri.org

providing ready access to a portion of the estimated average requirement for vitamin A deficient population whose diet is primarily rice-based.

The carotenoid biosynthetic pathway in immature rice endosperm functions up to the synthesis of geranylgeranyl diphosphate. While this compound is not solely devoted to carotenogenesis, it can be used as a substrate to produce the phytoene by a phytoene synthase heterologously expressed in the rice endosperm[7]. Completion of the pathway leading to the synthesis of ß-carotene via expression of the carotene desaturase I (*crtI*) gene from *Pantoea ananatis* (formerly *Erwinia uredovora*) that is fused in-frame at the 5′ terminus with the pea (*Pisum sativum*) RUBISCO SSU transit peptide encoding sequence (*SSU-crtI*), was described by Ye *et al.*[8]. Paine *et al.*[9] described improved constructs incorporating the *Zea mays* phytoene synthase gene (*Zmpsy1*) and phosphomannose isomerase (*pmi)* as selection marker for transformants, which produced elevated total carotenoid levels of up to 30 μg/g in the endosperm, of which *ca*. 80 percent were mixed isomers of ß-carotene. One of the resulting *Agrobacterium tumafaciens*- mediated transformants was GR2E using plasmid pSYN12424 containing the *SSU-crtI* (from *Pisum sativum* and *Pantoea ananatis*), *Zmpsy1* and *pmi* (from *E.coli*).

Completion of the food safety assessment is a necessary prerequisite to the conduct of bioefficacy studies with GR2E, which are an important component in demonstrating the value of GR2E rice in mitigating VAD. The totality of the assessment that provides assurance that GR2E is as safe as conventional rice involves characterization of introduced genes and proteins, which is discussed in this paper and also additionally, comparative analysis of the composition of rice used for food and feed[10]. Molecular characterization of a new transgenic event is important in providing an understanding of the nature of the genetic modification; including any genetic material introduced into the host genome, and provides context for the safety assessment[11]. Typically, the molecular characterization addresses three main aspects: a) the transformation method together with a detailed description of the DNA sequences introduced into the host genome; b) characterization of the inserted DNA, including any rearrangements that may have occurred as a consequence of the transformation; and c) the genetic stability of the inserted DNA and any expressed traits. The molecular characterization of GR2E rice included Southern blot analyses to confirm the number of sites of insertion of the transfer-DNA (T-DNA), the absence of plasmid backbone sequences, and the multi-generational stability of the inserted DNA. Nucleotide sequencing of the entire T-DNA insert, including a portion of the 5′ and 3′flanking host genomic region was carried out in order to demonstrate overall integrity of the insert, contiguity of the functional elements, and to detect any individual base-pair changes. Bioinformatic analyses were used to investigate the possibility of the unintended creation of any novel open reading frames (ORFs) spanning the inserted and genomic DNA junctions, potentially encoding allergens or toxins. Finally, segregation analyses served to demonstrate Mendelian inheritance of the introduced genes and traits.

Assessing the safety of newly expressed proteins produced in the edible portions of a genetically engineered food crop is an integral component of the overall safety assessment. There is currently no single criterion that is sufficiently predictive of potential toxicity or allergenicity. Therefore, a "weight-of-evidence" approach is recommended for hazard identification that considers the history of safe use, amino acid sequence similarity to known toxins or allergens, function or mode of action, digestibility under standardized *in vitro* conditions, stability to heat or processing, and expression levels and potential dietary exposure[12,13]. Depending on the results from early tier evaluation, additional characterization may include appropriate oral toxicity studies or other hypothesis-based toxicology studies. This approach was followed in assessing the safety of the phytoene synthase (*Zm*PSY1) and carotene desaturase (CRTI) proteins expressed in GR2E rice, and it is described in the following sections.

Based on the presence of PMI in a wide range of food and feedstuffs derived from genetically engineered maize lines, and on the extensive history of prior regulatory reviews[14] in some countries, additional characterization of the protein was unnecessary. Previously submitted safety studies[15–19] reviewed in the context of other genetically engineered plant events are directly applicable to the safety assessment of PMI protein expressed in GR2E rice.

The data on the molecular characterization and safety assessment of the proteins in GR2E was part of the regulatory dossier that is intended to meet the regulatory requirements in target countries such as Philippines, Bangladesh, and Indonesia. Other countries may have other requirements.

Results

Southern blot characterization of GR2E rice. *Hin*dIII and *Sph*I both with unique restriction sites within the T-DNA (Supplementary Fig. 1, adapted from submitted GR2E-FFP (Food and Feed or for Processing) study reports) were used for determining the copy number of the introduced DNA within the GR2E genome. Hybridization of *Hin*dIII-digested genomic DNA from homozygous plants of GR2E in four genetic backgrounds (Kaybonnet, PSBRc82, BRRI *dhan*29, and IR64) with the *Zmpsy1* or *pmi* probes resulted in the detection of a single fragment of *ca*. 7900 bp, and hybridization with the probe specific for *SSU-crtI* yielded a single *ca*. 7200 bp fragment (Supplementary Table 2, Supplementary Fig. 2, lanes 16–19, adapted from GR2E-FFP submitted study reports). Hybridization of *Sph*I-digested GR2E genomic DNA with the *Zmpsy1* or *SSU-crtI* probes gave a single fragment of *ca*. 6900 bp, and upon hybridization with the *pmi* probe a single fragment of *ca*. 5500 bp was observed (Supplementary Fig. 2, lanes 6–9, adapted from GR2E-FFP submitted study reports; Supplementary Table 2). Weak hybridization between the *Zmpsy1* probe and sequences derived from the endogenous rice *psy1* gene was detected for restriction enzyme digests of control Kaybonnet and GR2E DNA samples (Supplementary Fig. 2, panel A, adapted from GR2E-FFP submitted study reports). Hybridizing fragments of *ca*. 5600 bp and *ca*. 4900 bp were detected in Southern blots of *Sph*I and *Hin*dIII-digested DNA samples, respectively. This was not an unexpected finding considering the high degree of sequence identity, ca. 83 percent, shared between the *Zmpsy1* and *Oryza sativa psy* genes.

Southern blot analyses of *Asc*I + *Xma*I-digested GR2E rice DNA were used to investigate the integrity of the T-DNA insert containing the *Zmpsy1*, *SSU-crtI*, and *pmi* gene cassettes. The T-DNA contains a single *Asc*I restriction site located at position 199 and a single *Xma*I site at position 8946 (Supplementary Fig. 1, adapted from GR2E-FFP submitted study reports). Insertion of an intact copy of the T-DNA should thus result in the detection of an 8747 bp *Asc*I + *Xma*I fragment with the *Zmpsy1*, *SSU-crtI*, and *pmi* probes. The results of Southern analyses (Supplementary Fig. 2, lanes 11–14, adapted from GR2E-FFP submitted study reports) demonstrate that the correct size fragment was detected with all of the hybridization probes (Supplementary Table 2).

Hybridizing fragments were not detected when backbone probes were tested against samples of *Asc*I + *Xma*I-digested GR2E genomic DNA (Supplementary Fig. 3, lanes 6–8, panels A and B, adapted from GR2E-FFP submitted study reports), confirming the absence of plasmid backbone sequences. Positive control samples of wild-type Kaybonnet genomic DNA spiked with pSYN12424 plasmid DNA did result in detection of the expected-size 4349 bp fragment using backbone probe 5 (Supplementary Fig. 3, panel B, lane 3, adapted from GR2E-FFP submitted study reports) and two fragments of 1243 bp and 4349 bp, respectively, using the mixture of backbone probes 1–4 (Supplementary Fig. 3, panel A, lane 3, adapted from GR2E-FFP submitted study reports).

Thus, multiple Southern hybridization analyses clearly demonstrate the insertion of the T-DNA into a single site and the absence of sequences derived from the plasmid backbone.

Stability of the introduced trait across multiple generations. The stability of the inserted DNA across multiple generations was assessed by Southern blot analyses of genomic DNA samples prepared from a selfed generation of GR2E in Kaybonnet background (T_n) and three back-cross generations of GR2E (BC_3F_5, BC_4F_3, and BC_5F_3) for each recurrent parents (BRRI *dhan* 29, IR64 and PSBRc82). Digestions with *Hin*dIII, *Sph*I and *Asc*I + *Xma*I were separated by gel electrophoresis and blots were probed with probes specific for *Zmpsy1*, *SSU-crtI*, or *pmi* genes, respectively (Fig. 1, adapted from GR2E-FFP submitted study reports). Single hybridizing fragments of ~7900 bp, ~6900 bp, or 8747 bp were detected using the *Zmpsy1, SSU-crtI*, or *pmi* probes, respectively, in corresponding blots of *Hin*dIII, *Sph*I, or *Asc*I + *Xma*I digests of genomic DNA from each generation of GR2E rice (Supplementary Table 3).

Concentrations of total carotenoids were determined in grain samples collected from GR2E plants in Kaybonnet germplasm, the BC_3F_5 generations in PSBRc82 and IR64 backgrounds, and the BC_4F_3 and BC_5F_3 generations in PSBRc82, IR64, and BRRI *dhan* 29 germplasm backgrounds (Table 1). Carotenoid accumulation in the endosperm was positively correlated with the presence of the T-DNA insert as previously established by Southern blot characterization of the same generations and germplasm backgrounds of GR2E rice. Some variation in the concentrations of total carotenoids was observed depending on the germplasm background, with Kaybonnet and BRRI *dhan* 29 GR2E attaining the highest levels.

Mendelian inheritance of the inserted DNA. The inheritance pattern of the T-DNA insert within GR2E rice was investigated using a polymerase chain reaction (PCR)-based zygosity test. Segregation of the insert within three segregating generations (BC_4F_2, BC_5F_1, and BC_5F_2) in each of three genetic backgrounds was determined. Chi-square analysis resulted in no statistically significant differences between the observed and expected segregation ratios for the three segregating generations of GR2E in PSBRc82, BRRI *dhan*29, and IR64 genetic backgrounds (Supplementary Table 5).

Nucleotide sequence analysis of the inserted DNA and flanking regions. The nucleotide sequence of the plasmid T-DNA, together with preliminary sequence information from the 5′ and 3′ flanking genomic DNA, was used to design seven sets of oligonucleotide primers to amplify the insert and flanking regions from GR2E genomic DNA as seven individual overlapping fragments (Supplementary Fig. 4, adapted from GR2E-FFP submitted study reports).

In total, 12,772 bp of GR2E genomic sequence was determined, comprising 1,988 bp of the 5′ genomic border sequence, 1,788 bp of the 3′ genomic border sequence, and 8,996 bp of the inserted T-DNA (Supplementary Fig. 4, adapted from GR2E-FFP submitted study reports). The T-DNA in GR2E rice was found to have a 23 bp deletion at the right border end and an 11 bp deletion at the left border end, which is common for *Agrobacterium*-mediated transformation events[20]. All remaining sequence was intact and identical to that of the T-DNA region of plasmid pSYN12424.

Basic local alignment search tool 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) localized the T-DNA on chromosome 3 within the intergenic region between LOC_Os03g43980 (3′ proximal) and LOC_Os03g43990 (5′ proximal; Fig. 2, adapted from GR2E-FFP submitted study reports).

To investigate the possibility of creating new ORFs as a consequence of the T-DNA insertion in GR2E, an open reading frame analysis was conducted to look for potential start-to-stop ORFs that spanned either the 5′ or 3′ junctional regions. This analysis examined each of three possible reading frames in both orientations (i.e., six possible reading frames in total) for potential ORFs capable of encoding sequences of 30 or more amino acids. An allergen usually contains at least two epitopes, each of which will be a minimum of approximately 15 amino acid residues long, in order that antibody binding could occur. This implies a lower size limit for protein allergens of approximately 30 amino acid residues[21], although currently there is no consensus among scientist on such size limit. Two ORFs were identified, one in the reverse orientation that spanned the 5′ T-DNA insert–genomic DNA border (Supplementary Fig. 4, adapted from GR2E-FFP submitted study reports; ORF-1, 207 bp, 68 amino acids), and one in the forward orientation that spanned the 3′ T-DNA insert–genomic DNA border (Supplementary Fig. 4, adapted from GR2E-FFP submitted study reports; ORF-2, 240 bp, 79 amino acids).

To search for potential similarity to known toxins, the amino acid sequence of each ORF was queried against a toxin database using the FAST All sequence alignment tool 36 (FASTA36) to identify possible significant sequence

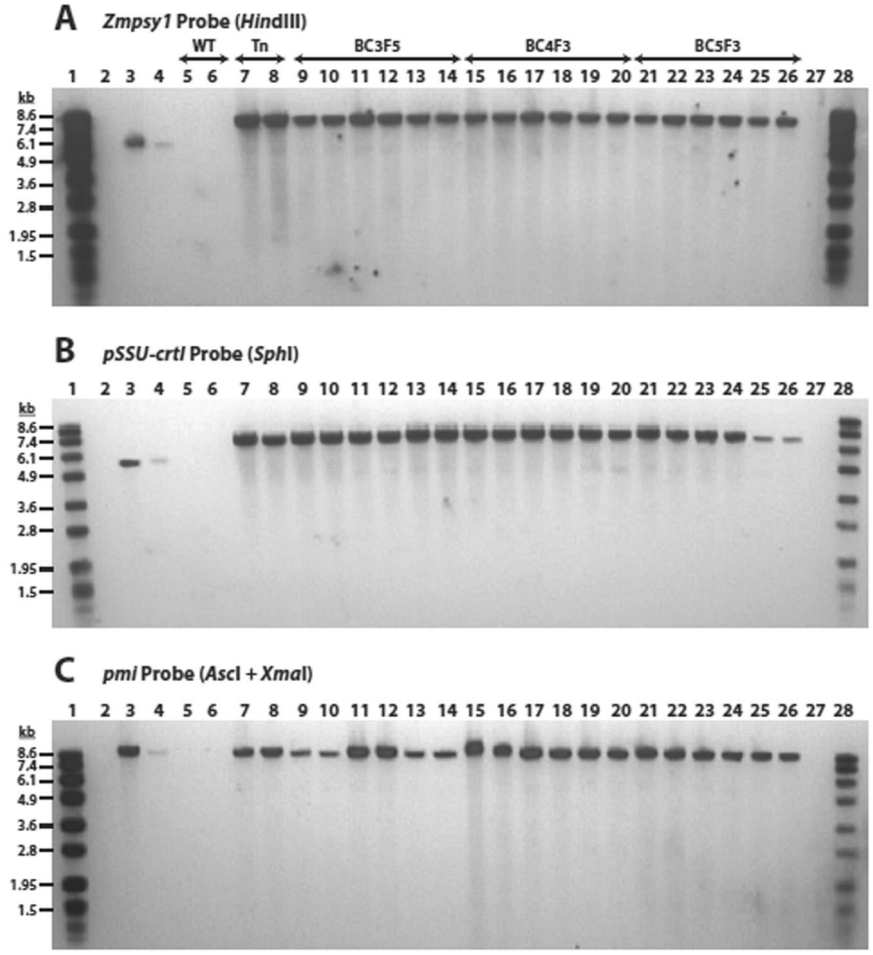


Figure 1. Samples of DNA from individual plants of event GR2E in Kaybonnet (Tn; lanes 7–8, as replicates), BRRI dhan 29 (BC_3F_5, lanes 9–10; BC_4F_3, lanes 15–16; and BC_5F_3, lanes 21–22), IR64 (BC_3F_5, lanes 11–12; BC_4F_3, lanes 17–18; and BC_5F_3, lanes 23–24), and PSB Rc82 (BC_3F_5, lanes 13–14; BC_4F_3, lanes 19–20; and BC_5F_3, lanes 25–26) germplasm backgrounds and negative control DNA from Kaybonnet rice (lanes 5–6) were subjected to Southern blot analysis. For this, 5 μg genomic DNA was digested with HindIII (panel A), SphI (panel B), or AscI plus XmaI (panel C) followed by agarose gel electrophoresis and transfer onto nylon membrane. Positive control samples consisted of negative control Kaybonnet rice containing either one (lane 3) or 0.2 (lane 4) copy equivalents of pSYN12424 plasmid DNA digested with SphI (panels A and B) or AscI + XmaI (panel C), Blots were hybridized with DIG-labelled probes for Zmpsy1 (panel A), pSSU-crtI (panel B), or pmi (panel C). Following washing, hybridized probes and DIG-labelled molecular weight markers VII (lanes 1 and 28) were visualized using a chemiluminescent detection. Lanes 2 and 27 were blank on all gels (adapted from GR2E-FFP submitted study reports).

Breeding Generation	Total Carotenoids (μg/g FWT)[†]			
	Kaybonnet	PSB Rc82	IR64	BRRI *dhan* 29
T(n)	30.50 ± 2.49	—	—	—
BC_3F_5	—	12.84 ± 3.32	20.11 ± 8.71	ND[a]
BC_4F_3	—	9.09 ± 2.26	19.40 ± 2.00	24.50 ± 2.85
BC_5F_3	—	14.12 ± 0.11	13.23 ± 1.16	29.33 ± 3.14

Table 1. Concentrations of total carotenoids in different generation and germplasm backgrounds of GR2E rice. [†]Values shown are mean values ± SD (standard deviation) determined spectrophotometrically for total carotenoids in grain samples after one day of storage at 16 °C. All concentrations are on a fresh weight (FWT) basis, not corrected for moisture content. [a]ND = Not determined. Due to poor quality of remnant seed from the BC_3F_5 generation of BRRI *dhan* 29 containing GR2E, there was no seed germination and plants could not be produced for grain sampling.

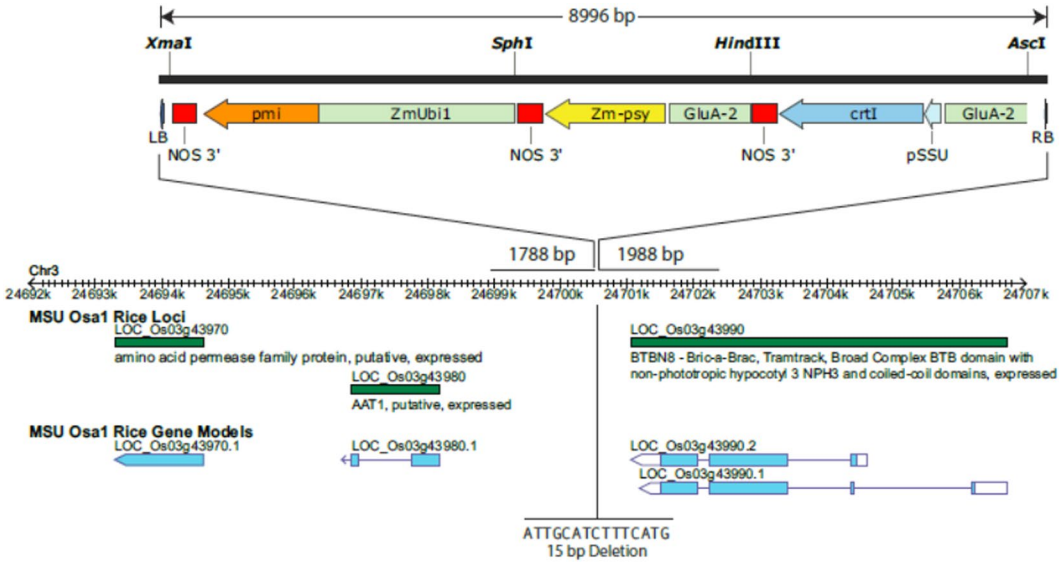


Figure 2. Map position is indicated according to the MSU Rice Genome Annotation Project Release 7 (Nipponbare). The locations of the LB and RB flanking sequences correspond to positions 24,698,762–24,700,549 and 24,700,565–24,702,552, respectively. The insertion of the pSYN12424 T-DNA was within an intergenic region between loci LOC_Os03g43980 and LOC_Os03g43990, and resulted in the deletion of 15 bp of host genomic DNA in addition to truncations of the LB and RB regions of 11 bp and 23 bp, respectively (adapted from GR2E-FFP submitted study reports).

similarity with known or potential toxins. An *E*-score criterion of 1×10^{-5} was used to identify sequences from the toxin database with potential for significant sequence similarity to the query sequences of each ORF. Typically, alignments between two sequences require an E-score of 1×10^{-5} or less to be considered to have sufficient sequence similarity to infer homology. The FASTA36 search resulted in no significant hits returned (Supplementary Table 6).

To assess the potential for allergenicity, the amino acid sequence of each ORF was compared to a peer-reviewed database of 2129 known and putative allergens and celiac protein sequences residing in the Food Allergy Research and Resource Program (FARRP) dataset version 19 at the University of Nebraska. A criteria of $>35\%$ identity over any segment of 80 or more amino acids as an indication of possible cross-reactivity for allergens was adopted by the Codex[12] as the primary sequence search criteria for use in flagging proteins that might be of some concern of cross-reactivity for genetically modified plants. No identity matches of greater than 35 percent over 80 residues were observed for either ORF-1 or ORF-2. Each query sequence was also evaluated for any eight contiguous identical amino acid matches to the allergens contained in the FARRP database. There were no eight contiguous identical amino acid matches observed for either ORF-1 or ORF-2 (Supplementary Table 6).

Novel protein expression. The tissue specificity of *Zm*PSY1, CRTI, and PMI expression was confirmed by immunoblot analysis of various tissues sampled from GR2E rice. The patterns of expression of proteins in GR2E tissues were consistent with the activity of the endosperm-specific rice *GluA-2* promoter of the *Zmpsy1* and *crtI* genes, and use of the constitutive maize polyubiquitin promoter for the *pmi* gene. Expression of *Zm*PSY1 and CRTI was detected only in milk, dough, and mature stage grain (Supplementary Fig. 5, lanes 3–5, panels B and D, adapted from GR2E-FFP submitted study reports) and not in samples of bran, hulls, leaf, stem, or root tissue. In comparison, PMI expression was detected in all rice tissues tested (Supplementary Fig. 5, lanes 3–10, panel F, adapted from GR2E-FFP submitted study reports).

In order to estimate potential human and animal dietary exposure to the *Zm*PSY1, CRTI, and PMI enzymes expressed in GR2E, the protein concentrations in plant tissues were determined by quantitative enzyme-linked immunosorbent assay (ELISA). Three replicated samples of grains (milky, dough, mature) and straw were collected from GR2E rice grown at four locations in the Philippines during two growing seasons in 2015–16. Expression of the *Zm*PSY1 and CRTI proteins in GR2E is driven by the endosperm-specific rice *GluA-2* promoter and measurable concentrations of both these proteins were found in all grain developmental stages but not in stem tissue (straw; Supplementary Table 7). For each protein, the highest concentrations were measured in samples of dough-stage grain, ranging between *ca.* 308–359 ng/g fresh weight tissue (FWT) and between ca. 54–68 ng/g FWT for *Zm*PSY1 and CRTI, respectively, across both growing seasons. Across the four locations and two growing seasons, the highest concentrations of *Zm*PSY1 and CRTI measured in samples of mature grain were *ca.* 245 ng/g FWT and 30 ng/g FWT, respectively.

Concentrations of PMI protein were significantly higher than either *Zm*PSY1 or CRTI in samples from all grain growth stages (Fig. 3, adapted from GR2E-FFP submitted study reports), and were highest in dough-stage grain, averaging *ca.* 2015 ng/g FWT across the four locations over both growing seasons. The mean PMI concentration in mature GR2E rice grain samples was *ca.* 1282 ng/g FWT across both growing seasons (Supplementary

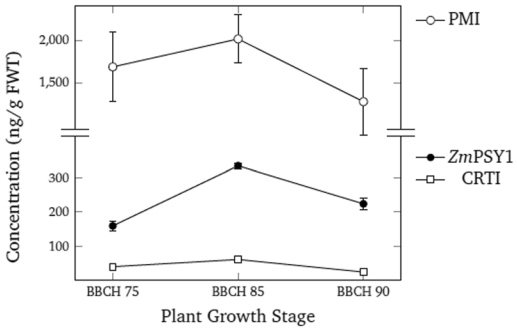


Figure 3. Samples of GR2E grain were collected at different developmental stages [BBCH 75 (milk stage), BBCH 85 (dough stage), and BBCH 90 (mature stage)] from four locations over two growing seasons in 2015–16 and the concentrations of ZmPSY1, CRTI, and PMI were determined by quantitative ELISA and are given in ng g/fresh weight tissue (FWT). Values represent the mean concentration across locations and years for each protein, and the error bars represent the range of concentrations measured across locations over both growing seasons. In some cases, the size of the error bars was less than the symbol size used for plotting (adapted from GR2E-FFP submitted study reports).

Table 7). Since expression of the PMI protein was under control of the constitutive maize polyubiquitin promoter, it was also present in straw samples at concentrations ranging between 320–796 ng/g FWT depending on location and growing season. The average concentration of PMI protein in GR2E straw across both growing seasons was *ca*. 482 ng/g FWT.

Estimated human daily dietary exposure to ZmPSY1, CRTI, and PMI proteins. Two approaches were followed to obtain estimates of daily rice consumption. First, historic rice utilization data for the highest rice-consuming countries in Asia, in comparison with the United States, were obtained from the USDA Production Supply and Distribution database and converted to per capita utilization estimates using the FAOSTAT population database. These values for 2011–2015 are presented in Supplementary Table 8. Projected utilization values for the same countries were obtained from the International Rice Outlook: International Rice Base Projections[22], also presented in Supplementary Table 8 and Supplementary Fig. 6 (adapted from GR2E-FFP submitted study reports).

Using the highest projected per capita rice utilization in Cambodia of 253 kg/yr and an estimated average adult body weight of 57.7 kg in Asia[23], the maximum daily rice intake was calculated as shown in this equation:

$$\text{Daily Rice Intake} = \frac{253\ (\text{kg/yr})}{365 \times 57.7\ (\text{kg BW})} \text{X}\ 1000\ (\text{g/kg}) = 12.0\ (\text{g/kg body weight})$$

The second approach utilized data from the Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) Chronic Individual Food Consumption Database summary statistics (CIFOCOss) currently containing summary statistics of 37 surveys from 26 countries. The CIFOCOss was initially developed to be used by FAO/WHO scientific committees for dietary exposure assessment. Available data for Asian countries are shown in Supplementary Table 9. A further comparison of consumption data between Asian countries and selected European, African, and South American countries is shown in Supplementary Fig. 7 (adapted from GR2E-FFP submitted study reports).

Based upon consideration of both approaches, a value of 12.5 g/kg body weight was chosen as the upper limit of mean daily dietary intake of rice. This value was judged as sufficient to account for consumption by all population subgroups, including children. In deriving estimates of maximum potential daily dietary exposure to the *Zm*PSY1, CRTI, and PMI proteins expressed in GR2E rice, the following assumptions were used: (i) The mean daily dietary rice consumption is 12.5 g/kg body weight, (ii) 100% percent of the dietary rice intake is from GR2E rice and (iii) the grain concentrations of *Zm*PSY1, CRTI, and PMI used for estimation are the highest values measured. This is the case in samples of dough-stage grain collected from any individual trial site location in either 2015 or 2016. These concentrations were significantly higher than those measured in mature grain at harvest. Using these assumptions, the estimated maximum potential daily dietary exposure from GR2E rice to each novel protein is shown in Table 2. They are estimated to be *ca*. 4.5, 0.85, and 30 μg/kg body weight to *Zm*PSY1, CRTI, and PMI proteins, respectively.

Bioinformatic analysis of the ZmPSY1 and CRTI protein amino acid sequences. PSY plays a pivotal role in the carotenoid biosynthesis pathway as it catalyzes the first committed step and controls flux through the pathway[24,25]. Phytoene undergoes consecutive modifications such as desaturation reactions by carotene desaturases and cis-trans isomerization reactions to form all-trans-lycopene, which is cyclized to α- and β-carotene.

Potential identities between the *Zm*PSY1 query sequence and proteins in the allergen database were evaluated with the FASTA35 sequence alignment tool using the default parameters. A criteria of > 35% identity over any segment of 80 or more amino acids as an indication of possible cross-reactivity to allergens was adopted by the Codex[12] as the primary sequence search criteria for use in flagging proteins that might be of some concern

Protein	Concentration (ng/g FWT)	Daily Dietary Exposure (μg/kg body weight)†
*Zm*PSY1	359	4.49
CRTI	68	0.85
PMI	2397	29.96

Table 2. Estimated maximum potential daily dietary exposure to *Zm*PSY1, CRTI, and PMI. †Daily dietary exposures (μg/kg body weight) to *Zm*PSY1, CRTI, and PMI were calculated based on daily *per capita* rice consumption of 12.5 g/kg body weight.

of cross-reactivity for genetically modified plants. No identity matches of > 35 percent over 80 residues were observed. Also, there were no instances of eight contiguous identical amino acid matches observed between the amino acid sequence of *Zm*PSY1 when compared with the sequences of known allergenic proteins. To search for similarity to known or potential toxins, the amino acid sequence of the *Zm*PSY1 was queried against a toxin database using the FASTA36 algorithm. The *Zm*PSY1 query sequence did not return any entries with *E-score* less than 1×10^{-5}. Therefore, there are no sequence homology alerts for potential toxicity of the *Zm*PSY1 protein.

Potential identities between the CRTI query sequence and proteins in the allergen database were evaluated and no identity matches of > 35 percent over 80 residues were observed, nor were there any instances of eight contiguous identical amino acid matches observed between the CRTI amino acid sequence and sequences of known allergenic proteins. However, a search using the CRTI query sequence returned two protein accessions from the toxin database with an *E-score* less than 1×10^{-5}. The two sequence alignments (Supplementary Fig. 8, adapted from GR2E-FFP submitted study reports) were to the conserved N-terminal FAD (flavin adenine dinucleotide) -binding regions of L-amino acid oxidase (LAAO) enzymes from two species of venomous snakes: *Bungarus multicinctus* (many-banded krait, also known as the Taiwanese krait or the Chinese krait) and *B. fasciatus* (banded krait). Homology of these proteins will be discussed later.

Rapid digestion of ZmPSY1 and CRTI in simulated gastric fluid (SGF). Rapid gastric and intestinal digestion is known to be correlated to the allergenic potential of proteins[26]. The *in vitro* pepsin resistance of native i.e. enzymatically active *Zm*PSY1 protein was investigated. Samples were removed at the given stated time points and subjected to SDS-PAGE analysis. 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) was evident as assessed by either SDS-PAGE or western immunoblot analysis (Supplementary Fig. 9, lane 4 in panels A and B, respectively, adapted from GR2E-FFP submitted study reports). Faint, low molecular mass degradation products were visible by Coomassie staining in samples removed up to two minutes of digestion (Supplementary Fig. 9, lanes 4–6, panel A, adapted from GR2E-FFP submitted study reports), but not at later time points, and these were not detected in the western blot.

Similar results were obtained with CRTI. Following exposure to SGF containing pepsin for 30 seconds, the earliest time point sampled during the digestion, no intact CRTI protein was evident as assessed by either SDS-PAGE or western immunoblot analysis (Fig. 4, lane 4 in panels A and B, respectively, adapted from GR2E-FFP submitted study reports), and there was no evidence of stable lower molecular mass proteolytic fragments derived from CRTI.

Heat stability of ZmPSY1 and CRTI protein. The thermal stability of the *Zm*PSY1 protein was evaluated by measuring enzymatic activity, i.e. the conversion geranylgeranyl diphosphate (GGPP) into 15-*cis*-phytoene, as monitored by HPLC analysis. The GGPP substrate was produced with GGPP-synthase from its precursor molecules dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP)[24,27]. Thermal instability, i.e. rapid denaturation greatly increases the chance for proteolytic cleavage, adding to the safety of the expressed proteins. Proteins that are labile to temperatures used in cooking and processing are likely to have negligible dietary exposure. The expressed purified *Zm*PSY1 catalyzed 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 (Fig. 5, adapted from GR2E-FFP submitted study reports). Enzyme activity was irreversibly destroyed upon heat treatment, with 50 percent loss of activity following pre-incubation at *ca*. 42 °C for 15 minutes and complete loss of activity at 50 °C for 15 minutes.

The thermal stability of the CRTI protein was evaluated by measuring enzymatic activity using a spectrophotometric assay to monitor the conversion of liposome-incorporated 15-*cis*-phytoene to all-*trans*-lycopene according to assay conditions[28]. The purified CRTI was enzymatically active, catalyzing the conversion of liposome-incorporated phytoene to all-*trans*-lycopene at the rate of *ca*. 5.4 pmol all-*trans*-lycopene μg^{-1} min^{-1} under the assay conditions used (Supplementary Fig. 10, adapted from GR2E-FFP submitted study reports). Enzyme activity was irreversibly destroyed upon heat treatment, with 50 percent loss at *ca*. 51 °C for 15 minutes and complete loss of activity following pre-incubation at 55 °C for 15 minutes.

Lack of acute toxicity of CRTI protein. The potential for acute toxicity resulting from a single oral exposure to CRTI was investigated in mice. Groups of five male and five female mice were dosed orally by gavage with: formulation buffer; bovine serum albumin or purified microbial-expressed CRTI (100 mg/kg body weight actual dose), at a volume of 13.45 ml/kg body weight, administered in two separate doses approximately 4 hours apart on test day 1. All animals survived until the scheduled end of the study period on day 15 and there were no clinical signs (abnormal behavior, general appearance and mortality/moribundity) of toxicity observed during the test

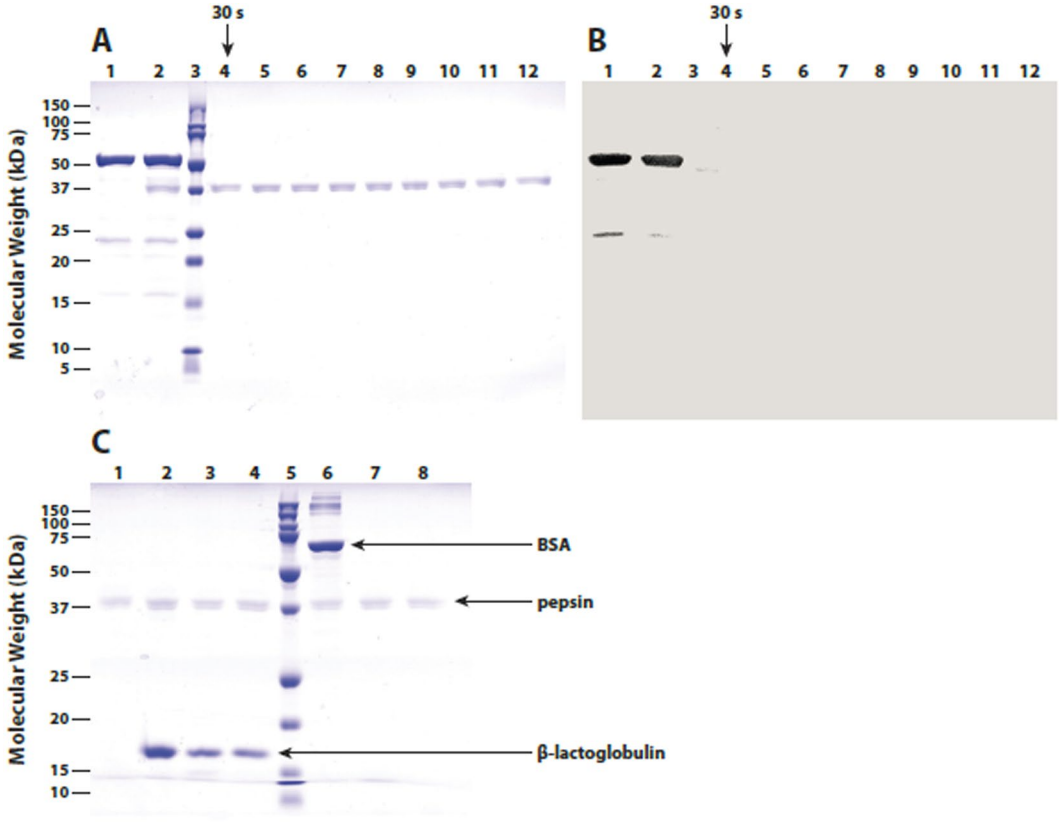


Figure 4. Panels A and B: Samples of CRTI protein purified from recombinant E. coli (Lot No. M20454-02) were incubated in the presence of SGF pH 1.2 containing pepsin for 0 min (lane 2) and 0.5, 1, 2, 5, 10, 20, 30 or 60 min at 37 °C (lanes 4–11) and then analyzed by SDS-PAGE. Gels were either stained for protein with colloidal blue G250 (panel A) or subjected to western immunoblot analysis (panel B) using rabbit anti-CRTI immunoglobulin (1:1000) and horseradish peroxidase-conjugated goat anti-rabbit IgG followed by precipitating substrate development. Control samples included CRTI protein diluted in gastric control fluid without pepsin (lane 1) and SGF solution containing pepsin (lane 12). Molecular weight standards are shown in lane 3 (adapted from GR2E-FFP submitted study reports).

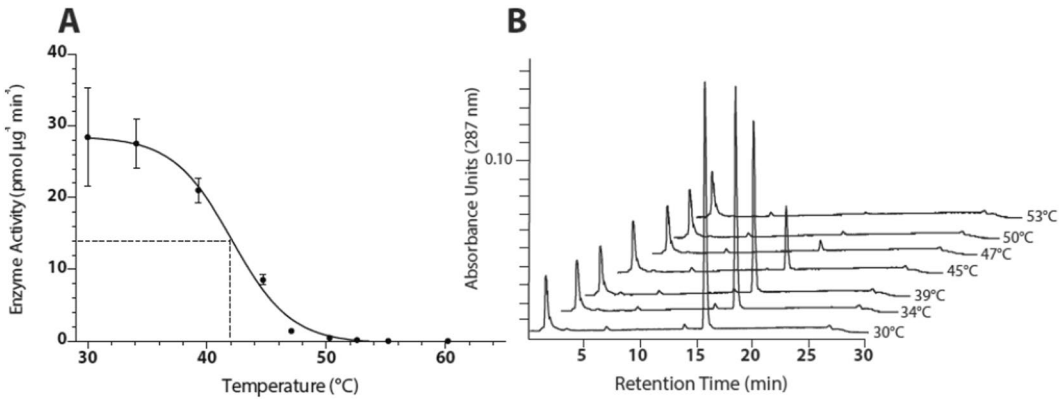


Figure 5. Individual samples of ZmPSY1 protein purified from recombinant E. coli (Lot No. M20452-05) were heated for 15 minutes at a designated temperature ranging from 30–65 °C. Following this treatment, enzymatic production of 15-cis-phytoene was measured by HPLC. Panel A shows enzymatic activity (pmol µg-1 min-1) versus pre-incubation temperature, where the values are means + / − standard deviation of two technical replicates. Panel B shows HPLC chromatograms (287 nm) of chloroform:methanol extracts of selected ZmPSY1 activity assays. The area under the phytoene peak (retention time = 15.7 min) was used to calculate phytoene concentration. Pre-incubation temperatures are indicated to the right of each chromatogram trace (adapted from GR2E-FFP submitted study reports).

period, nor were any gross lesions found in the mice at necropsy. There were no treatment-related effects on body weights for male or female mice over the study duration and all mice experienced net weight gain by test day 15 compared with test day 1 (pre-fast).

Discussion

The purpose of this evaluation of GR2E rice was to determine whether the use of GR2E rice could raise any safety concerns relative to conventional rice. This assessment was not intended to address questions related to the efficacy of GR2E rice in helping combat VAD in affected population sub-groups.

The characterization of the inserted T-DNA by Southern blot analyses and nucleotide sequencing demonstrated that the inserted expression cassettes for the *Zmpsy1*, *crtI*, and *pmi* genes were intact and without sequence changes. They were introduced at a location in the genome that is unlikely to affect the expression of any endogenous genes. Moreover, the two novel ORFs that were created were unlikely to have the potential to encode toxins or allergens. The inserted genes and production of carotenoids were stably inherited across multiple generations according to Mendelian patterns of segregation, consistent with a single locus.

In the pepsin digestion, heat stability assay and ELISA, both *Zm*PSY1 and CRTI used were microbial-expressed proteins. Based on the combination of physiochemical (N-terminal sequence and mass spectral analyses) and enzyme activity analyses, both proteins were functionally equivalent to the *in planta* expressed counterpart, and were suitable surrogate proteins for conducting relevant safety studies.

Zea mays (maize) was the source of the *psy1* gene[29] Maize, a food crop with a long history of safe use, is cultivated worldwide as the third most planted crop after wheat and rice. No significant endogenous toxins are reported to be associated with the genus *Zea*[30]. Food allergy to maize is relatively rare, and the only significant reported food allergen is a nonspecific lipid transfer protein[31]. The *Zmpsy1* gene does not have amino acid sequence similarity to known allergenic and toxic proteins, is not resistant to *in vitro* gastric digestion, and is heat labile. This resulted in the conclusion that further hazard characterization by animal toxicity testing was unnecessary.

Pantoea ananatis was the source of the *crtI* gene[32] and is found in a wide range of natural environments, including water, soil, as part of the epi- and endo-phytic flora of various plant hosts[33]. *Pantoea ananatis* is a ubiquitous bacterium which is found on fresh fruit and vegetables[34]. Many members of the *Enterobacteriaceae*, belonging to the genera *Serratia, Enterobacter, Pantoea, Proteus*, and *Hafnia*, often contribute to meat spoilage[35]. The ubiquity of *P. ananatis* suggests that it has adapted to proliferate in a wide range of environments, and its isolation from both plant and animal hosts indicate that it has adapted for cross-kingdom colonization and pathogenesis[36]. Although strains of *P. ananatis* have been found to be pathogenic on a broad range of plant hosts as well as humans, *crtI* is not part of the genomic material that is associated with the organism's pathogenicity and virulence. There is a possibility of some dietary exposure to CRTI protein at low levels as a consequence of adventitious presence of *P. ananatis* on foods. However, a history of dietary exposure is difficult to establish.

Due to the non-food source of the *crtI* gene, acute oral toxicity testing of CRTI protein in mice was conducted as an additional assurance of safety. Oral administration of CRTI test substance at a concentration representing at least an 115,000-fold margin of exposure relative to any realistically conceivable human dietary intake produced no test substance-related clinical signs of toxicity, body weight losses, gross lesions, or mortality, further substantiating the predicted lack of acute oral toxicity of CRTI. Additionally, CRTI does not have amino acid sequence similarity to proteins known to be allergens or toxic via the oral route of exposure and was readily degraded in the presence of pepsin and inactivated when exposed to heat. A 90 day feeding trial to rodents was not mandatory for approval in the target countries and in some regulatory agencies.

The limited sequence similarity between the *P. ananatis* CRTI protein and the LAAO accessions retrieved from the toxin database was due to homology between N-terminal motifs involved in FAD (flavin adenine dinucleotide) binding, and was not considered to be a structural alert for potential toxicity. LAAOs are flavoenzymes belonging to the class of oxidoreductases that catalyze the stereospecific oxidative deamination of L-amino acids, and snake venom LAAOs are usually homodimeric with cofactors FAD or FMN (flavin mononucleotide) covalently linked to their chemical structure. CRTI also belongs to the flavoprotein superfamily comprising, for instance, protoporphyrinogen IX oxidoreductase and monoamine oxidase and has an absolute requirement for FAD as the sole cofactor in CRTI-mediated phytoene desaturation. Thus, the limited sequence similarity between the CRTI protein and the two L-amino acid oxidase accessions is not surprising and is due to homology between N-terminal motifs involved in FAD binding. Indeed, similar homologies exist between the native rice phytoene desaturase (*Os*PDS) and LAAOs. The two desaturases, CRTI and *Os*PDS, have likely evolved resulting in two different approaches to achieving similar catalytic goals[28].

In summary, the molecular-genetic characterization of GR2E rice, including the assessment of potential toxic or allergic reaction to the newly expressed *Zm*PSY1 and CRTI proteins, when considered together with the comparative compositional assessment[10], support the conclusion that food derived from rice varieties containing event GR2E is as safe as food derived from conventional rice varieties. All the data presented in the regulatory dossier is mainly intended to meet the regulatory requirements in Philippines, Bangladesh, and Indonesia. Other countries may have other requirements.

Methods

Molecular genetic characterization of GR2E rice. Genomic DNA extraction was performed as described by Murray and Thompson[37]. For Southern analyses, genomic DNA samples extracted from selected homozygous GR2E and control rice plants were digested with *Asc*I + *Xma*I to investigate the integrity of the inserted T-DNA, and with either *Hin*dIII or *Sph*I to determine the number of copies of the inserted DNA. Samples were prepared from GR2E in four different germplasm backgrounds: Kaybonnet, BRRI *dhan* 29, IR64, and PSB Rc82. The stability of the inserted DNA across multiple generations of GR2E rice was assessed by DNA blot

analyses of genomic DNA samples prepared from the Tn generation and the BC_3F_5, BC_4F_3, and BC_5F_3 generations of GR2E in BRRI *dhan* 29, IR64, and PSB Rc82 germplasm. Probe DNA was synthesized by PCR according to the procedures supplied in the PCR DIG Probe Synthesis Kit (Roche). Probes for the *Zmpsy1*, *pSSU-crtI*, and *pmi* genes were used to detect genetic elements within the insertion. Probes covering the backbone region of plasmid were used to verify absence of plasmid backbone DNA in GR2E rice (Supplementary Fig. 11, adapted from GR2E-FFP submitted study reports). DNA fragments of the probe elements were generated by PCR from plasmid using specific primers (Supplementary Table 1).

Segregation analysis. Analyses were performed on individual plants from three segregating generations representing the BC_4F_2, BC_5F_1, and BC_5F_2 generations of GR2E rice in PSBRc82, BRRI dhan 29, or IR64 genetic background. DNA samples were analyzed by multiplex PCR to determine zygosity of the T-DNA insert as illustrated in Supplementary Fig. 12 (adapted from GR2E-FFP submitted study reports). Individual PCR reaction mixtures contained 300 ng genomic DNA, 0.5 μM of each primer (ZD-E1-P1, ZD-E1-P2, and Zd-E1-P3), 0.2 mM deoxynucleotide triphosphates, 1.5 mM $MgCl_2$, and 0.5 units KAPA Taq DNA polymerase (KAPA Biosystems) in 1x KAPA Taq buffer. After an initial 5 min denaturation at 95 °C, PCR amplifications were performed for 35 cycles consisting of denaturation (95 °C, 45 sec), annealing (55 °C, 45 sec), and extension (72 °C, 45 sec). A final extension step (72 °C, 8 min) was included following the last PCR cycle, after which samples were held at 25 °C. Following PCR, reaction mixtures were analyzed by agarose gel electrophoresis in 1.5 percent agarose and amplified fragments were visualized with SYBR® Safe DNA Gel Stain (Thermo Fisher Scientific). A Chi-square analysis was performed on the segregation results of each GR2E rice generation to compare the observed segregation ratio to the expected Mendelian segregation ratio. The critical value to reject the hypothesis at the 5 percent level is 3.84.

Nucleotide sequence analysis of GR2E rice. The procedures for cloning the DNA fragments from GR2E and its sequencing are described in Supplementary Method 1.

Enzyme-linked immunosorbent assay. Samples of rice grain were collected at the milk, dough, and mature stages of development[38]. Each individual sample was a composite of material obtained from at least five representative plants. Approximately 3–4 panicles were collected from each representative plant and placed in a pre-labelled net bag, one per block. Following collection of all plants, grains for each sample and growth stage was removed from the panicles, mixed, and ca. 100 g placed in 50 ml screw-top Falcon tubes. Samples of rice straw were collected at harvest. Each individual sample was a composite of material obtained from at least five representative plants. Following collection, straw was chopped into small pieces ca. 12 cm in length, mixed, and ca. 100 g placed in a pre-labelled sample bag. Concentrations of the ZmPSY1, CRTI, and PMI proteins were determined using specific quantitative ELISA methods (GR2E-FFP submitted study reports).

Immunoblot analysis of ZmPSY1, CRTI, and PMI expression in different plant tissues. Frozen tissue samples were ground to a powder in liquid nitrogen using a mortar and pestle. Weighed amounts (200 mg) were re-suspended in either 950 μl (grain, bran, and hulls) or 425 μl (stem, leaf, and root tissue) of 1 × Laemmli sample buffer containing 350 mM dithiothreitol (DTT), vortexed (1 min), and placed on ice for 30 minutes. Tissue extracts were centrifuged (12000 g, 10 min at 4 °C) and supernatant fractions were transferred to new 1.5-ml tubes. The total protein concentration of each sample was determined using the bicinchoninic acid (BCA) protein assay kit (Pierce, Thermo Scientific). Sample extracts containing either 40 μg (ZmPSY1 and CRTI blots) or 7 μg (PMI blots) total protein were subjected to sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) on 10 percent Tris-glycine polyacrylamide gels at 40 V for 30 min followed by 50 V for three hours. For each sample set, one gel was stained with colloidal blue G250 and another was electroblotted onto polyvinylidene fluoride (PVDF) membrane for immuno-labelling. Each sample set contained a positive control sample consisting of non-transgenic Kaybonnet dough-stage grain extract spiked with either purified ZmPSY1 (M20452-05; 2.5 ng), CRTI (M20454-02; 25 ng), or PMI (21038 G; 6.25 ng), and a negative control sample of non-transgenic Kaybonnet dough-stage grain extract.

Membrane blots were incubated for 30 min at room temperature in Tris-buffered saline pH 7.4, 0.05% Tween-20 (TBST) containing 5 percent non-fat dry milk (NFDM) to block non-specific protein binding sites. Membranes were then incubated with either mouse monoclonal antibody specific for ZmPSY1 (2 μg/ml; ELX1048, Envirologix) or CRTI (2 μg/ml; ELX1043, Envirologix) overnight at 4 °C with shaking followed by rinsing with five changes of TBST (5 min each). Rinsed blots were incubated with alkaline phosphatase (AP)-conjugated goat anti-mouse IgG (1:10,000 in TBST + 5% NFDM, 120 min) followed by rinsing as before with TBST. Bound antibody-conjugate was detected using CDP-Star chemiluminescent substrate (Roche) and exposure on Kodak Biomax MS film followed by electronic image capture using a white light transilluminator (Vilber Lourmat, France). For detection of PMI protein, blocked membranes were incubated with horseradish peroxidase (HRP)-conjugated rabbit anti-PMI antiserum (1:100; Romer Labs) followed by chemiluminescent substrate development using luminol substrate (GE Healthcare, USA).

Amino acid sequence similarity search between proteins and known and putative protein toxins and allergens. A FASTA36 bioinformatic search using the ZmPSY1 and CRTI amino acid sequences as the query sequence was performed against a toxin database to identify possible significant sequence similarity with known or potential toxins. The toxin database was created from a subset of sequences derived from the UniProt Knowledgebase, comprised of 550,116 manually annotated and reviewed sequences from Swiss-Prot and 55,270,679 automatically annotated, un-reviewed sequences from TrEMBL[39], that were selected using a keyword search on toxins (KW800). The collection contained a total of 24,098 sequences as of 21 January 2016, comprising 6,588 reviewed sequences from Swiss-Prot and 17,510 un-reviewed sequences from TrEMBL. The BLOSUM50

similarity scoring matrix was used for FASTA36 alignments[40]. An E-score acceptance criterion of 1×10^{-5} was used to identify sequences with potential for significant sequence similarity to the proteins.

To assess the potential for allergenic cross-reactivity, the amino acid sequence encoded by the *Zmpsy1* and *crtI* genes was compared to a database of 2129 known and putative allergen and celiac protein sequences residing in the FARRP19 dataset at the University of Nebraska. Potential identities between the query sequence and proteins in the allergen database were evaluated with the FASTA35 sequence alignment tool using the default parameters. The recommended 35 percent or greater identity threshold over any 80-amino acid length sequence alignment between the query sequence and an allergen was used to indicate the potential for cross-reactivity.

Sequence similarity searches to known Allergens using ORF-1 and ORF-2 sequences as queries are described in Supplementary Method 2.

***In vitro* pepsin digestions.** The *in vitro* pepsin resistance of ZmPSY1 and CRTI protein were investigated by incubating purified ZmPSY1 and CRTI protein for 0, 0.5, 1, 2, 5, 10, 20, 30, and 60 minutes at 37 °C in SGF pH 1.2 containing pepsin. Control digestions with bovine serum albumin (BSA) and beta-lactoglobulin were performed for 0, 1, and 60 minutes under the same conditions. Samples were removed at stated time points and subjected to SDS-PAGE and immunoblot analyses.

Heat stability of ZmPSY1 and CRTI protein. The thermal stability of the ZmPSY1 protein was evaluated by measuring enzymatic activity using a HPLC method to monitor the production of 15-*cis*-phytoene from *in situ* produced geranylgeranyl diphosphate (adapted from GR2E-FFP submitted study reports). Samples of microbial-expressed, purified ZmPSY1 protein were subjected to heat treatment over a temperature incubation range of ca. 30–65 °C for 15 minutes and then used in activity assays. Chloroform extracts of individual reaction mixtures were separated by reverse-phase HPLC and ZmPSY1 enzyme activity was calculated based on phytoene peak area measurements.

The thermal stability of CRTI protein was evaluated by measuring enzymatic activity using a spectrophotometric assay to monitor the conversion of liposome-incorporated 15-*cis*-phytoene to all-*trans*-lycopene. Samples of microbial-expressed, purified CRTI protein (6 μg) were subjected to heat treatment over a temperature incubation range of ca. 30–60 °C for 15 minutes, following measurement of enzyme activity at 37 °C in the presence of 7 μM phytoene, 150 μM flavin adenine dinucleotide, 50 mM Tris-HCl pH 8.0, and 200 mM NaCl.

Acute toxicity of CRTI protein. Groups of five male and five female Crl:CD1(ICR) mice were dosed orally by gavage with: formulation buffer (vehicle control; 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP)); bovine serum albumin (BSA; negative control; target dose 100 mg/kg body weight dissolved in vehicle control solution); or purified microbial-expressed CRTI (100 mg/kg body weight actual dose purified protein dissolved in vehicle control solution), at a volume of 13.45 ml/kg body weight, administered in two divided doses ca. 4 hours apart on test day 1. Body weights were evaluated on test days 1 (pre-fast and shortly prior to administration of the first dose), 2, 3, 5, 8, and 15. Clinical signs (abnormal behavior, general appearance and mortality/moribundity) were evaluated seven times on test day 1 (distributed before and after each dose) and daily thereafter. On test day 15, all mice were euthanized and subject to a gross pathological examination.

GR2E provitamin A biofortified rice regulatory dossier. The complete package of GR2E provitamin A biofortified rice regulatory dossier submissions can be found at https://www.dropbox.com/sh/qpiz0cftefcaceq/AAByIpj_HED3zgqH7ufW7A-ta?dl=0[41]; https://www.foodstandards.gov.au/code/applications/Documents/A1138%20Application_Redacted.pdf[42]

Data availability

All the data has been provided in the manuscript.

Received: 19 July 2019; Accepted: 30 December 2019;
Published online: 28 January 2020

References

1. Scrimshaw, N. S. & SanGiovanni, J. P. Synergism of nutrition, infection, and immunity: an overview. *Am. J. Clin. Nutr.* **66**, 464S–477S (1997).
2. Christian, P., Schulze, K., Stoltzfus, R. J. & West, K. P. Hyporetinolemia, illness symptoms, and acute phase protein response in pregnant women with and without night blindness. *Am. J.f Clin. Nutr.* **67**, 1237–43 (1998).
3. ICDDR. National micronutrients status survey 2011–2012. (UNICEF, Dhaka, 2013).
4. de Pee, S. Proposing nutrients and nutrient levels for rice fortification. *Ann. N. Y. Acad. Sci.* **1324**, 55–66 (2014).
5. La Frano, M., Woodhouse, L., Burnett, D. & Burri, B. Biofortified cassava increases β-carotene and vitamin A concentrations in the TAG-rich plasma layer of American women. *Br. J. Nutr.* **110**, 310–20 (2013).
6. De Moura, F. *et al*. Are Biofortified Staple Food Crops Improving Vitamin A and Iron Status in Women and Children? *New Evidence from Efficacy Trials. Adv. Nutr.* **5**, 568–70 (2014).
7. Burkhardt, P. *et al*. Transgenic rice (*Oryza sativa*) endosperm expressing daffodil (*Narcissus pseudonarcissus*) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. *Plant J.* **11**, 1071–1078 (1997).
8. Ye, X. *et al*. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**, 303–305 (2000).
9. Paine., J. *et al*. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* **23**, 482–487 (2005).
10. Swamy, B. P. M. *et al* Compositional analysis of genetically engineered GR2E "Golden Rice" compared to conventional rice. *J. Agric. Food Chem*. https://doi.org/10.1021/acs.jafc.9b01524 (2019).

11. OECD. Consensus Document on Molecular Characterisation of Plants Derived from Modern Biotechnology. https://www.oecd.org/env/ehs/biotrack/46815346.pdf. (accessed on 2019-10-24). (2010).
12. Codex Alimentarius Commission. Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants CAC/GL 45–2003 (Rome: Codex Alimentarius, Joint FAO/WHO Food Standards Program, 2009).
13. Delaney, B. *et al*. Evaluation of Protein Safety in the Context of Agricultural Biotechnology. *Food Chem. Toxicol.* **46**, 872–897 (2008).
14. FDA. New protein consultation: phosphomannose isomerase (PMI). Technical report, US Food and Drug Administration (2005).
15. Nelson, A. *In vitro* digestibility of phosphomannose isomerase (PMI) as contained in test substance PMI-0105 under simulated mammalian gastric conditions. Technical report, No. SSB-034-07 (unpublished, Syngenta Biotechnology, Inc., North Carolina, 2008).
16. Korgaonkar, C. PMI-0105 - Single-dose oral (gavage) toxicity study in mice with a 14-day observation period. Technical report, No. WIL-639011 (unpublished, Syngenta Biotechnology, Inc., North Carolina, 2009).
17. Mims, G. Effect of temperature on phosphomannose isomerase as contained in test substance PMI-0105 as assessed by enzymatic activity. Technical report, No. SSB-023-09 (unpublished, Syngenta Biotechnology, Inc., North Carolina, 2009).
18. Harper, B. Phosphomannose isomerase (Entrez® accession number AAA24109): assessment of amino acid sequence similarity to known or putative toxins. Technical report, No. SSB-145-11. (unpublished, Syngenta Biotechnology, Inc., North Carolina, 2011a).
19. Harper, B. Phosphomannose isomerase (Entrez® database accession number AAA24109): assessment of amino acid sequence similarity to known or putative allergens. Technical report, No. SSB-102-1. (unpublished, Syngenta Biotechnology, Inc., North Carolina, 2011b).
20. Kim, S.-L., Veena & Gelven, S. Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. *Plant Journal* **51**, 779–791 (2007).
21. Huby, R., Dearman, R. & Kimber, I. Why are some proteins allergens. *Toxicological Sciences* **55**, 235–246 (2000).
22. Wailes, E. & Chavez, E. International Rice Outlook: International Rice Baseline Projections 2014-2024. Technical report (University of Arkansas, Fayetteville, 2015).
23. Walpole, S. *et al*. The weight of nations: an estimation of adult human biomass. *BMC Public Health* **12**, 439 (2012).
24. Welsch, R. *et al*. Provitamin A accumulation in cassava (Manihot esculenta) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* **22**, 3348–3356 (2010).
25. Cazzonelli, C. & Pogson, B. Source to sink: regulation of carotenoid biosynthesis in plants. *Trends in Plant Sci.* **15**, 266–274 (2010).
26. Astwood, J., Leach, J. & Fuchs, R. Stability of food allergens to digestion *in vitro*. *Nature Biotechnol.* **14**, 1269–1273 (1996).
27. Camagna, M. *et al*. Enzyme fusion removes competition for geranylgeranyl diphosphate in carotenogenesis. *Plant Physiol.* **179**, 1013–1027 (2019).
28. Schaub, P. *et al*. On the structure and function of the phytoene desaturase CRTI from *Pantoea ananatis*, a membrane-peripheral and FAD-dependent oxidase/isomerase. *PLoS One* **7**, 6 (2012).
29. Buckner, B., San Miguel, P., Janick-Buckner, D. & Bennetzent, J. The y1 gene of maize codes for phytoene synthase. *Genetics* **143**, 479–488 (1996).
30. IFBC. Biotechnologies and food: assuring the safety of foods produced by genetic modification. Regul. Toxicol. Pharmacol. 12,S1–S196 (1990).
31. Pastorello, E. *et al*. The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. *J. Allergy Clin. Immunol.* **106**, 744–751 (2000).
32. Misawa, N. *et al*. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J. Bacteriol.* **172**, 6704–6712 (1990).
33. Coutinho, T. & Venter, S. *Pantoea ananatis*: an unconventional plant pathogen. *Mol. Plant Pathol.* **10**, 325–335 (2009).
34. Leff, J. & Fierer, N. Bacterial Communities Associated with the Surfaces of Fresh Fruits and Vegetables. *PLoS One* **8**, 1–9 (2013).
35. Ercolini, D., Russo, F., Torrieri, E., Masi, P. & Villani, F. Changes in the spoilage-related microbiota of beef during refrigerated storage under different packaging conditions. *Appl. Environ. Microbiol.* **72**, 4663–4671 (2006).
36. De Maayer, P. *et al*. Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Gen.* **15**, 404 (2014).
37. Murray, M. G. & Thompson, W. F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**, 4321–4326 (1980).
38. Lancashire, P. D. *et al*. A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology.* **119**, 561–601 (1991).
39. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research*, 43, D204–212 (2014).
40. Pearson, W. R. Selecting the Right Similarity-Scoring Matrix. *Current Protocols in Bioinformatics* **43**, 1–12 (2013).
41. Provitamin A Biofortified Rice Event GR2E https://www.dropbox.com/sh/qpiz0cftefcaceq/AAByIpj_HED3zgqH7ufW7A-ta?dl=0; (2017).
42. Provitamin A Biofortified Rice Event GR2E https://www.foodstandards.gov.au/code/applications/Documents/A1138%20Application_Redacted.pdf (2016).

Acknowledgements

This work was made possible through support of grants provided from Bill and Melinda Gates Foundation and US Agency for International Development. The GR2E regulatory dossier submitted can be found at https://www.dropbox.com/sh/qpiz0cftefcaceq/AAByIpj_HED3zgqH7ufW7A-ta?dl=0; https://www.foodstandards.gov.au/code/applications/Documents/A1138%20Application_Redacted.pdf.

Author contributions

N.O. and M.F.C.R. conducted the protein studies. M.F.C.R., N.O. and K.R.T. characterized GR2E plants. K.R.T. and I.S.L. sequenced and analyzed the DNA insertion. D.M. facilitated the acute toxicity studies and made the bioinformatic analyses. M.Swamy and M.Samia did the heritability and trait stability studies. R.W., P.S. and P.B. made the heat stability studies. N.O., M.F.C.R. and D.M. wrote the manuscript. R.R. and R.B. managed the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-57669-5.

Correspondence and requests for materials should be addressed to D.M., I.S.-L. or B.P.M.

Reprints and permissions information is available at www.nature.com/reprints.

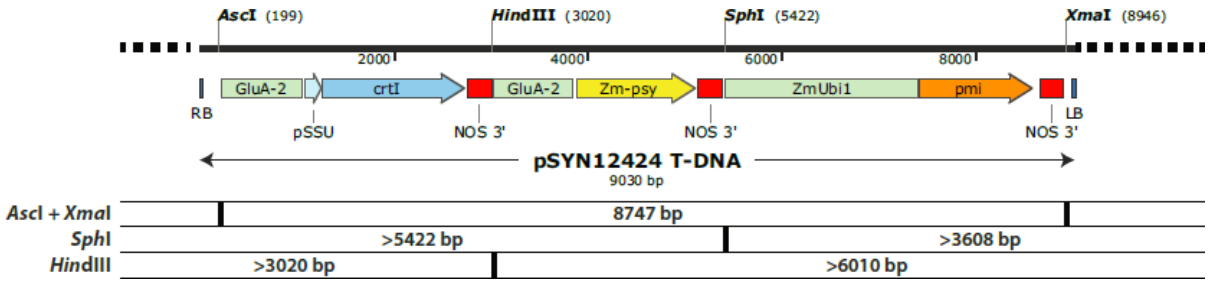
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

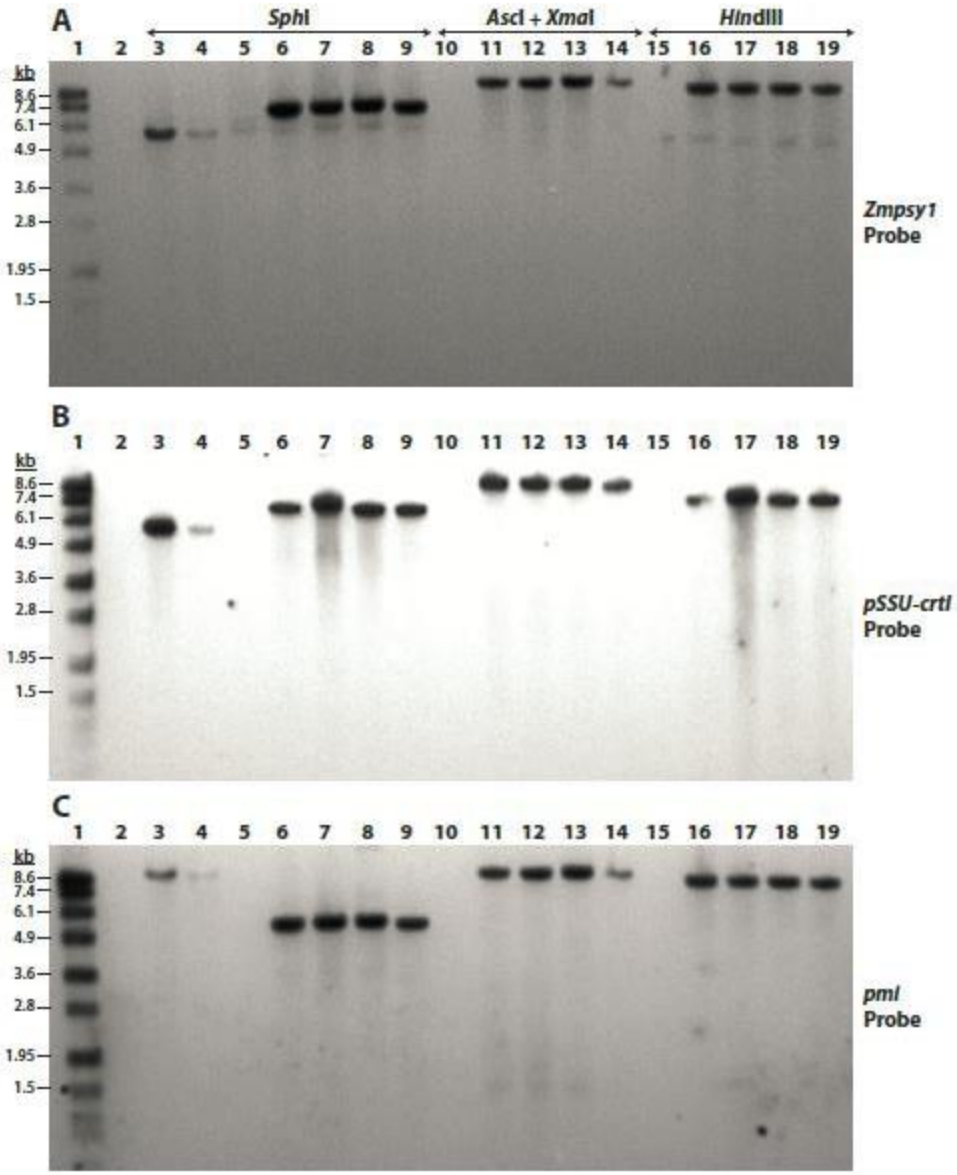
© The Author(s) 2020

Molecular characterization and safety assessment of biofortified provitamin A rice

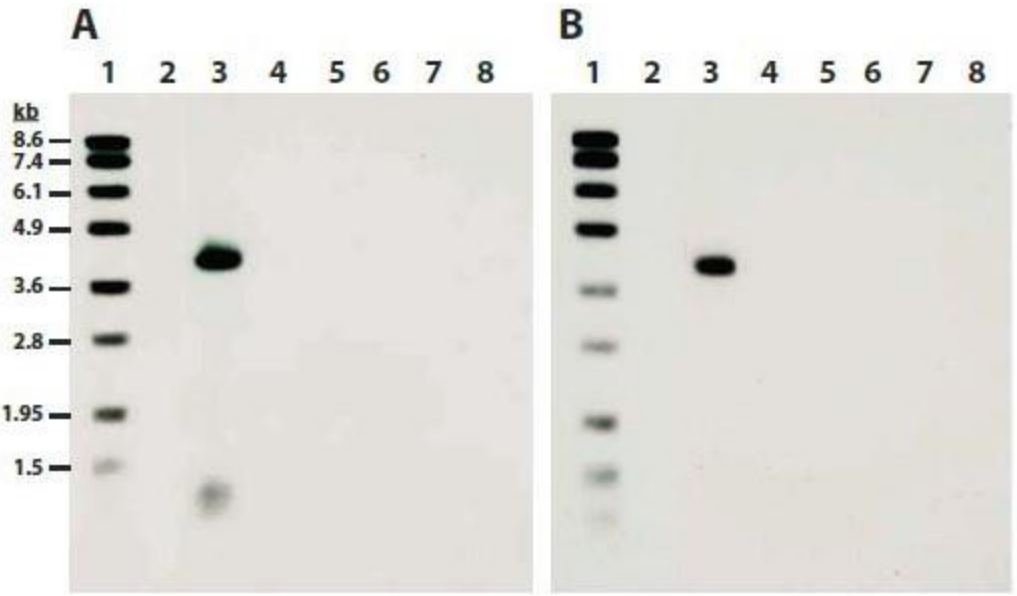
Norman Oliva, Maria Florida Cueto-Reaño, Kurniawan Trijatmiko, Mercy Samia, Ralf Welsch, Patrick Schaub, Peter Beyer, Donald Mackenzie, Raul Boncodin, Russell Reinke, Inez Slamet-Loedin, B.P. Mallikarjuna Swamy



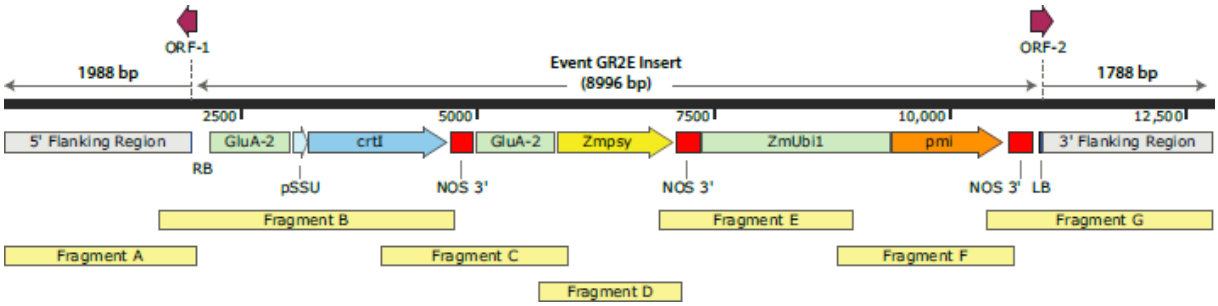
Supplementary Fig. 1 Schematic map of the pSYN12424 T-DNA showing the *Asc*I, *Hin*dIII, *Sph*I, and *Xma*I restriction endonuclease sites used for Southern Hybridization characterization of GR2E rice. The flanking rice genome is represented by the horizontal dotted line. Based on the insertion of a single copy of the T-DNA at a single site, the predicted sizes of restriction fragments following digestion with *Asc*I plus *Xma*I (8747 bp), *Sph*I (>5422 bp and >3608 bp), or *Hin*dIII (>3020 bp and >6010 bp) are shown. Numbering shown on the map is relative to the pSYN12424 T-DNA (adapted from GR2E-FFP submitted study reports)



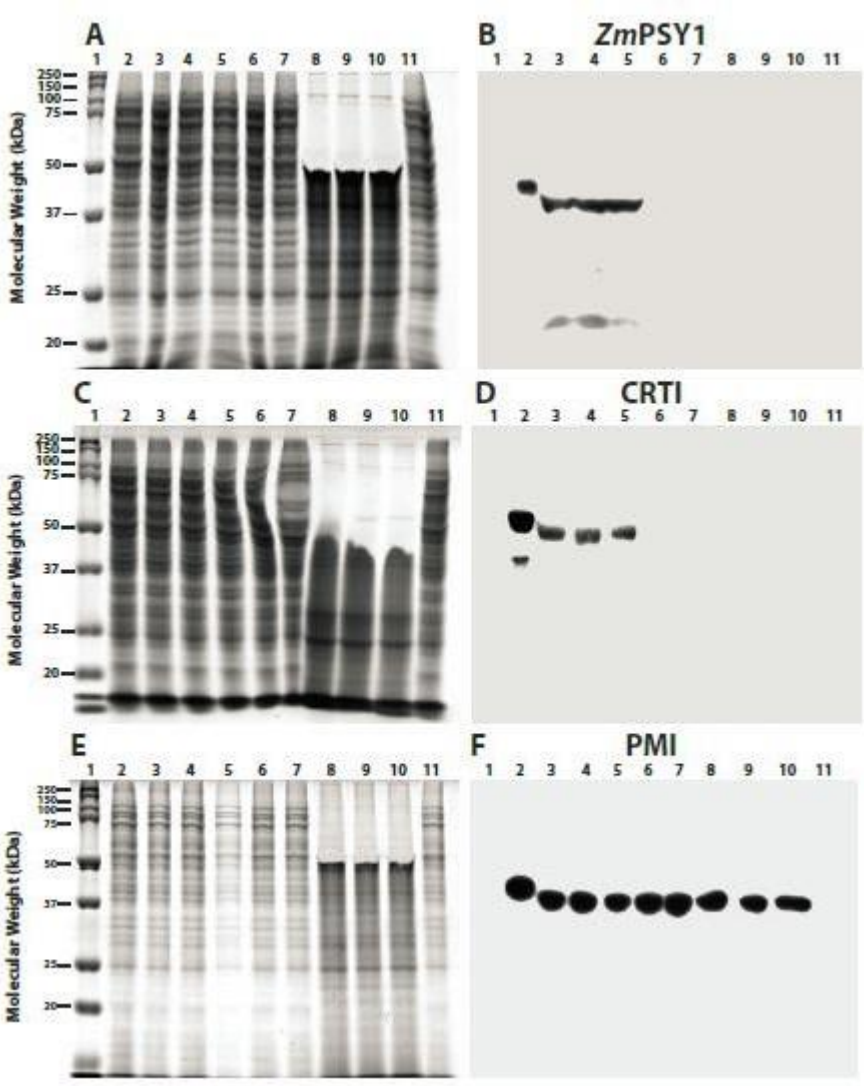
Supplementary Fig. 2 Samples of genomic DNA (ca. 7 µg) from individual plants of event GR2E in Kaybonnet (Tn; lanes 6, 11, and 16), BR29 (BC_5F_3; lanes 7, 12, and 17), IR64 (BC_5F_3; lanes 8, 13, and 18), and PSB Rc82 (BC_5F_3; lanes 9, 14, and 19) germplasm backgrounds; negative control DNA from Kaybonnet rice (lanes 5, 10, and 15); and negative control Kaybonnet rice containing either ca. one (lane 3) or 0.2 (lane 4) copy equivalents of pSYN12424 plasmid DNA, were subjected to restriction endonuclease digestion with *Sph*I (lanes 3–9), *Asc*I plus *Xma*I (lanes 10–14), or *Hin*dIII (lanes 15–19) followed by agarose gel electrophoresis and transfer onto nylon membrane. Blots were hybridized with DIG-labelled probes specific for *Zmpsy1* (panel A), *SSU-crtI* (panel B), or *pmi* (panel C). Following washing, hybridized probes and DIG-labelled molecular weight markers VII (lane 1) were visualized using a chemiluminescent detection system followed by electronic image capture. Lane 2 was blank on all gels (adapted from GR2E-FFP submitted study reports)



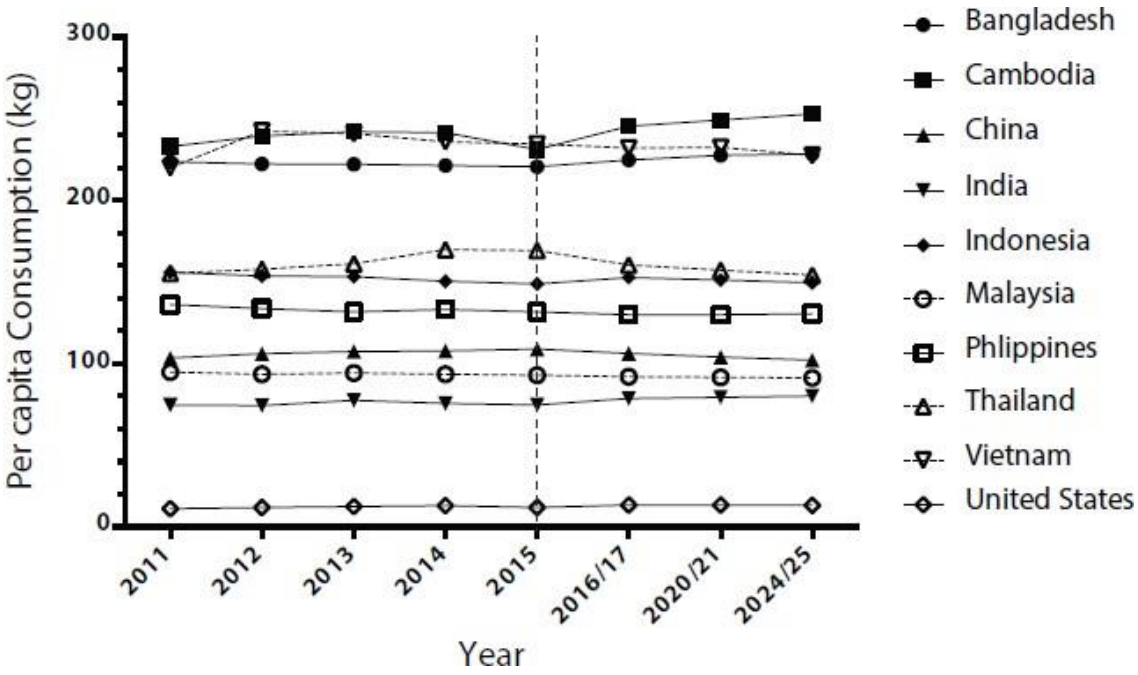
Supplementary Fig. 3 Samples of genomic DNA (ca. 7 µg) from control Kaybonnet rice (lanes 4 and 5), event GR2E in Kaybonnet (Tn; lanes 6–8), and negative control Kaybonnet rice containing three copy equivalents of pSYN12424 plasmid DNA (lane 3), were subjected to restriction endonuclease digestion with *Asc*I plus *Xma*I followed by agarose gel electrophoresis and transfer onto nylon membrane. Blots were hybridized with the mixture of DIG- labelled backbone probes 1–4 (panel A) or backbone probe 5 (panel B). Following washing, hybridized probes and DIG-labelled molecular weight markers VII (lane 1) were visualized using a chemiluminescent detection system followed by electronic image capture. Lane 2 was blank on all gels (adapted from GR2E-FFP submitted study reports)



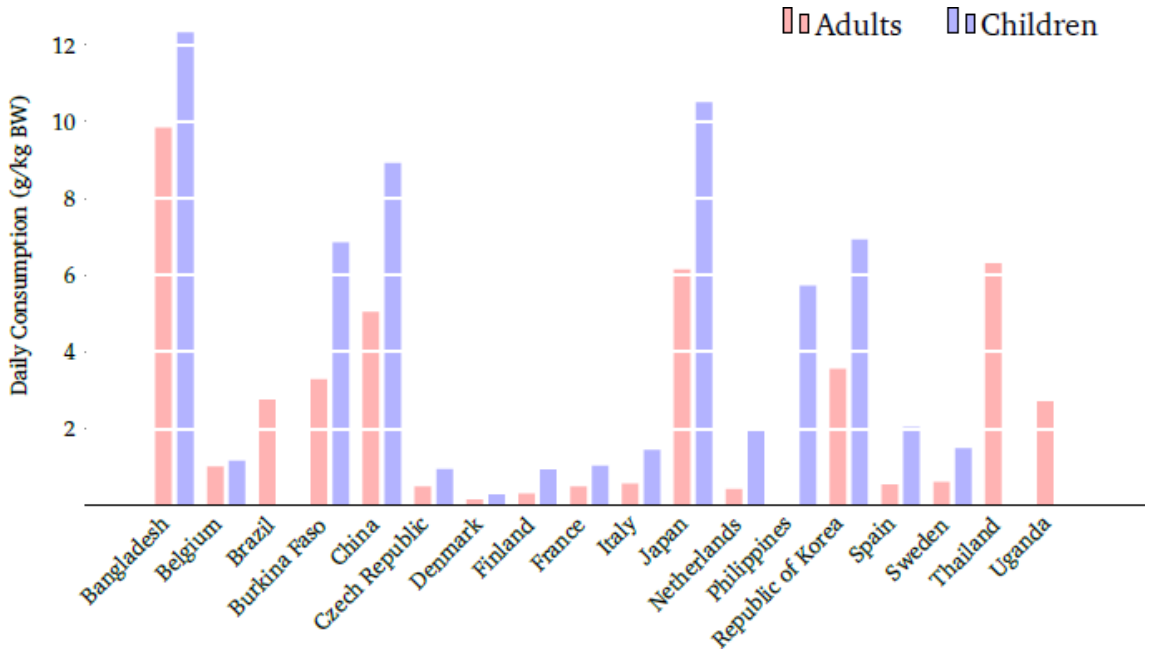
Supplementary Fig. 4 Schematic diagram of the GR2E T-DNA insert indicating the *crtI*, *Zmpsy1*, and *pmi* genes along with their respective regulatory elements. Approximate locations of the seven overlapping polymerase chain reaction (PCR) fragments subjected to nucleotide sequencing are indicated as A through G. The location of two putative ORFs spanning the 5' and 3' junctions of the T-DNA insert and the rice genomic DNA are indicated as ORF-1 and ORF-2, respectively (adapted from GR2E-FFP submitted study reports)



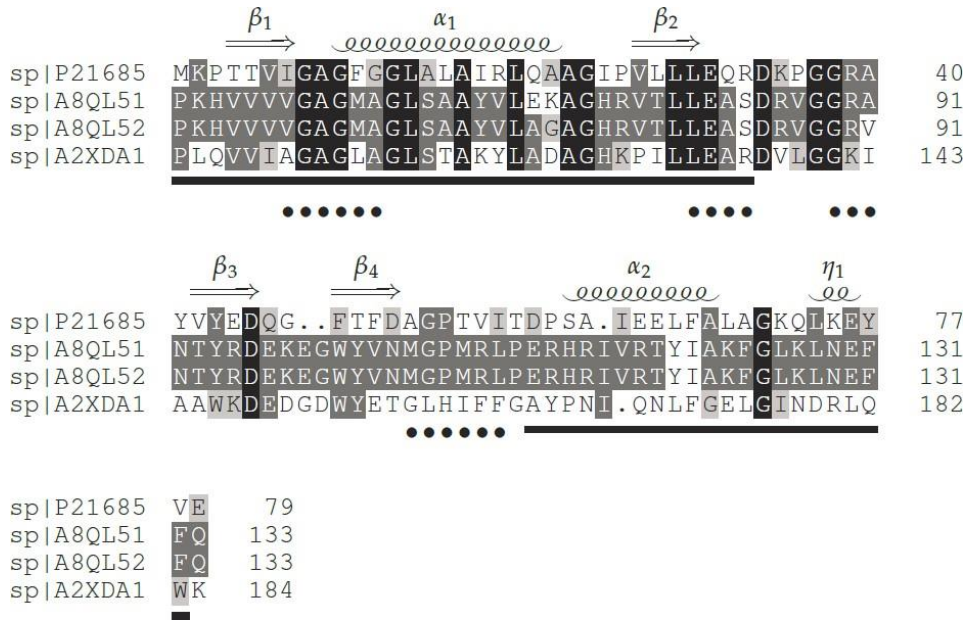
Supplementary Fig. 5 Total protein extracts were prepared from dough (BBCH 75; lane 3), milk (BBCH 85; lane 4), and mature (BBCH 90; lane 5) grain, bran (lane 6), hulls (lane 7), leaf (lane 8), stem (lane 9), and root tissue (lane 10) obtained from Kaybonnet rice containing event GR2E. Similar extracts were prepared from control non-transgenic Kaybonnet dough-stage grain (lane 11). For each analysis, the positive control sample (lane 2) consisted of control Kaybonnet extract spiked with 2.5 ng of purified ZmPSY1 protein (Lot No. M20452-05; panels A and B), 25 ng of purified CRTI protein (Lot No. M20454-02; panels C and D), or 6.25 ng purified PMI protein (Lot No. 21038G; panels E and F). Samples (ca. 40 µg total protein, panels A–D, or 7 µg total protein, panels E and F) were subjected to sodium dodecylsulfate (SDS)-polyacrylamide gel electrophoresis (PAGE). Gels were either stained for total protein with colloidal blue G250 (panels A, C, and E) or subjected to western immunoblot analysis (panels B, D, and F). For ZmPSY1 (panel B) and CRTI (panel D), the blots were labelled with monoclonal anti-ZmPSY1 (ELX1048; 2 µg/ml) and monoclonal anti-CRTI (ELX1043; 2 µg/ml), respectively, followed by alkaline phosphatase (AP)- conjugated goat anti-mouse IgG (1:10000). For PMI (panel F), the blot was labelled with horseradish peroxidase (HRP)-conjugated rabbit anti-PMI (1:100). For all blots, detection of bound antibody-conjugate was via chemiluminescent substrate development. Molecular weight standards are shown in lane 1 (panels A, C, and E) (adapted from GR2E-FFP submitted study reports).



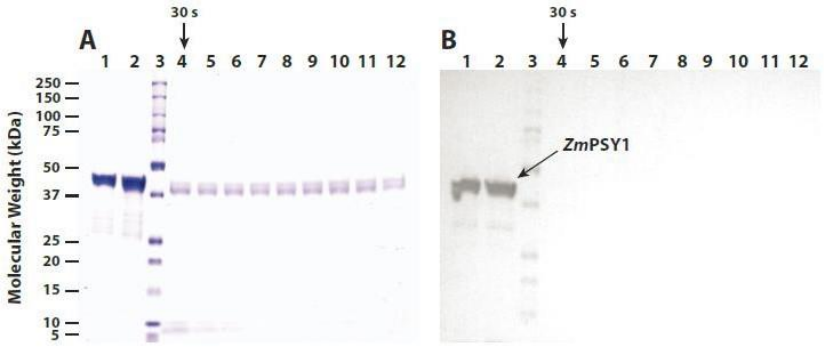
Supplementary Fig. 6 Per capita rice consumption in the highest-consuming countries in Asia compared with the United States. Values from 2011–2015 were derived using the USDA Production, Supply, and Distribution online database (USDA-FAS, 2016) and the FAOSTAT population database, while values after 2015 are projections taken from Wailes and Chavez (2015) (adapted from GR2E-FFP submitted study reports).



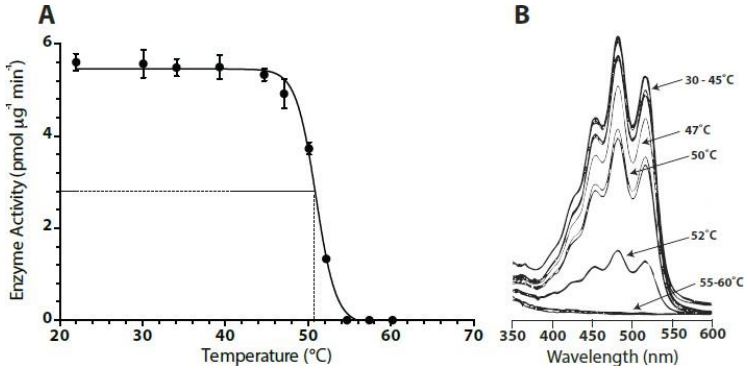
Supplementary Fig. 7 Daily individual rice consumption rates for adults and children in selected countries using data from the FAO/WHO Chronic Individual Food Consumption Database (adapted from GR2E-FFP submitted study reports)



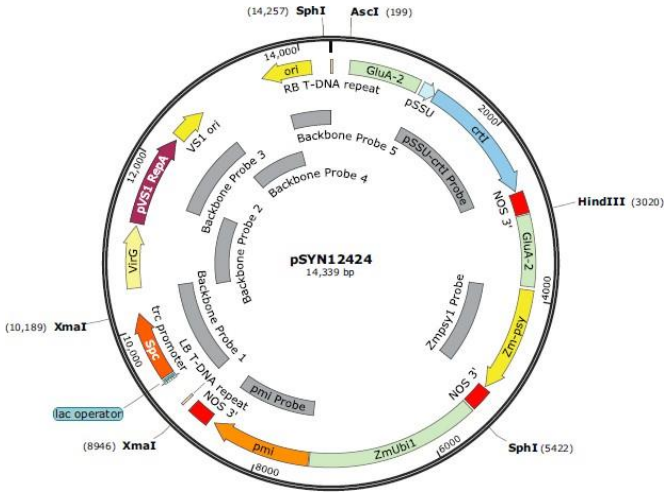
Supplementary Fig. 8 Alignment of the N-terminal region of the *P. ananatis* CRTI protein (spjP21685, residues 1–79) with the corresponding regions of the rice (*O. sativa*) phytoene desaturase (OsPDS, spjA2XDA1) and the two L-amino acid oxidase enzymes identified from the toxin database search. These were from *Bungarus multicinctus* (OXLA_BUNMU, spjA8QL51) and *B. fasciatus* (OXLA_BUNFA, spjA8QL52). The secondary structure elements of CRTI have been indicated above the alignment and the shaded bar underneath the alignment indicates the FAD-binding domains within the N-terminal region, with bullets indicating putative FAD binding residues, as described in Schaub et al. (2012). This figure was generated with TEXshade (Beitz, 2000) (adapted from GR2E-FFP submitted study reports).



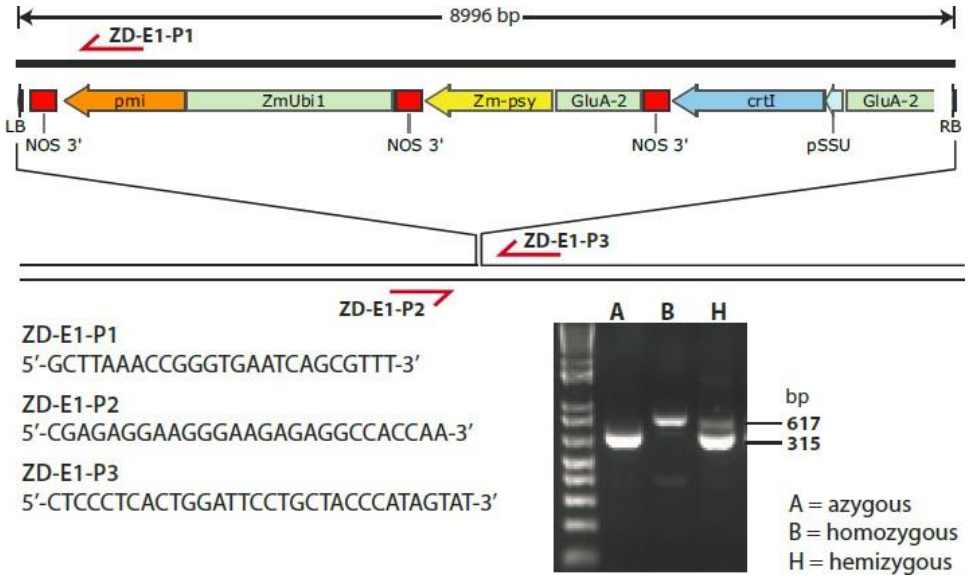
Supplementary Fig. 9 Panels A and B: Samples of *Zm*PSY1 protein purified from recombinant *E. coli* (Lot No. M20452-05) were incubated in the presence of SGF pH 1.2 containing pepsin for 0 min (lane 2) and 0.5, 1, 2, 5, 10, 20, 30 or 60 min at 37°C (lanes 4–11) and then analyzed by SDS-PAGE. Gels were either stained for protein with colloidal blue G250 (panel A) or subjected to western immunoblot analysis (panel B) using rabbit anti-*Zm*PSY1 immunoglobulin and horseradish peroxidase-conjugated goat anti-rabbit IgG followed by precipitating substrate development. Control samples included *Zm*PSY1 protein diluted in gastric control fluid without pepsin (lane 1) and SGF solution containing pepsin (lane 12). Molecular weight standards are shown in lane 3. The apparent molecular weight of *Zm*PSY1 protein corresponds to ca. 42 kDa (adapted from GR2E-FFP submitted study reports).



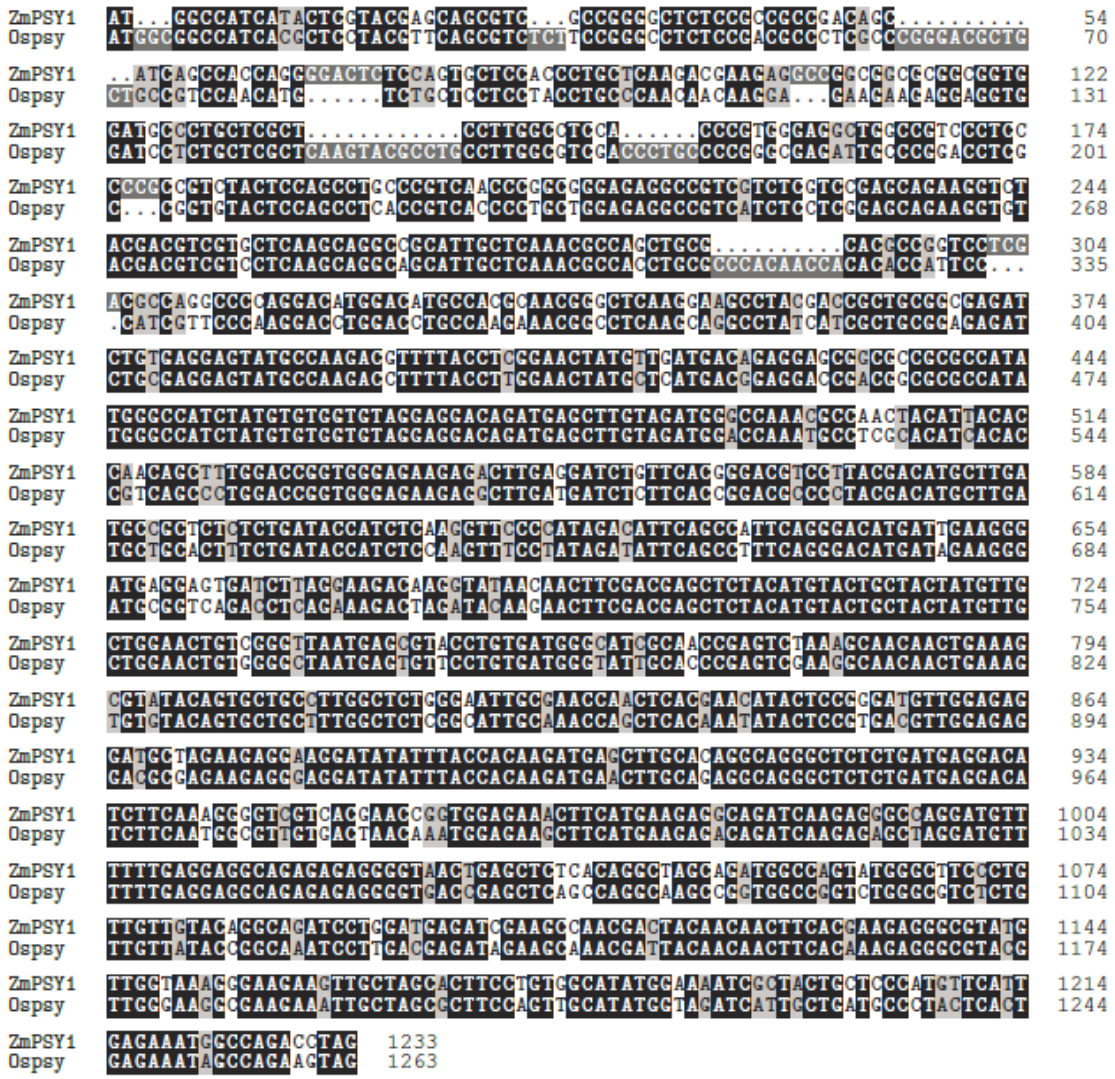
Supplementary Fig. 10 Individual samples of CRTI protein purified from recombinant E. coli (Lot No. M20454-02) were heated for 15 minutes at a designated temperature ranging from 30–60°C. Following this treatment, enzymatic conversion of 15-cis-phytoene to all-trans-lycopene was measured with a spectrophotometric assay. **Panel A** shows enzymatic activity (pmol $\mu g^{-1} min^{-1}$) versus pre-incubation temperature, where the values are means of two technical replicates and error bars represent the standard deviation around the mean value. **Panel B** illustrates the UV-visible spectral scans (350–600 nm) of the chloroform:methanol extracts of CRTI activity assays following different pre-incubation temperature regimes. The maximum absorbance at 483 nm was used to calculate the amount of all-trans-lycopene product (adapted from GR2E-FFP submitted study reports).



Supplementary Fig. 11 Schematic map of plasmid pSYN12424 showing the location of the DNA probes used for Southern hybridization characterization of GR2E rice (adapted from GR2E-FFP submitted study reports).



Supplementary Fig. 12 Genomic DNA isolated from individual plants was analyzed by multiplex PCR amplification employing three primers. Two primers, ZD-E1-P2 and ZD-E1-P3, were complementary to sequences in the left border (LB) and right border (RB) flanking host genomic regions, respectively, and primer ZD-E1-P1 was complementary to pmi gene sequences within the inserted T-DNA. Null segregants lacking the pSYN12424 T-DNA resulted in amplification of a single 315 bp fragment. Homozygous GR2E plants resulted in amplification of a single 617 bp fragment, and hemizygous plants resulted in the amplification of two fragments (315 bp and 617 bp) (adapted from GR2E-FFP submitted study reports).



Supplementary Fig. 13 Alignment of the nucleotide sequences of the *Zea mays psy1* (*Zmpsy1*; GI:1098664) and *Oryza sativa psy* (*Ospsy*; GI:47678186) genes. These two genes share 86.9 percent similar residues and 82.8 percent identical residues. This figure was generated with TEXshade (Beitz, 2000)

Supplementary Table 1 Location and sizes of DNA probes used for Southern hybridization

Probe	Primers (5' to 3')	Position on pSYN12424 (bp to bp)	Length (bp)
Zmpsy1	ATGGCCATCATACTCGTACGA CTAGGTCTGGCCATTTCTCAA	3893-3913 5105-5125	1233
SSU-crtI	ATGGCTTCTATGATATCCTCTTCC TCAAATCAGATCCTCCAGCA	1087-1110 2717-2736	1650
pmi	ATGCAAAAACTCATTAACTCAGTGC TTACAGCTTGTTGTAAACACGC	7429-7453 8583-8604	1176
Backbone 1	CCACCAGCCAGCCAACAG CTTHTTCACAGCGCTTGAGA	9028-9045 10418-10437	1410
Backbone 2	GCCGCAATTCTGACGAACTG CGGTTCATGGATTCGGTTAG	10355-10374 11664-11683	1329
Backbone 3	GTGATCCGCTACGAGCTTCC CTCTCATCAACCGTGGCTCC	11559-11578 12821-12840	1282
Backbone 4	CTACGGCCAGGCAATCTACC TGGAGCGAACGACCTACACC	12681-12700 13751-13770	1090
Backbone 5	ATCTCAGTTCGGTGTAGGTC CTAAGAGAAAAGAGCGTTTATTAG	13741-13760 14316-14339	599

(adapted from GR2E-FFP submitted study reports)

Supplementary Table 2 Correlation of predicted and observed fragment sizes bases on a single site of insertion of the pSYN12424 T-DNA within GR2E rice

		Event GR2E Genomic DNA Fragment Size (bp)		Plasmid pSYN12424 Fragment Size (bp)	
Probe	Enzyme	Predicted	Observed[a]	Predicted	Observed
Zmpsy1	*Asc*I + *Xma*I	8747	8747	-	-
	*Sph*I	>5422	~6900 ~5600[b]	5504	5504
	*Hin*dIII	>6010	~7900 ~4900[b]	-	-
SSU-crtI	*Asc*I + *Xma*I	8747	8747	-	-
	*Sph*I	>5422	~6900	5504	5504
	*Hin*dIII	>3020	~7200	-	-
pmi	*Asc*I + *Xma*I	8747	8747	-	-
	*Sph*I	>3608	~5500	8835	8835
	*Hin*dIII	>6010	~7900	-	-

[a] Observed fragment sizes were approximated from the digoxigenin (DIG)-labelled DNA molecular size marker VII fragments on the Southern blots. Due to the inability to determine exact sizes on the blot, all values are approximations to the nearest 200 bp. The exception to this was when the observed fragment had equal migration as the control plasmid hybridization fragment, in which case the predicted size was assigned (e.g., 5504 bp in the case of *Sph*I digestions probed with DIG-labelled *Zmpsy1* or *SSU-crtI*).

[b] These fragments were due to hybridization with sequences from the endogenous rice *psy* gene, which is ca. 83 percent identical to the *Zmpsy1* gene . (Supplementary Fig.13,) (adapted from GR2E-FFP submitted study reports).

Supplementary Table 3 Primers used for nucleotide sequencing of the inserted DNA within GR2E rice

PCR Fragment	Primer	Primer Sequences (5'-3')	Target Sequence Location (bp to bp)†	Amplified Fragment Size (bp)
A	GR2E_F1 GR2E_R1	TCTGGAGGATCTACTGCTGC CGCTCATGATCAGATTGTCG	1-20 2022-2041	2041
B	GR2E_F2 GR2E_R2	CCCAGTGAGTAGTATTTTCAGTC GT AAGACCGGCAACAGGATTC	1642-1666 4780-4778	3137
C	GR2E_F3 GR2E_R3	TAGAGGACGGTCGCAGGTTC GGCCTCTTCGTCTTGAGCAG	3998-4017 5945-5964	1967
D	GR2E_F4 GR2E_R4	TCCACCTTTCGTGTACCACA AAGACCGGCAACAGGATTC	5662-5681 7149-7168	1506
E	GR2E_F5 GR2E_R5	GGATGAGATCGAAGCCAACG CCACATCATCACAACCAAGC	6952-6971 8820-8969	2018
F	GR2E_F6 GR2E_R6	GGATCTGTATGTGTGTGCCA AAGACCGGCAACAGGATTC	8820-8839 10668-10686	1868
G	GR2E_F7 GR2E_R7	ACCACCATTAGCCAGCAGAG AGGAAGAATGCATCTGGAGC	10390-10409 12753-12772	2383

† Numbering is relative to the sequence of the insert and flanking regions from event GR2E rice. Nucleotides corresponding to the 5' flanking sequence (1-1988), T-DNA insert (1989-10984), and 3' flanking sequence (10985-12772). (adapted from GR2E-FFP submitted study reports)

Supplementary Table 4 Thermocycling parameters

Fragment	Denaturation	Annealing	Extension	Cycles
A,B,G	3 min; 95°C 20 sec; 98°C	15 sec; 55°C	2 min; 72°C 7 min; 72°C	1 35 1†
C	3 min; 95°C 20 sec; 98°C	15 sec; 55°C	2 min; 72°C 7 min; 72°C	1 35 1
D	3 min; 95°C 20 sec; 98°C	15 sec; 55°C	90 sec; 72°C 7 min; 72°C	1 35 1
E	3 min; 95°C 20 sec; 98°C	15 sec; 65°C	2 min; 72°C 7 min; 72°C	1 35 1
F	3 min; 95°C 20 sec; 98°C	15 sec; 60°C	2 min; 72°C 7 min; 72°C	1 35 1

† Following completion of the final extension incubation, samples were held at 4°C. (Adapted from GR2E-FFP submitted study reports)

Supplementary Table 5 Mendelian Inheritance of the pSYN12424 T-DNA insert within multiple generations of GR2E rice

		Expected		Observed			
Generation[a]	Total	Present	Absent	Present	Absent	Chi Square[c]	p-Value
BC4F2*1 (3:1)	64	48	16	44	20	1.333	0.2482
BC4F2*2 (3:1)	51	38.25	12.75	38	13	0.007	0.9356
BC4F2*3(3:1)	63	47.25	15.75	43	20	1.529	0.2162
BC5F1*1 (1:1)	59	29.5	29.5	32	27	0.424	0.5151
BC5F1*2 (1:1)	49	24.5	24.5	25	24	0.020	0.8864
BC5F1*3 (1:1)	127	63.5	63.5	68	59	0.638	0.4245
BC5F2*1 (3:1)	100	75	25	74	26	0.053	0.8174
BC5F2*2(3:1)	99	74.25	24.75	72	27	0.273	0.6015
BC5F2*3(3:1)	100	75	25	70	30	1.333	0.2482

[a] Generation identified with the superscript "*1", "*2", and "*3" designation represented different GR2E breeding lines where the recurring non-transgenic parent was PSB Rc82, BRRI *dhan* 29, and IR64, respectively.
[b]Multiplex PCR-based zygosity testing was performed on each plant from each generation.
[c]The analysis tested the hypothesis that the introduced DNA was segregating in a Mendelian fashion. The critical value to reject the hypothesis at the five percent level is 3.84. Since the Chi-square (χ^2) value was less than 3.84 within each generation, the observed differences were not statistically significant(adapted from GR2E-FFP submitted study reports).

Supplementary Table 6 Location and identity of the ORFs spanning the 5' and 3' T-DNA–genomic DNA junctions in GR2E rice

ID	Nucleotide Location	Strand	Length (amino acids)	Deduced Amino Acid Sequence	No. of Allergen Hits	No. of FASTA Hits[a]
ORF-1	1830-2036	-	68 (7.9 kDa)	MIRLSFPAFSLNYQCL EKKRRFISLPVHLHQF CNCNSGCVYVSIHIHT HTLNFQTILNRLSILSP VLP	None	None
ORF-2	10870-11109	+	79 (8.8 kDa)	MLLDLLALQEIYRCPG GQHGRIRNVLLSCLSV NLFTPQFLGNGRCNQS MTTRLESKQAQILDQF IGGLSSLPLAFLLLQ	None	None

[a]Sequences were queried against a protein toxin database using FASTA with an *E*-value cutoff of 1 x 10^{-5} (adapted from GR2E-FFP submitted study reports).

Supplementary Table 7 Concentrations of *Zm*PSY1, CRTI, and PMI in samples of grain and straw collected from GR2E rice grown at four locations in the Philippines during two growing seasons in 2015-16

	Concentration (ng/g FWT)[†]					
	*Zm*PSY1		CRTI		PMI	
Samples	2015	2016	2015	2016	2015	2016
Grain (milk stage)	156.6 (126.4-180.8)	161.7 (138.5-184.6)	43.6 (23.3-52.2)	36.6 (31.4-40.0)	1548.1 (1097-1798)	1828.4 (1352-2083)
Grain (dough stage)	338.6 (326.3-358.9)	328.9 (308.2-356.4)	61.8 55.3-67.6)	60.3 (53.8-68.0)	1897.9 (1565-2197)	2130.7 (1912-2397)
Grain (mature stage)	226.6 (196.7-244.8)	220.2 (195.7-239.6)	26.1 (23.0-29.7)	22.8 (16.6-27.4)	1369.0 (915-1891)	1195.0 (765-1780)
Straw (mature stage)	<LOQ[a]	<LOQ	<LOQ	<LOQ	466.7 (320-617.5)	498.0 (339-795.5)

[†] Values represent the mean across four locations of three replicate samples collected from each location in the Philippines where event GR2Erice was grown during the rainy season in 2015 and again during the dry season in 2016 (n=12 for each season). The lowest and highest individual means location values for each year are shown in parentheses. Concentrations are uncorrected for extraction effici ency and expressed in ng protein per gram fresh weight tissue (FWT).

[a]LOQ = Limit of quantification, which for the *Zm*PSY ELISA with straw (stem) samples was 66.4 ng/g FWT and for CRTI with straw samples was 6.2 ng/g FWT. (adapted from GR2E-FFP submitted study reports)

Supplementary Table 8 Historic and projected per capita rice consumption in Asia and in the US

	Annual per capita Utilization (kg)[a]					Projected per capita Utilization (kg)[b]		
Country	2011	2012	2013	2014	2015	2016/17	2020/21	2024/25
Bangladesh	223.6	222.2	222.1	221.3	220.5	224.8	227.6	228.5
Cambodia	233.0	239.3	242.1	241.4	231.1	245.4	249.1	252.9
China	103.5	106.2	107.4	107.8	109.0	106.3	104.1	102.1
India	74.8	74.4	77.5	76.7	74.7	78.6	79.4	80.1
indonesia	156.0	153.7	153.2	150.5	148.7	152.7	151.2	149.5
Malaysia	94.8	93.5	94.2	93.6	93.0	92.1	91.7	91.3
Philippines	136.1	133.8	131.7	133.1	131.6	130.0	130.0	130.5
Thailand	155.4	157.8	161.2	169.8	169.2	160.4	157.2	154.2
Vietnam	220.0	242.4	240.8	235.9	234.4	232.0	232.5	227.6
United States	11.2	12.0	12.5	13.0	12.0	13.5	13.5	13.3

[a]Consumption estimates were derived using data from the USDA PSD (Production, Supply and Distribution) online database and the FAOSTAT population database.

[b]Projection excerpted from Table 68 in Wailes and Chavez (2015). (adapted from GR2E-FFP submitted study reports)

Supplementary Table 9 FAO/WHO CIFOCOss per capita daily rice consumption data for selected countries in Asia

Country	Population Subgroup	Number of Consumers Sampled	Mean Daily Consumption (g/kg body weight)[a]
Bangladesh	Adult women	474	9.84
	Children	545	12.32
China	General population	50011	5.04
	Children	2,233	8.92
Japan	General population	2711	6.15
	Children	71	10.5
Philippines	Children	1114	5.72
Republic of Korea	General population	9225	3.56
	Children	648	6.92
Thailand	General population	15695	6.31

[a]Individual consumption data for rice (excluding wild rice: GC 0649) were obtained from the FAO/WHO Chronic Individual Food Consumption Database summary statistics (CIFOCOss) accessed on 16 May 2016. Available at http://www.who.int/foodsafety/databases/en/. (adapted from GR2E-FFP submitted study reports)

Supplementary Method 1 Nucleotide Sequence Analysis of GR2E Rice

The nucleotide sequence of the plasmid pSYN12424 T-DNA, together with preliminary sequence information from the 5' and 3' flanking genomic DNA (Syngenta), was used to design seven sets of oligonucleotide primers that were used to amplify the insert and flanking regions from GR2E rice genomic DNA as seven individual overlapping fragments (Supplementary Table 3). PCR amplifications were carried out using the KAPA HiFi HotStart PCR Kit (Kapa Biosystems, Wilmington, MA, USA) and a G-Storm GS1 thermocycler using the parameters in Table S4.

The PCR amplification fragments were cloned into pCR4-TOPO vector via blunt-end ligation and transformed into *E. coli* strain DH5. Three colonies from each transformation were randomly selected and separately transferred into liquid culture and grown overnight at 37°C with shaking. Plasmid DNA prepared from each culture was submitted to Macrogen Incorporated (Seoul, Korea) for dye-terminator nucleotide sequencing using the ABI3730XL analyzer with ABI BigDye 3.1 terminator chemistry. Sequence analysis was performed using the Phred, Phrap, and Consed packages (University of Washington) to an error rate of less than 1 in 10,000 bases (Ewing and Green, 1998). Nucleotide sequences from each individual clone were aligned using AlignX™ (Vector NTI Advance™ v. 11.0) to obtain the final consensus sequence for each PCR amplification fragment.

Supplementary Method 2 Sequence Similarity Searches to Known Allergens using ORF-1 and ORF-2 Sequences as Queries

The deduced amino sequences of the two junction-spanning open reading frames (ORFs) identified in GR2E rice were used as query sequences to search for significant sequence similarities with 2129 known and putative allergen and celiac protein sequences residing in the FARRP19 dataset at the University of Nebraska. Search routines included a search for full-length alignments using FASTA to identify any significant sequence alignments (E-value threshold $1X10^{-5}$) that would be indicative of possible cross-reactivity, as well as the sliding 80-mer search to look for >35 percent sequence identity over all possible 80 amino acid segments and a search for exact matches of 8 contiguous amino acids.

Supplementary Reference

Beitz, E. TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics (Oxford, England)*, 16(2):135–139 (2000).