Toxicology Reports 2 (2015) 1233–1245

Contents lists available at ScienceDirect

Toxicology Reports

journal homepage: www.elsevier.com/locate/toxrep

Genome-wide analysis of potential cross-reactive endogenous allergens in rice (*Oryza sativa* L.)☆

Fang Chao Zhu[a,b,1], Rui Zong Jia[a,*,1], Lin Xu[a], Hua Kong[a], Yun Ling Guo[a], Qi Xing Huang[a], Yun Judy Zhu[a,c], An Ping Guo[a,*]

[a] *State Key Biotechnology Laboratory for Tropical Crops, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China*
[b] *College of Agriculture, Hainan University, Haikou, Hainan 570228, China*
[c] *Hawaii Agriculture Research Center, Kunia, HI 96759, USA*

ARTICLE INFO

Article history:
Received 22 April 2015
Received in revised form 4 July 2015
Accepted 20 July 2015
Available online 29 July 2015

Keywords:
Rice
Allergen
Conserved domain
Phylogenetic
Cross-reacting endogenous allergens

ABSTRACT

The proteins in the food are the source of common allergic components to certain patients. Current lists of plant endogenous allergens were based on the medical/clinical reports as well as laboratory results. Plant genome sequences made it possible to predict and characterize the genome-wide of putative endogenous allergens in rice (*Oryza sativa* L.). In this work, we identified and characterized 122 candidate rice allergens including the 22 allergens in present databases. Conserved domain analysis also revealed 37 domains among rice allergens including one novel domain (histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, PF13589) adding to the allergen protein database. Phylogenetic analysis of the allergens revealed the diversity among the Prolamin superfamily and DnaK protein family, respectively. Additionally, some allergens proteins clustered on the rice chromosome might suggest the molecular function during the evolution.

1. Introduction

Food allergy is typically defined as an immunoglobulin E (IgE) or non-IgE mediated immune response to food proteins, which can induce symptoms such as itching, wheezing, vomiting, nausea, urticaria, diarrhea, oral allergy syndrome, abdominal pain and even systematic anaphylaxis. With growing evidence of an increase in prevalence, food endogenous allergens affect nearly 5% adults and 8% of children [1]. Rice is one of the most important crops cultivated worldwide. Despite wide consumption, rice is commonly regarded as hypoallergenic, and recommended as diet substitute for some cereal sensitive patients. After the first allergic reaction to rice reported in 1979, it has attracted increasingly public attentions [2]. Hereafter, a number of clinical cases on rice allergy that triggered either by contacting with raw rice, inhaling of rice powders or vapors, or by ingesting of rice have been reported [3–6]. In Japan, the prevalence of IgE-mediated hypersensitivity to rice is higher rate in atopic subjects, while it is much lower in Europe and America [7]. An Indian group reported that IgE-mediated rice allergy affects about 0.8% of asthma and rhinitis cases [8].

So far, several putative allergenic components in rice have been described. Among the reported rice allergens, β-expansin (35 kDa, Ory s 1) and profilin A (14 kDa, Ory s 12) listed the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-Committee (www.allergen.org). Other suspected rice allergens have already been reported, but further tests need to be conducted for validation. Six of the most abundant pollen-specific candidate transcripts: Ory s 2, calcium-binding protein/polcalcin (Ory s 7), Ory s 11, Ory s 23, glycosyl hydrolase family 28 (polygalacturonase) and FAD binding proteins, were suggested to be listed as putative allergens [9]. ELISA and immunoblotting analysis results confirmed the allergenicity of the Ory s 1 and extensin [10]. A group of 14–16 kDa proteins contains 11 isoallergens belongs to α-amylase/trypsin inhibitor family [7]. The 26 and 33 kDa seed proteins were characterized as α-globulin and glyoxalase I respectively [11]. The cross-activities between rice lipid transfer protein (LTP, 14 kDa, Ory s 14) and peach/apple LTPs also have been demonstrated [12]. An abundant 56 kDa protein involved in rice allergy was identified as granule-bound starch synthase [13]. Employing proteomic analysis and immunoblotting, two globulin-like proteins (a 52 kDa protein and a

☆ Genome-wide screening finds 122 rice allergens, including the 22 allergens in database, a novel domain also suggested to add to the current allergen database.

* Corresponding authors.
E-mail addresses: jiaruizong@itbb.org.cn (R.Z. Jia), gap211@126.com (A.P. Guo).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.toxrep.2015.07.017

63 kDa protein, both homologous to Cupin superfamily) were identified as novel IgE-binding proteins [14]. Golias et al. [15] identified six new thermostable putative rice allergens: glutelin C precursor, granule-bound starch synthase 1 protein, disulfide isomerase-like 1-1 protein, hypothetical protein OsI_13867, putative acid phosphatase precursor 1, and a protein encoded by locus Os02g0453600.

Although considerable research has been devoted to epidemiology, pathogenesis, and therapy of food allergy over last few decades, the specificity of diagnosis indicating definite allergen according to certain symptoms is still difficult. On one hand, a number of novel allergens have not been discovered or fully certified; on the other hand, closely related allergenic proteins from entirely irrelevant resources can induce IgE-binding cross reactivity due to similarities in overall sequence and structure [16]. Research indicated that 80% of patients with food and pollen allergies have increased IgE antibodies against rice proteins [17]. In addition, ethnic backgrounds, environmental factors and dietary habits also partially account for hypersensitivity. So, it is important to distinguish allergens from non-allergenic proteins, and to predict the potential IgE-mediated cross reactivity. It was reported that an alignment between a query sequence and an allergen having more than 70% identity throughout the length of protein commonly indicated a cross-activity, and 50–70% identity posed a moderate risk of cross reactivity [18]. Another indicator, the expectation value (*E*-value), reflects the degree of similarity and the relationship in evolutionary terms between the query sequence and known allergen. According to the website of Food Allergy Research and Resource Program (FARRP, www.allergenonline.org), it recommended that the *E*-value cutoff (1E-30) was used to judge the candidate protein likely to be allergic cross-reactive.

In this study, we analyzed the rice genome-wide protein sequences with in-house designed pipeline by using WHO/IUIS database and FARRP database to identify putative cross-reactive allergens proteins. By characterizing these conservative protein sequences, we investigated the potential biological functions, the distributions across the chromosomes and the genetic divergence roles of the rice cross-reactive allergens during the evolution.

2. Materials and methods

2.1. Rice protein sequences resources

The rice genome (*Oryza sativa* L. ssp. *japonica* cv. Nipponbare) release 7.0 was publically available on Phytozome v10 (www.phytozome.jgi.doe.gov). The rice genome is 372 Mb genome in 12 chromosomes, 55,986 loci containing protein-coding transcripts including transposable element (TE) genes and 66,338 protein-coding transcripts [19].

2.2. Allergen protein database

The Allergen Nomenclature Database, maintained by the WHO/IUIS Allergen Nomenclature Sub-Committee, contains 797 approved and officially recognized allergens (www.allergen.org). The Food Allergy Research and Resource Program (FARRP) Database provides access to a peer reviewed allergen list of 1706 sequences entries (v 14, released on January 20, 2014) (www.allergenonline.org). Combining the two databases, we retrieved 2194 allergens and fragments by removing the duplicate-entries. The amino acid sequences and other molecular information were obtained via NCBI (www.ncbi.nlm.nih.gov) and UniProt (www.uniprot.org). Among the listed allergens, 22 entries were rice endogenous allergens.

2.3. Identification of potential cross-reacting allergen proteins

The potential rice allergens were identified by local BLAST-P (packaged in BioEdit software, v 7.2.5) performing full-length sequences alignments for similarity to the known allergens, by setting values on BLAST-P: a matrix of BLOSUM62 and an *E*-value of 1 [20]. First, rice genome was set as query to BLAST-P with allergens from the Allergen Nomenclature and FARRP databases, and then the allergens from the Allergen Nomenclature and FARRP databases were set as query to BLAST-P with rice genome. After two rounds BLAST-P, the retrieved candidate proteins were further finalized by two comprehensive threshold values: Identity ≥70% [18] and *E*-value ≤1e-30 (www.allergenonline.org).

2.4. Motif-based potential cross-reacting allergen sequences analysis

The retrieved candidate proteins were submitted to the MEME online server (Multiple Em for Motif Elicitation, v4.9.1, www.meme.nbcr.net) with the aim of discovering novel motifs [21], and further confirmed these sequences by using Pfam-A family database (v27.0, www.pfam.xfam.org) using Hidden Markov Models method [22]. The parameters of MEME: the occurrence of a single motif, the minimum motif width and the maximum motif width were all set to the defaults. The maximum number of different motifs found within the sequences was set to 6.

2.5. Gene ontology (GO) analysis

GO analysis was used to determine the biologic functions most frequently found among allergens. GO accession numbers of candidate proteins were submitted to Web Gene Ontology Annotation Plot (WEGO, www.wego.genomics.org.cn) for visualizing and plotting GO annotation results according previous work [23]. The biochemical functions of allergens most frequently found were limited to hydrolysis of proteins, polysaccharides, and lipids; binding of metal ions and lipids; storage; and cytoskeleton association [24]

2.6. Phylogenetic analysis of rice allergen proteins

The retrieved candidate proteins were aligned by using Clustal W program. Pairwise distance of aligned allergen proteins was calculated by using PAUP (v4.0b10, Sinauer Associates, Inc.) to construct Neighbor-Joining tree by setting bootstrap at 1000. The TreeView (Win32, v1.6.6) was used to view the phylogenetic tree.

2.7. Analysis of putative rice allergens across the chromosomes

The information of the location of a retrieved candidate protein on a chromosome was obtained from Phytozome v10. Physical map of potential rice allergens across the chromosomes was constructed by MapChart 2.2. Further, all candidate proteins were classified into groups according to their functions and structures and were showed in the map.

2.8. Quantitative PCR validation of the putative allergen proteins gene expression

Rice stems and leaves of post-tillering phase were collected and fine-ground in liquid nitrogen. Total DNA was extracted by using Gentra Puregen DNA Extraction Kit (Qiagen). The concentration and integrity of genome DNA was measured with NanoDrop™ Spectrophotometer (Thermo Fisher). Total RNA was extracted by using RNeasy Plant Mini Kit (Qiagen). Primers were designed by

Table 1
The list of rice genome-wide putative allergens.

No.	Gene name [a]	E-value	ID [b]	GI#	Species	AN [c]	Protein name	Featured conserved domains [d]
Allgn1	LOC_Os01g04360.1	3.0E-59	75.3	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn2	LOC_Os01g04370.1	4.0E-60	75.3	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn3	LOC_Os01g04380.1	2.0E-60	76.0	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn4	LOC_Os01g05490.1	7.0E-125	86.6	11124572	*T. aestivum*	ND	Triticum triosephosphate isomerase	
Allgn5	LOC_Os01g05490.2	5.0E-91	87.0	11124572	*T. aestivum*	ND	Triticum triosephosphate isomerase	
Allgn6	LOC_Os01g41710.1	3.0E-133	85.8	1769849	*A. graveolens*	Api g 3	Chlorophyll a-b binding protein	
Allgn7	LOC_Os01g52240.1	6.0E-132	85.9	1769849	*A.graveolens*	Api g 3	Chlorophyll a-b binding protein	
Allgn8	LOC_Os01g60740.1	1.0E-38	82.0	283476400	*T. aestivum*	Tri a 14	Nonspecific lipid transfer protein 1	
Allgn9	LOC_Os01g62290.1	0	76.8	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn10	LOC_Os01g62290.2	0	74.0	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn11	LOC_Os01g62420.1	8.0E-116	80.1	11124572	*T. aestivum*	ND	Triticum triosephosphate isomerase	
Allgn12	LOC_Os01g62420.2	1.0E-44	78.3	11124572	*T. aestivum*	ND	Triticum triosephosphate isomerase	
Allgn13	LOC_Os01g62420.3	6.0E-84	80.3	11124572	*T. aestivum*	ND	Triticum triosephosphate isomerase	
Allgn14	LOC_Os01g62420.4	8.0E-78	82.5	11124572	*T. aestivum*	ND	Triticum triosephosphate isomerase	
Allgn15	LOC_Os02g02410.1	0	88.5	10944737	*C. avellana*	Cor a 10	Luminal binding protein	
Allgn16	LOC_Os02g02890.1	2.0E-80	81.8	373939374	*D. carota*	ND	Daucus cyclophilin	
Allgn17	LOC_Os02g07490.1	3.0E-143	75.0	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn18	**LOC_Os02g17920.1**	2.0E-119	70.0	84029333	*O. sativa*	ND	Oryza glyoxalase I	
Allgn19	LOC_Os02g38920.1	2.0E-163	84.5	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn20	**LOC_Os03g01610.1**	2.0E-159	100	109913547	*O. sativa*	Ory s 1	Beta-expansin	
Allgn21	**LOC_Os03g01630.1**	3.0E-123	81.7	109913547	*O. sativa*	Ory s 1	Beta-expansin	

Table 1 (*Continued*)

Allgn22	**LOC_Os03g01640.1**	1.0E-161	100	8118439	*O. sativa*	ND	Oryza Ory s 1	
Allgn23	**LOC_Os03g01650.1**	2.0E-159	100	109913547	*O. sativa*	Ory s 1	Beta-expansin	
Allgn24	LOC_Os03g14450.1	0	88.5	9581744	*H. brasiliensis*	Hev b 9	Enolase	
Allgn25	LOC_Os03g14450.2	0	89.2	9581744	*H. brasiliensis*	Hev b 9	Enolase	
Allgn26	LOC_Os03g15960.1	6.0E-62	73.9	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn27	LOC_Os03g16020.1	2.0E-61	77.3	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn28	LOC_Os03g16030.1	4.0E-62	74.5	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn29	LOC_Os03g16040.1	2.0E-59	71.0	46359518	*C. sativa*	Cas s 9	Cytosolic class I small heat shock protein	
Allgn30	LOC_Os03g16860.1	0	75.5	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn31	LOC_Os03g16860.2	0	73.1	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn32	LOC_Os03g16920.1	0	75.3	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn33	LOC_Os03g18454.2	6.0E-77	73.0	283480515	*T. aestivum*	Tri a 27	Thiol reductase	
Allgn34	LOC_Os03g22810.1	1.0E-74	84.2	160962549	*O. europaea*	ND	Olea Ole e 5	
Allgn35	LOC_Os03g25350.1	6.0E-37	74.4	66840998	*T. aestivum*	ND	Triticum 5a2 protein	
Allgn36	LOC_Os03g30470.1	7.0E-139	70.6	17932710	*M. acuminata*	Mus a 2	Class 1 chitinase	
Allgn37	LOC_Os03g39610.1	6.0E-117	76.2	1769849	*A. graveolens*	Api g 3	Chlorophyll a-b binding protein	
Allgn38	LOC_Os03g39610.2	7.0E-112	86.6	1769849	*A. graveolens*	Api g 3	Chlorophyll a-b binding protein	
Allgn39	LOC_Os03g41419.1	3.0E-102	72.9	5734506	*T. aestivum*	Tri a 33	Serpin	
Allgn40	LOC_Os03g45960.1	6.0E-88	70.7	88191901	*M. acuminata*	Mus a 4	Thaumatin-like protein	
Allgn41	LOC_Os03g46070.1	6.0E-93	74.5	88191901	*M. acuminata*	Mus a 4	Thaumatin-like protein	
Allgn42	LOC_Os03g50250.1	0	71.7	10944737	*C. avellana*	Cor a 10	Luminal binding protein	
Allgn43	LOC_Os03g51600.1	0	71.2	685432814	*D. farinae*	Der f 33	Alpha-tubulin	
Allgn44	LOC_Os03g51600.2	0	71.2	685432814	*D. farinae*	Der f 33	Alpha-tubulin	

Table 1 (*Continued*)

Allgn45	LOC_Os03g60620.1	0	75.5	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70
Allgn46	LOC_Os04g32680.1	9.0E-62	70.6	1588669	*Z. mays*	ND	Zea pollen specific protein
Allgn47	LOC_Os04g40950.1	7.0E-174	89.0	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase
Allgn48	LOC_Os04g40950.2	1.0E-150	88.1	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase
Allgn49	LOC_Os04g40950.3	4.0E-167	89.4	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase
Allgn50	**LOC_Os05g14194.1**	4.0E-117	71.1	84029333	*O. sativa*	ND	Oryza glyoxalase I
Allgn51	**LOC_Os05g14194.2**	6.0E-75	72.9	84029333	*O. sativa*	ND	Oryza glyoxalase I
Allgn52	LOC_Os05g18604.1	0	77.2	125987805	*T. aestivum*	ND	Triticum serine carboxypeptidase II
Allgn53	LOC_Os05g18604.2	0	75.6	125987805	*T. aestivum*	ND	Triticum serine carboxypeptidase II
Allgn54	LOC_Os05g18604.3	0	76.0	125987805	*T. aestivum*	ND	Triticum serine carboxypeptidase II
Allgn55	LOC_Os05g18604.4	2.0E-163	74.8	125987805	*T. aestivum*	ND	Triticum serine carboxypeptidase II
Allgn56	LOC_Os05g18604.5	4.0E-75	82.3	125987805	*T. aestivum*	ND	Triticum serine carboxypeptidase II
Allgn57	LOC_Os05g18604.6	3.0E-56	81.0	125987805	*T. aestivum*	ND	Triticum serine carboxypeptidase II
Allgn58	LOC_Os05g25850.1	2.0E-102	75.8	149786150	*P. vera*	Pis v 4	Superoxide dismutase [Mn]
Allgn59	LOC_Os05g25850.2	1.0E-60	75.5	5777414	*H. brasiliensis*	Hev b 10	Superoxide dismutase [Mn]
Allgn60	LOC_Os05g35400.1	0	73.9	10944737	*C. avellana*	Cor a 10	Luminal binding protein
Allgn61	LOC_Os05g37330.1	9.0E-41	72.8	111013714	*P. dulcis*	Pru du 5	60s acidic ribosomal protein P2
Allgn62	LOC_Os05g38530.1	0	74.7	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70
Allgn63	LOC_Os06g04510.1	0	88.8	9581744	*H. brasiliensis*	Hev b 9	Enolase
Allgn64	LOC_Os06g04510.2	0	89.8	9581744	*H. brasiliensis*	Hev b 9	Enolase
Allgn65	LOC_Os06g05880.1	5.0E-67	86.9	207366248	*T. aestivum*	Tri a 12	Profilin
Allgn66	LOC_Os06g35300.1	2.0E-138	75.3	674275739	*S. halepense*	Sor h 13	Exopolygalacturonase 28
Allgn67	LOC_Os06g35320.1	0	75.6	674275739	*S. halepense*	Sor h 13	Exopolygalacturonase 28
Allgn68	LOC_Os06g35370.1	0	75.6	674275739	*S. halepense*	Sor h 13	Exopolygalacturonase 28

Table 1 (*Continued*)

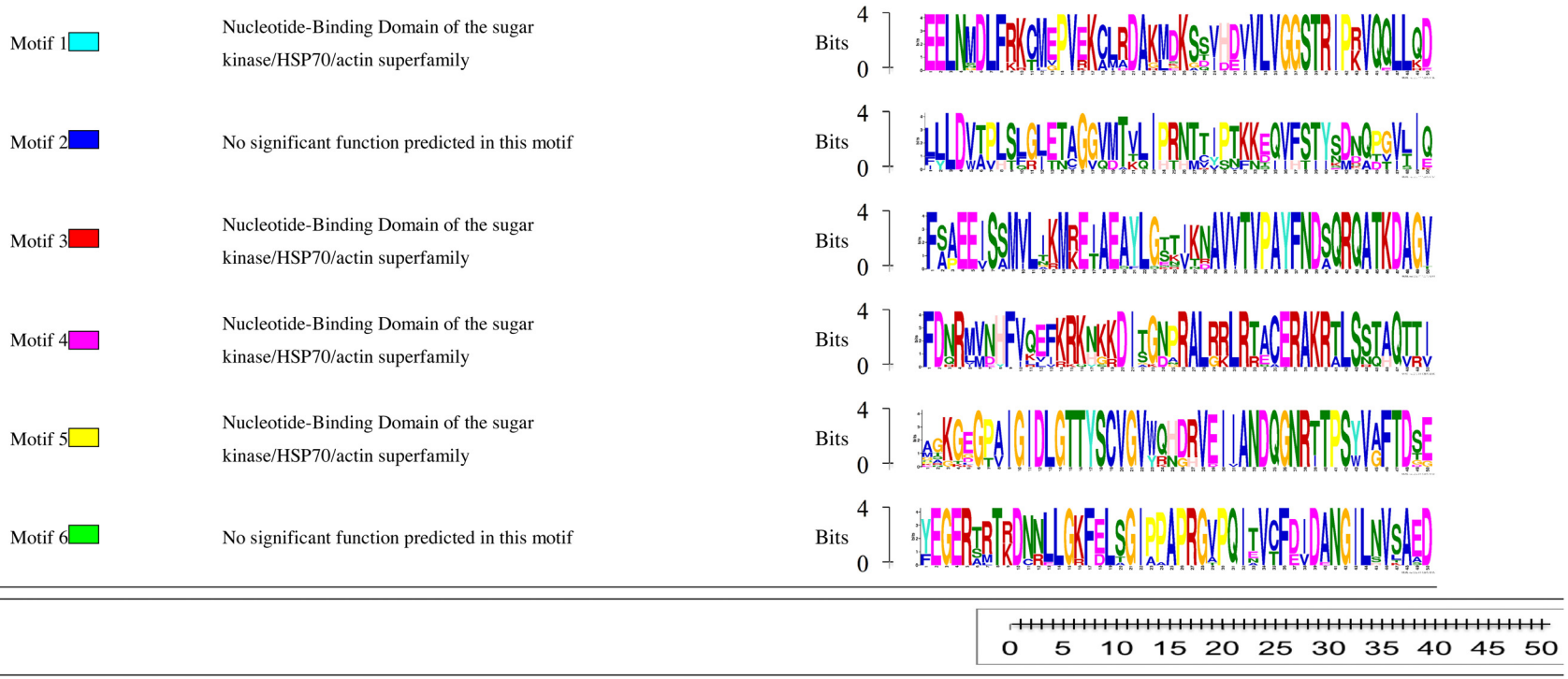
Allgn69	LOC_Os06g36240.1	8.0E-62	77.7	23452313	*P. pratense*	Phl p 11	Ole e 1-related protein	
Allgn70	LOC_Os06g45590.1	4.0E-140	73.2	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn71	LOC_Os06g49480.1	7.0E-67	71.7	373939374	*D. carota*	ND	Daucus cyclophilin	
Allgn72	LOC_Os06g49480.2	7.0E-67	71.7	373939374	*D. carota*	ND	Daucus cyclophilin	
Allgn73	LOC_Os06g51050.1	2.0E-141	73.9	17932710	*M. acuminata*	Mus a 2	Class 1 chitinase	
Allgn74	LOC_Os06g51060.1	1.0E-147	74.6	17932710	*M. acuminata*	Mus a 2	Class 1 chitinase	
Allgn75	**LOC_Os07g11330.1**	1.0E-93	100	23616947	*O. sativa*	ND	Oryza trypsin alpha-amylase inhibitor	
Allgn76	**LOC_Os07g11360.1**	1.0E-96	100	114152864	*O. sativa*	ND	Oryza trypsin alpha-amylase inhibitor	
Allgn77	**LOC_Os07g11380.1**	3.0E-98	100	114152865	*O. sativa*	ND	Oryza trypsin alpha-amylase inhibitor	
Allgn78	**LOC_Os07g11380.2**	2.0E-57	89.3	114152865	*O. sativa*	ND	Oryza trypsin alpha-amylase inhibitor	
Allgn79	**LOC_Os07g11410.1**	9.0E-92	100	23616954	*O. sativa*	ND	Oryza trypsin alpha-amylase inhibitor	
Allgn80	**LOC_Os07g11510.1**	1.0E-93	100	23495787	*O. sativa*	ND	Oryza trypsin alpha-amylase inhibitor	
Allgn81	LOC_Os07g38730.1	0	72.0	685432814	*D. farinae*	Der f 33	Alpha-tubulin	
Allgn82	LOC_Os07g44430.1	1.0E-109	87.0	34539782	*T. aestivum*	Tri a 32	1-cys-Peroxiredoxin	
Allgn83	LOC_Os07g44440.1	7.0E-98	76.3	190684059	*T. aestivum*	ND	Triticum Tri a 32 peroxiredoxin	
Allgn84	LOC_Os07g46990.1	2.0E-75	84.9	39840779	*O. europaea*	Ole e 5	Superoxide dismutase [Cu-Zn]	
Allgn85	LOC_Os07g46990.2	2.0E-75	84.9	39840779	*O. europaea*	Ole e 5	Superoxide dismutase [Cu-Zn]	
Allgn86	LOC_Os08g03290.1	0	92.6	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn87	LOC_Os08g03290.2	1.0E-158	92.2	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn88	LOC_Os08g03290.3	1.0E-156	92.1	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn89	LOC_Os08g03290.4	4.0E-116	92.6	253783729	*T. aestivum*	Tri a 34	Glyceraldehyde-3-phosphate-dehydrogenase	
Allgn90	**LOC_Os08g09250.1**	1.0E-163	100	84029333	*O. sativa*	ND	Oryza glyoxalase I	
Allgn91	**LOC_Os08g09250.2**	2.0E-170	100	84029333	*O. sativa*	ND	Oryza glyoxalase I	
Allgn92	**LOC_Os08g09250.3**	2.0E-101	100	84029333	*O. sativa*	ND	Oryza glyoxalase I	

Table 1 (*Continued*)

Allgn93	LOC_Os08g09770.1	0	73.9	10944737	*C. avellana*	Cor a 10	Luminal binding protein	
Allgn94	LOC_Os08g39140.4	1.0E-91	72.7	1930153	*A. fumigatus*	Asp f 12	Heat shock protein P90	
Allgn95	**LOC_Os08g44660.1**	1.0E-40	98.8	45736119	*O. sativa*	ND	Polcalcin Phl p 7	
Allgn96	LOC_Os09g17740.1	2.0E-131	85.1	1769849	*A. graveolens*	Api g 3	Chlorophyll a-b binding protein	
Allgn97	LOC_Os09g39780.1	2.0E-76	76.8	373939374	*D. carota*	ND	Daucus cyclophilin	
Allgn98	LOC_Os09g39780.2	2.0E-76	76.8	373939374	*D. carota*	ND	Daucus cyclophilin	
Allgn99	LOC_Os10g08550.1	0	88.3	14423687	*H. brasiliensis*	ND	Hevea Hev b 9	
Allgn100	LOC_Os10g08550.3	0	87.9	14423687	*H. brasiliensis*	ND	Hevea Hev b 9	
Allgn101	LOC_Os10g08550.5	0	89.1	14423687	*H. brasiliensis*	ND	Hevea Hev b 9	
Allgn102	**LOC_Os10g17660.1**	9.0E-75	100	11141757	*O. sativa*	Ory s 12	Profilin A	
Allgn103	**LOC_Os10g17680.1**	9.0E-75	100	11141757	*O. sativa*	Ory s 12	Profilin A	
Allgn104	LOC_Os10g40090.1	6.0E-124	73.2	28630919	*Z. mays*	ND	Zea m 1 beta-expansin	
Allgn105	LOC_Os11g02369.1	5.0E-44	74.2	128388	*Z. mays*	Zea m 14	Nonspecific lipid-transfer protein	
Allgn106	LOC_Os11g02389.1	3.0E-42	70.8	128388	*Z. mays*	Zea m 14	Nonspecific lipid-transfer protein	
Allgn107	LOC_Os11g14220.1	0	70.7	685432814	*D. farinae*	Der f 33	Alpha-tubulin	
Allgn108	LOC_Os11g24070.1	2.0E-44	72.3	128388	*Z. mays*	Zea m 14	Nonspecific lipid-transfer protein	
Allgn109	LOC_Os11g37950.1	2.0E-49	70.2	2832430	*H. brasiliensis*	Hev b 6	Hevein precursor	
Allgn110	LOC_Os11g47760.1	0	75.3	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn111	LOC_Os11g47760.2	0	73.1	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn112	LOC_Os11g47760.3	0	77.8	685432788	*D. farinae*	Der f 28	Heat shock protein P70	
Allgn113	LOC_Os11g47760.4	0	75.3	1498496	*P. citrinum*	Pen c 19	Heat shock protein P70	
Allgn114	LOC_Os11g47760.5	0	77.8	685432788	*D. farinae*	Der f 28	Heat shock protein P70	
Allgn115	LOC_Os11g47760.6	1.0E-174	80.1	442565876	*D. farinae*	ND	Heat shock protein P70	
Allgn116	LOC_Os12g02310.1	2.0E-43	73.3	128388	*Z. mays*	Zea m 14	Nonspecific lipid-transfer protein	

Table 1 (*Continued*)

Allgn117	LOC_Os12g02310.2	7.0E-43	73.7	128388	*Z. mays*	Zea m 14 Nonspecific lipid-transfer protein	
Allgn118	LOC_Os12g02320.1	2.0E-42	70.8	128388	*Z. mays*	Zea m 14 Nonspecific lipid-transfer protein	
Allgn119	LOC_Os12g42876.1	0	88.1	225810599	*S. kali*	ND	Salsola Sal k 3 pollen allergen
Allgn120	LOC_Os12g42884.1	0	87.7	225810599	*S. kali*	ND	Salsola Sal k 3 pollen allergen
Allgn121	LOC_Os12g42884.2	0	87.7	225810599	*S. kali*	ND	Salsola Sal k 3 pollen allergen
Allgn122	LOC_Os12g42884.3	0	87.7	225810599	*S. kali*	ND	Salsola Sal k 3 pollen allergen



[a] entries in bold indicates the rice allergens that recorded by allergen databases.

[b] ID indicates identity in percentage (%) of the most similar allergen database entries aligned with BLAST-P.

[c] AN indicates IUIS Allergen Nomenclature designation described by allergen database, where ND means official nomenclature of allergen remains unassigned.

[d] presents the de-nove conserved domains caculated by MEME program (v 4.9.1, www.meme.nbcr.net).

Full-length sequences alignments were performed between rice genome and the allergens from Allergen Nomenclature and FARRP databases. 122 putative rice allergens were finally identified, their most similar allergen databases entries and predicted conserved domains were listed.

[a]Entries in bold indicates the rice allergens that recorded by allergen databases.

[b]ID indicates identity in percentage (%) of the most similar allergen database entries aligned with BLAST-P.

[c]AN indicates IUIS Allergen Nomenclature designation described by allergen database, where ND means official nomenclature of allergen remains unassigned.

[d]Presents the de-nove conserved domains caculated by MEME program (v 4.9.1, www.meme.nbcr.net).

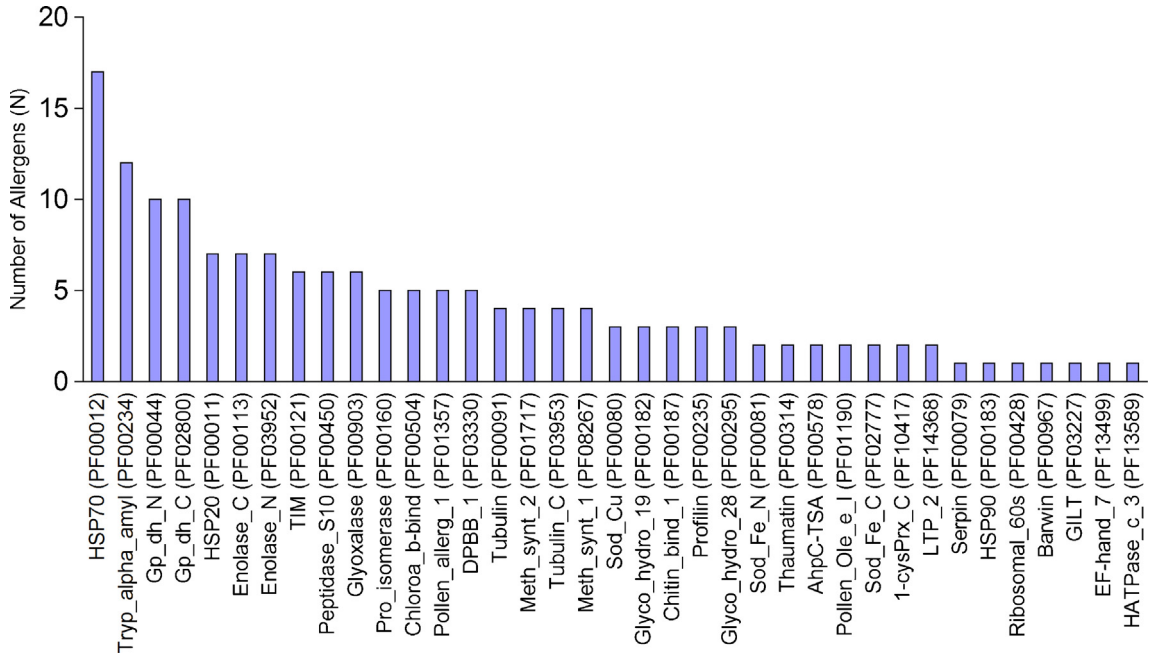


Fig. 1. Conserved protein domains among the putative rice allergens. All rice allergen sequences were searched in the Pfam-A database for the matching family. The number of allergens that contained specific protein domain was counted. The letters in the brackets stand for the Pfam accession numbers.

using NCBI on-line tool Primer-BLAST [25] (Supplementary Table 1) and tested by using Thermocycler (Bio-rad) (Supplementary Fig. 1). First-strand cDNA were synthesized by using GoScript™ Reverse Transcription System (Promega). Quantitative PCR were performed in StrataGene Mx3005P (Agilent).

The statistical analysis of the gene expression was performed by using SAS software package (SAS INC). Duncan multiple range test and *T* test were calculated with significant level ($P = 0.05$), and extreme significant level ($P = 0.01$).

3. Results

3.1. Identification of putative rice allergens

In order to investigate endogenous allergens in rice and their cross reactivity, we performed a full-length sequences alignment. One hundred and twenty-two protein-coding transcripts were identified as candidate allergens (Table 1), comparing to 22 entries of known rice allergens from the Allergen Nomenclature and FARRP

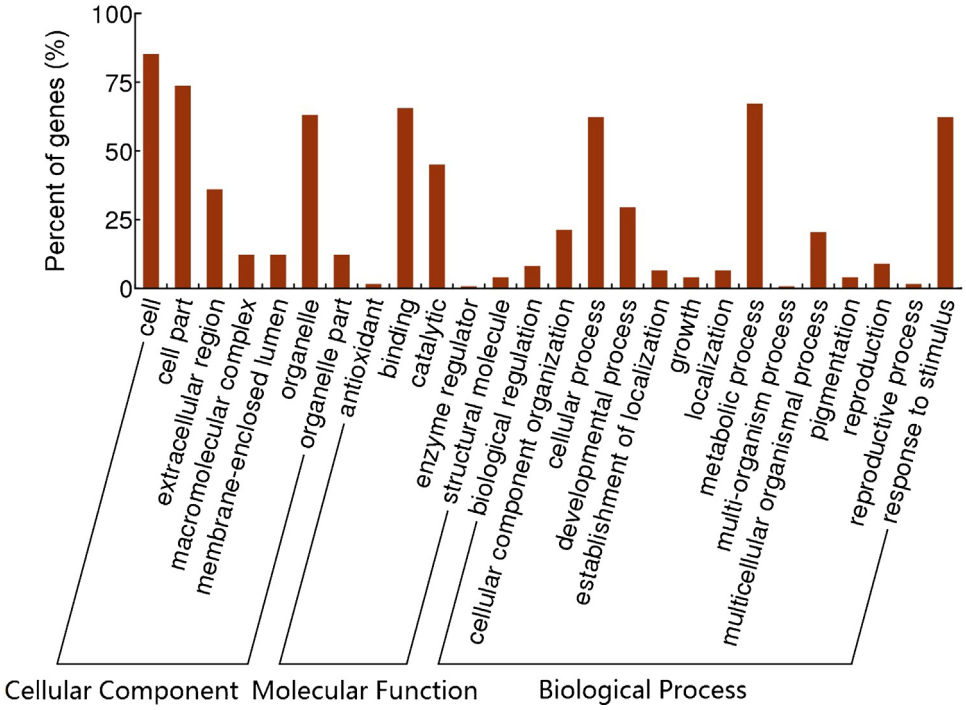


Fig. 2. GO classification of putative rice allergens. GO analysis was performed using WEGO tool in order to reveal the biological functions most frequently found among identified rice allergens. Of the 122 allergens, 116 contained GO annotations which were classified into three categories: cellular component, molecular function, and biological process.

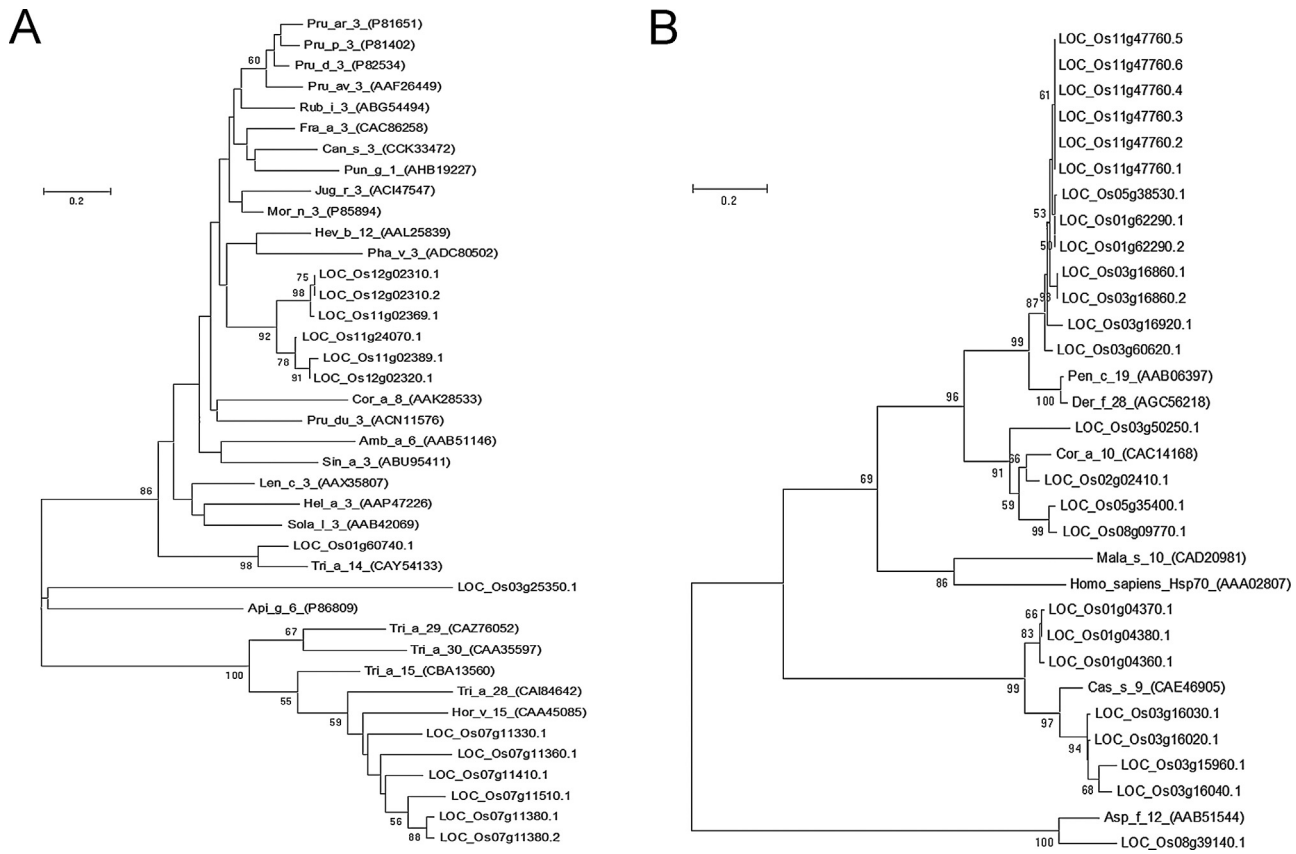


Fig. 3. Sequence conservation of Prolamin superfamily (A) and heat shock proteins (B). Prolamin superfamily comprises alpha-amylase/trypsin inhibitors and nonspecific lipid transfer proteins. Reference allergenic sequences of each family were adopted from the systematic allergen nomenclature database. Bootstrapped (1000 replicates) neighbor-joining trees were constructed using the PAUP. Numbers less than 50% (clustering 50 out of 100 times) were omitted due to possible collapse of the branch. Branch lengths proportional to genetic distance are indicated in the *scale bar*.

databases. Because of similarity, the 22 entries were matched with 19 proteins (highlighted in Table 1), which were distributing on five rice allergen groups: four proteins belong to Ory s 1, two proteins belong to Ory s 12, six proteins belong to Oryza glyoxalase I, six proteins belong to Oryza trypsin alpha-amylase inhibitor and one protein belongs to Oryza putative polcalcin Phl p 7, respectively.

Six conserved motifs were detected in the putative rice allergens using MEME analysis. These motifs were found in 34 out of the 122 sequences and their amino acid residues showed highly conserved. Four motifs were aligned with nucleotide-binding domain of the sugar kinase/HSP70/actin superfamily, the relation between its function and allergenicity should to be confirmed. No significant function was predicted in other two motifs.

3.2. Protein family distribution of potential rice allergens

Thirty-seven protein families (167 domains) were found from the 122 putative rice allergens (Fig. 1). Seventeen putative rice allergens contained HSP70 domain (PF00012). The HSP70s (70 kDa heat shock proteins) played an important role in assisting other proteins to fold properly and protecting cells from heat stress and toxic chemicals. Although proteins with similar structure ubiquitously expressed in bacteria and eukarya, validated allergenic HSP70s were only reported in dust mite (Der f 28) and fungi (Alt a 3, Pen c 19 and Mala s 10) and acted as airborne allergens. Tryp_alpha_amyl domain (protease inhibitor/seed storage/plant lipid transfer protein family, PF00234) was found in twelve allergens. Most known allergenic proteins in rice seed belong to this family [26]. Gp_dh_N (glyceraldehyde 3-phosphate dehydrogenase, GAPDH, NAD binding domain, PF00044) or Gp_dh_C (GAPDH, C-terminal domain, PF02800) were found in ten allergens, respectively. GAPDH in wheat (*Triticum aestivum*) named Tri a 34 had been listed in allergen database. Comparing with the result of allergen database-wide analysis, one novel protein family domain termed HATPase_c_3 (histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, PF13589) was added to the allergen-specific protein families. This domain was found in LOC_Os08g39140.1 next to a HSP90 domain, while needs to be verified in the further study.

3.3. Biological processes among those potential rice allergens

We utilized the GO annotation to characterize the putative allergens based on the biological functions. Of the 122 potential rice allergens, 116 contained GO annotations that belong to 92 different GO terms categorized into cellular component, molecular function and biological process (Fig. 2). For the ontology type of cellular component, 85.2% of 122 allergens were located in cell (GO: 0005623), followed by 73.8% in cell part (GO: 0044464) and 63.1% in organelle (GO: 0043226). As the molecular function, 65.6% of them were inferred to binding activity (GO: 0005488), 45.1% possess catalytic activity, a few proteins (4.1%) possess structural molecule activity (GO: 0005198), antioxidant activity (1.6%, GO: 0016209) and enzyme regulator activity (0.8%, GO: 0030234). The potential rice allergens were involved in fourteen subcategories of biological processes. The top three were metabolic process (67.2%, GO: 0008152), cellular process (62.3%, GO: 0009987) and response to stimulus (62.3%, GO: 0050896). Other important biological processes include developmental process (29.5%, GO: 0032502), cellular component organization (21.3%, GO: 0016043), and multicellular organismal process (20.5%, GO: 0032501).

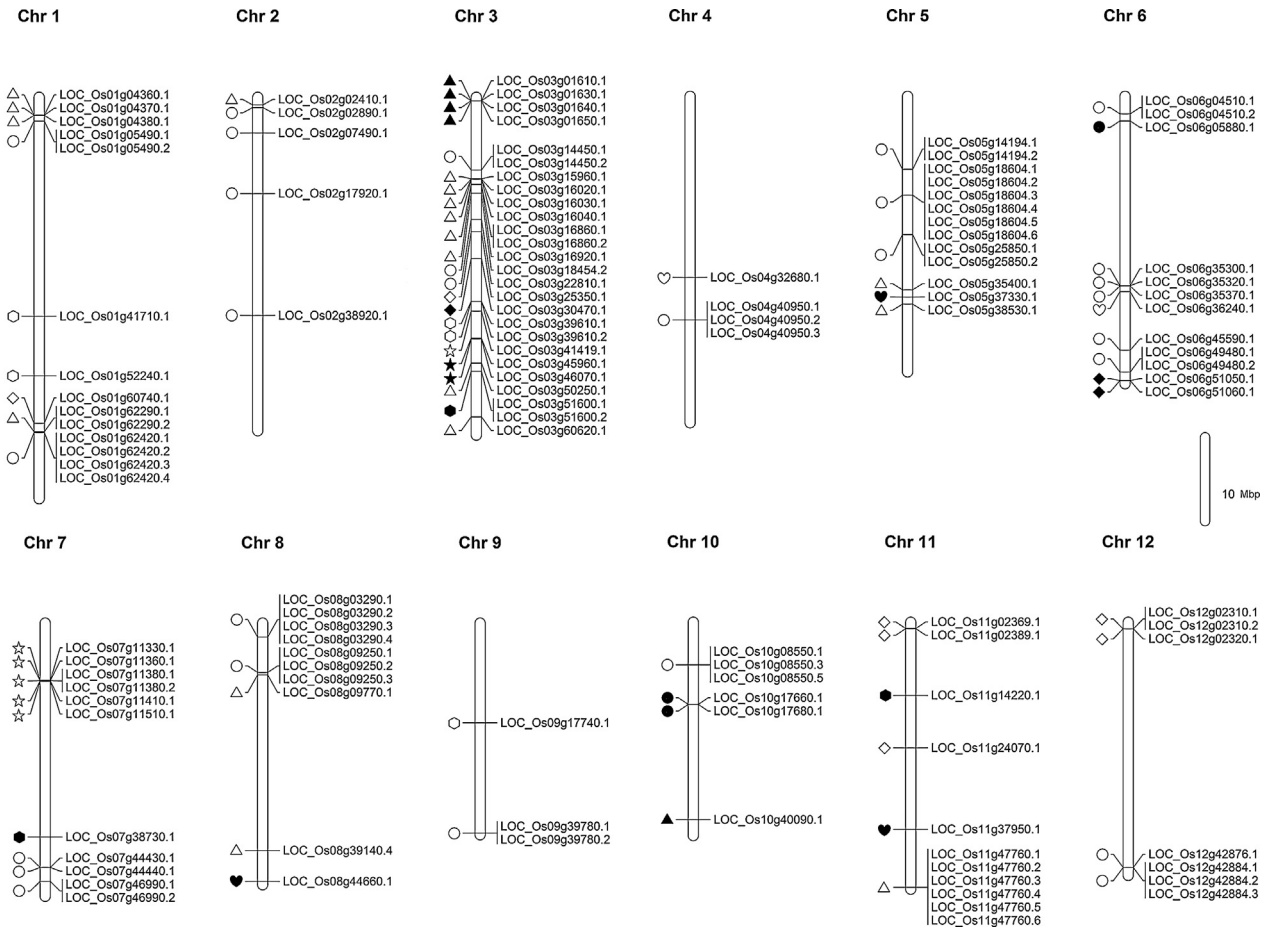


Fig. 4. Distribution of putative rice allergen genes across chromosomes. The hollow triangle for DnaK family protein, the hollow circle for metabolic enzyme, the hollow diamond for nonspecific lipid-transfer protein, the hollow pentacle for protease inhibitor, the hollow hexagon for chlorophyll binding protein, the hollow heart for pollen Ole e I allergen and extensin family protein, the solid triangle for expansin, the solid circle for profilin, the solid diamond for chitinase, the solid pentacle for thaumatin-like protein, the solid hexagon for tubulin and the solid heart for other proteins.

3.4. Phylogenetic diversity of allergens

The amino acid sequences of Prolamin superfamily (including alpha-amylase trypsin inhibitor and nonspecific lipid transfer protein) and DnaK protein family (or heat shock protein) were separately aligned to investigate their sequence diversities and phylogenetic relations (Fig. 3). Prolamin superfamily were grouped into four subgroups: all of the alpha-amylase/trysin inhibitors were clustered with the allergen Hor v 15 independently, and six nsLTPs were clustered with Hev b 12 and Pha v 3; both subgroup 3 and subgroup 4 only had a single nsLTP that were clustered with Tri a 14 and Api g 6, respectively. All 25 DnaK family proteins of rice were also divided into four subgroups: 13 proteins were clustered with Pen c 19 and Der f 28, four proteins were clustered with Cor a 10, seven proteins were clustered with Cas s 9, and the remaining single protein was clustered with Asp f 12. Phylogenetic analysis revealed the great divergence among the rice proteins shared common allergenic-liked domain.

3.5. The rice allergens distribution among the chromosomes

The putative rice allergens were scattered in all 12 chromosomes (Fig. 4), 26 allergens of them were distributed on chromosome 3, then 14 ones on chromosome 1, and 13 ones on chromosomes 5. Twenty-two screened loci have multiple transcripts as a consequence of post-transcriptional alternative splicing of precursor messenger RNA (pre-mRNA), such as LOC_Os01g05490, LOC_Os05g18604 and LOC_Os08g03290 have 2, 6 and 4 screened transcripts respectively. The diversification of cellular and organismal functions is mostly due to the expression of different transcripts and proteins from the same genes through alternative splicing [27]. While different gene sites even located on different chromosomes may encode the same kind of protein presenting as a multigene family, just as the loci LOC_Os01g41710, LOC_Os01g52240, LOC_Os03g39610 and LOC_Os09g17740 all express chlorophyll a-b binding proteins and share high identities ranging from 76.21% to 86.57% with the food allergen Api g 3. Moreover, multiple allergen transcripts clustering are found in rice genome. The reported rice seed allergenic proteins that belonged to trypsin alpha-amylase inhibitors arrange on chromosome 7 in a row, and four expansins acted as rice pollen allergens were clustered on chromosome 3.

3.6. Putative allergic gene expression in rice

Putative allergic candidates were studied by literatures searching and found that all the genes transcripts were expressed during experimental studies in Plant Comparative Genomics database (http://phytozome.jgi.doe.gov). Randomly 30 putative allergic gene expressions in post-tillering stage rice stem and leaves were analysis (Fig. 5). Comparing with internal control gene B-actin, four genes were extremely downregulated ($P = 0.01$): one chlorophyll a–b binding protein (Allgn6, LOC_Os01g41710.1), two nonspecific lipid-transfer proteins (Allgn106, LOC_Os11g02389.1

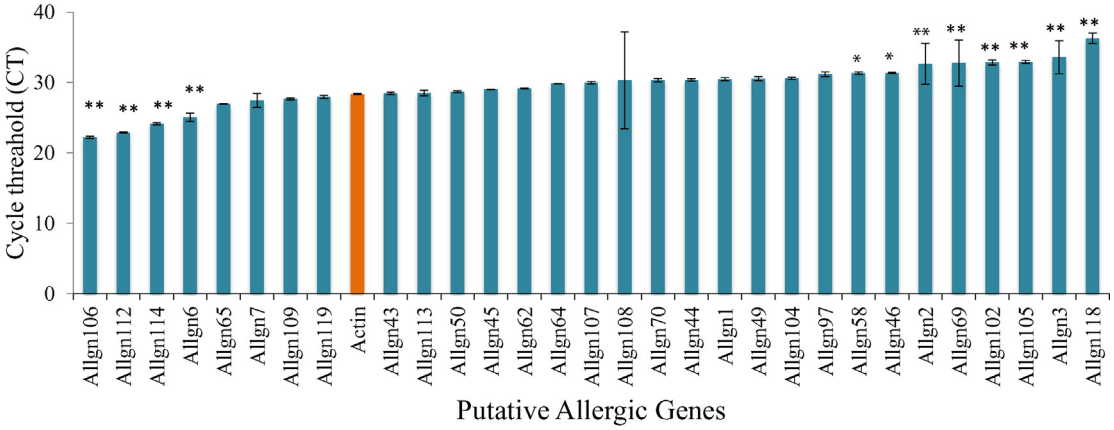


Fig. 5. qPCR results revealed some putative allergic protein transcripts expressed different in rice at post-tillering phase. ** Indicated extremely significant ($P = 0.01$), * indicated significant ($P = 0.05$).

and Allgn116, LOC_Os12g02310.1) and one heat shock protein P70 (Allgn112, LOC_Os11g47760.3). Eight transcripts were fund a higher expression than that of B-actin: two nonspecific lipid-transfer protein (Allgn118, LOC_Os12g02320.1 and Allgn105, LOC_Os11g02369.1), one Ole e 1-related protein (Allgn69, LOC_Os06g36240.1), one Zea pollen specific protein (Allgn46, LOC_Os04g32680.1), two cytosolic class I small heat shock protein (Allgn3, LOC_Os01g04380.1, and Allgn2, LOC_Os01g04370.1), one Profilin A (Allgn102, LOC_Os10g17660.1), and one Superoxide dismutase (Allgn58, LOC_Os05g25850.1).

4. Discussion

Identification of all the allergenic components in rice is necessary for the prediction of rice-related cross-reactivity and the diagnosis of rice allergy. In this study, we identified a total of 122 proteins as potential allergens in rice according to the similarity of amino acid sequences. Among them, a considerable part of proteins present as a multigene family across the rice genome, and even across a range of phylogenetic species. Post-transcriptional alternative splicing of allergen also produces a group of homologous proteins. Allergens from a single species with similar molecular weights, similar biochemical functions and more than 67% sequence identities are defined as isoallergens; isoforms or variants of isoallergen are typically defined as sequences with more than 90% identities [28]. The expression of isoallergens or isoforms showed a temporal and spatial transcript specificity, and could vary in different cultivars [29]. Both of alternative splicing and multigene family contribute to the diversity of rice allergen. A novel isoform of Pru av 1 (the major cherrry allergen) was identified and showed diverging IgE-binding properties with previously published Pru av 1 [30]. IgE binding capacity among the members of a single-allergen gene family in rice need to be further researched.

So far, 22 entries of rice allergens were accepted by Allergen Nomenclature and FARRP databases, including Oryza trypsin alpha-amylase inhibitor (15 entries), Ory s 1 (3 entries), Oryza glyoxalase I (2 entries), Oryza putative polcalcin Phl p 7 (1 entry) and Ory s 12 (1 entry). The rice allergenic protein (RA) from trypsin alpha-amylase inhibitor family was firstly cloned from cDNA libraries of maturing rice seeds [31]. Then, more cDNA clones and genomic clones encoding trypsin alpha-amylase inhibitor were isolated from rice, and listed by allergen databases. Some allergen entries may be isoallergens or incomplete allergenic protein fragments only. This is can be the reason why the number of database-recorded allergens is more than rice allergens that we identified. There seems to be a lack of homology to the 2S albumin family of proteins, which arguably is the most important food allergy group in nuts and seeds [32]. We also found less LTP family proteins. The LTPs of rice have been noted as being cross-reactive in IgE binding for a few patients with severe peach (prune family) allergy or maize allergy [33].

The screened rice allergens had highly similarities not only in overall sequence but also in specific motif with identified allergens. All allergen in database were grouped to only 130 of 9318 protein families as defined in Pfam A, they were highly biased toward certain families [34]. The specific motifs between query sequences and allergens can be useful in searches for potential allergens. But the classification of Pfam family is still unable to reveal the factors that determined allergenicity.

Although Table 1 only lists the optimal allergen of each alignment, the results of cross-reactive prediction were fairly complicated because every retrieved rice sequences share high identities with a cluster of allergenic proteins that may come from different species and may induce cross reactivities with each other. The prediction of probable cross reactivities between potential rice allergens and known allergens may explain the hypersensitivities happened in the reported clinical cases of allergy that activated by different approaches.

Author contribution statement

F.C. Zhu, R.Z. Jia and A.P. Guo designed and conducted the study. F.C. Zhu and R.Z. Jia contributed to the writing of the manuscript, L. Xu and Q.X. Huang contributed to the bioinformatic analysis, H. Kong, Y.L. Guo and Y.J. Zhu reviewed the allergen databses.

Conflict of interest

The authors declare that they have no conflict of interest.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgment

This work was supported by the Special Fund for Agro-Scientific Research in the Public Interest of the People's Republic of China (Grant No. 201403075).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.toxrep.2015.07.017.

References

[1] S.H. Sicherer, H.A. Sampson, Food allergy: epidemiology, pathogenesis, diagnosis, and treatment, J. Allergy Clin. Immunol. 133 (2014) 291–307, http://dx.doi.org/10.1016/j.jaci.2013.11.020.

[2] M. Shibasaki, H. Nemoto, T. Kuroume, Allergenicity and lymphocyte-stimulating property of rice protein, J. Allergy Clin. Immunol. 64 (1979) 259–265, http://dx.doi.org/10.1016/0091-6749(79)90141-6.

[3] B. Wüthrich, T. Scheitlin, B. Ballmer-Weber, Isolated allergy to rice, Allergy 57 (2002) 263–264, http://dx.doi.org/10.1034/j.1398-9995.2002.1n3587.x.

[4] M. Nambu, N. Shintaku, S. Ohta, Rice allergy, Pediatrics 117 (2006) 2331–2332, http://dx.doi.org/10.1542/peds.2006-0427.

[5] D. Villalta, G. Longo, G. Mistrello, S. Amato, R. Asero, A case of rice allergy in a patient with baker's asthma, Eur. Ann. Allergy Clin. Immunol. 44 (2012) 207.

[6] M. Yasutomi, T. Kosaka, A. Kawakita, H. Hayashi, S. Okazaki, H. Murai, K. Miyagawa, M. Mayumi, Y. Ohshima, Rice protein-induced enterocolitis syndrome with transient specific IgE to boiled rice but not to retort-processed rice, Pediatr. Int. 56 (2014) 110–112, http://dx.doi.org/10.1111/ped.12198.

[7] M. Besler, S. Tanabe, A. Urusu, Allergen data collection-update: rice (*Oryza sativa*), Internet Symp. Food Allerg. 3 (2001) 1615–1682.

[8] R. Kumar, P. Srivastava, D. Kumari, H. Fakhr, S. Sridhara, N. Arora, S.N. Gaur, B.P. Singh, Rice (*Oryza sativa*) allergy in rhinitis and asthma patients: a clinico-immunological study, Immunobiology 212 (2007) 141–147, http://dx.doi.org/10.1016/j.imbio.2006.11.006.

[9] S.D. Russell, P.L. Bhalla, M.B. Singh, Transcriptome-based examination of putative pollen allergens of rice (*Oryza sativa* ssp. japonica), Mol. Plant 1 (2008) 751–759, http://dx.doi.org/10.1093/mp/ssn036.

[10] K. Hirano, S. Hino, K. Oshima, T. Okajima, D. Nadano, A. Urisu, F. Takaiwa, T. Matsuda, Allergenic potential of rice-pollen proteins: expression, immuno-cross reactivity and IgE-binding, J. Biochem. 154 (2013) 195–205, http://dx.doi.org/10.1093/jb/mvt044.

[11] T. Kato, E. Katayama, S. Matsubara, Y. Omi, T. Matsuda, Release of allergenic proteins from rice grains induced by high hydrostatic pressure, J. Agric. Food Chem. 48 (2000) 3124–3129, http://dx.doi.org/10.1021/jf000180w.

[12] R. Asero, S. Amato, B. Alfieri, S. Folloni, G. Mistrello, Rice: another potential cause of food allergy in patients sensitized to lipid transfer protein, Int. Arch. Allergy Immunol. 143 (2006) 69–74, http://dx.doi.org/10.1159/000098226.

[13] H.B. Krishnan, M.H. Chen, Identification of an abundant 56 kDa protein implicated in food allergy as granule-bound starch synthase, J. Agric. Food Chem. 61 (2013) 5404–5409, http://dx.doi.org/10.1021/jf4014372.

[14] R. Satoh, R. Nakamura, A. Komatsu, M. Oshima, R. Teshima, Proteomic analysis of known and candidate rice allergens between non-transgenic and transgenic plants, Regul. Toxicol. Pharm. 59 (2011) 437–444, http://dx.doi.org/10.1016/j.yrtph.2011.01.008.

[15] J. Golias, Z. Humlova, P. Halada, V. Habova, I. Janatkoa, L. Tckova, Identification of rice proteins recognized by the IgE antibodies of patients with food allergies, J. Agric. Food Chem. 61 (2013) 8851–8860, http://dx.doi.org/10.1021/jf402759f.

[16] C.H. Schein, O. Ivanciuc, W. Braun, Bioinformatics approaches to classifying allergens and predicting cross-reactivity, Immunol. Allergy Clin. N. Am. 27 (2007) 1–27, http://dx.doi.org/10.1016/j.iac.2006.11.005.

[17] L. Tuckova, J. Golias, Z. Humlova, P. Halada, V. Habova, I. Janatkova, Rice proteins recognized by immunoglobulin E antibodies of patients with food allergies, Clin. Transl. Allergy 4 (2014) O6, http://dx.doi.org/10.1186/2045-7022-4-S2-O6.

[18] R.E. Goodman, S. Vieths, H.A. Sampson, D. Hill, M. Ebisawa, S.L. Taylor, R. van Ree, Allergenicity assessment of genetically modified crops-what makes sense? Nat. Biotechnol. 26 (2008) 73–81, http://dx.doi.org/10.1038/nbt1343.

[19] S.A. Goff, D. Rickel, T.H. Lan, et al., A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), Science 296 (2002) 92–100, http://dx.doi.org/10.1126/science.1068275.

[20] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucl. Acids Symp. Ser. 41 (1999) 95–98.

[21] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, W.S. Noble, MEME SUITE: tools for motif discovery and searching, Nucl. Acids Res. 37 (2009) W202–W208, http://dx.doi.org/10.1093/nar/gkp335.

[22] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, Nucl. Acids Res. 42 (2014) D222–D230, http://dx.doi.org/10.1093/nar/gkt1223.

[23] J. Ye, L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, J. Wang, S. Li, R. Li, L. Bolund, J. Wang, WEGO: a web tool for plotting GO annotations, Nucl. Acids Res. 34 (2006) W293–W297, http://dx.doi.org/10.1093/nar/gkl031.

[24] C. Radauer, M. Bublin, S. Wagner, A. Mari, H. Breiteneder, Allergens are distributed into few protein families and possess a restricted number of biochemical functions, J. Allergy Clin. Immunol. 121 (2008) 847–852, http://dx.doi.org/10.1016/j.jaci.2008.01.025.

[25] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinform. 13 (2012) 134.

[26] E.A. Pastorello, L. Farioli, V. Pravettoni, J. Scibilia, et al., Maize food allergy: lipid-transfer proteins, endochitinases, and alpha-zein precursor are relevant maize allergens in double-blind placebo-controlled maize-challenge-positive patients, Anal. Bioanal. Chem. 395 (2009) 93–102.

[27] E. de Klerk, P.A.C. t Hoen, Alternative mRNA transcription, processing, and translation insights from RNA sequencing, Trends Genet. 31 (2015) 128–139, http://dx.doi.org/10.1016/j.tig.2015.01.001.

[28] C. Radauer, A. Nandy, F. Ferreira, R.E. Goodman, J.N. Larsen, J. Lidholm, A. Pomes, M. Raulf-Heimsoth, P. Rozynek, W.R. Thomas, Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences, Allergy 69 (2014) 413–419, http://dx.doi.org/10.1111/all.12348.

[29] Z. Yang, Y. Ma, L. Chen, R. Xie, X. Zhang, B. Zhang, M. Lu, S. Wu, L.J.W.J. Gilissen, R. van Ree, Z. Gao, Differential transcript abundance and genotypic variation of four putative allergen-encoding gene families in melting peach, Tree Genet. Genomes 7 (2011) 903–916, http://dx.doi.org/10.1007/s11295-011-0383-5.

[30] A. Reuter, D. Fortunato, L.P. Garoffo, L. Napolitano, S. Scheurer, M.G. Giuffrida, S. Vieths, A. Conti, Novel isoforms of Pru av 1 with diverging immunoglobulin E binding properties identified by a synergistic combination of molecular biology and proteomics, Proteomics 5 (2005) 282–289, http://dx.doi.org/10.1002/pmic.200400874.

[31] H. Izumi, T. Adachi, N. Fujii, T. Matsuda, R. Nakamura, K. Tanaka, A. Urisu, Y. Kurosawa, Nucleotide sequence of a cDNA clone encoding a major allergenic protein in rice seeds, FEBS Lett. 302 (1992) 213–216, http://dx.doi.org/10.1016/0014-5793(92)80443-K.

[32] P.R. Shewry, M.J. Pandya, The 2S albumin storage proteins, Seed Proteins (1999) 563–586.

[33] I. Giangrieco, C. Alessandri, C. Rafaiani, M. Santoro, S. Zuzzi, et al., Structural features, IgE binding and preliminary clinical findings of the 70 kDa lipid transfer protein from tomato seeds, Mol. Immunol. 66 (2015) 154–163, http://dx.doi.org/10.1016/j.molimm.2015.02.025.

[34] O. Ivanciuc, T. Garcia, M. Torres, C.H. Schein, W. Braun, Characteristic motifs for families of allergenic proteins, Mol. Immunol. 46 (2009) 559–568, http://dx.doi.org/10.1016/j.molimm.2008.07.034.