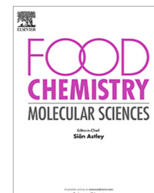
Food Chemistry: Molecular Sciences 1 (2020) 100002

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

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



Proteomic analysis of 2-chloroethanol extracts of rice (*Oryza sativa* L.) seeds

Songlin Ruan [a,b,*], Wenfei Xiao [a], Jieren Qiu [a], Weimin Hu [c], Wu Ying [a], Huizhe Chen [b], Jianxin Tong [a], Huasheng Ma [a,*]

[a] Institute of Crop Science, Hangzhou Academy of Agricultural Sciences, Hangzhou 310024, China
[b] State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China
[c] Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310012, China

ARTICLE INFO

Article history:
Received 28 February 2020
Received in revised form 21 May 2020
Accepted 5 July 2020
Available online 16 August 2020

Keywords:
Rice seed
Ethanol-soluble protein
Proteomic analysis
Seed storage protein
Protease inhibitors

ABSTRACT

Ethanol-soluble proteins, including prolamins, are one of the most important seed proteins in rice (*Oryza sativa* L.). However, little is known about the proteomic profile of ethanol-soluble protein fraction extracted from rice grain. In this work, the differential profile of ethanol-soluble proteins extracted by 2-chloroethanol and ethanol has been documented. Proteome analysis utilizing LC-MS/MS identified a total of 64 unique proteins in the 2-chloroethanol extract of rice seeds. The majority of these proteins had low molecular weight ranging from 10 to 25 kD and isoelectric point (pI) in mid-acidic (pH 5–pH 7) and mid-basic (pH 7–pH 9) ranges. Database searches combined with transmembrane domain (TMD) analysis revealed that >70% of identified proteins were hydrophobic, i.e., had at least one TMD. Gene ontology classification and enrichment analysis showed that the identified proteins were involved in13 types of biological processes, 5 types of cell components, and 17 types of molecular functions. These results were significant based on the hyper p-value of < 0.05. The most frequent categories of biological processes, cell components, and molecular functions were, respectively, type I hypersensitivity, extracellular space and extracellular region, and serine-type endopeptidase inhibitor activity. Interestingly, in addition to seed storage proteins such as prolamins and glutelins, certain allergen proteins, protease inhibitors, and lipid transfer proteins were identified in the extracts. Together, the collected data provide novel insights into the protein profile of 2-chloroethanol extract of rice seeds.

1. Introduction

Rice (*Oryza sativa* L.) is one of the most important cereal crops worldwide, and a food staple in many countries (Fei et al., 2019; Wang et al., 2019). Besides starch, rice grains contain abundant storage proteins, accounting for 8–13% of their dry weight. Based on the solubility in various solvents, rice seed storage proteins are classified into three groups: acid/alkaline-soluble glutelins, alcohol-soluble prolamins, and saline-soluble α-globulin (Yamagata & Tanaka, 1986). They account for, respectively, 60%–80%, 20%–30%, 2%–8% of the total protein content of rice grain (Kim, Lee, Yoon, Lim, & Kim, 2013). Prolamins are the second most abundant class of endosperm proteins in rice and are stored mostly in type I of protein body (PB-I)derived from the ER, whereas glutelins and α-globulin are accumulated in irregularly-shaped PB-II bodies derived from the protein storage vacuole (PSV) (Kim et al., 2013). The prolamins in PB-I are mainly composed of 13 kD polypeptides, while 10 kD and 16 kD polypeptides are relatively rare. Although prolamins are difficult to be digested, they affect the textural and pasting properties of rice flour and nutritional quality of rice grain (Baxter, Blanchard, & Zhao, 2004; Kim et al., 2013).

The commonly used solvents for rice prolamins are the mixtures of ethanol and water (e.g., 50–70% (v/v) ethanol) or 2-propanol and water (e.g., 55–70%(v/v) 2-propanol), with or without the addition of reducing agents (Gorinstein, Nue, & Arruda, 1991). Interestingly, pure 2-propanol is also an effective solvent for the extraction of prolamin from rice flour (Baxter et al., 2004). In other cereals, such as wheat and barley, 2-chloroethanol is frequently used for the extraction of prolamin and has been successfully used for cultivar identification (Cooke, 1984; Weiss, Postel, & Görg, 1991). However, 2-chloroethanol is rarely used in the extraction of rice prolamin. Ethanol-soluble cereal proteins are extremely difficult to purify and characterize due to their low solubility and tendency to polymerize. These proteins are most often separated by the reverse phase HPLC (RP-HPLC) due to the high resolution of this technique (Bietz, 1983; Paulis & Bietz, 1986; Pernollet et al., 1989). In addition to RP-HPLC, two types of 1-D polyacrylamide gel electrophoresis (PAGE), including SDS-PAGE and acid-PAGE, are also effective in separating ethanol-soluble proteins for the analysis of genetic diversity. In particular, acid-PAGE has been frequently used for cultivar identification (Bean & Lookhart, 2000; Cooke, 1984; Weiss et al., 1991). Interestingly, the combination of gel electrophoresis and

* Corresponding authors at: Institute of Crop Science, Hangzhou Academy of Agricultural Sciences, Hangzhou 310024, China.
E-mail addresses: ruansl1@hotmail.com, (S. Ruan), huwm168@zju.edu.cn, (W. Hu), 2789572094@qq.com. (H. Ma).

http://dx.doi.org/10.1016/j.fochms.2020.100002
2666-5662/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

chromatographic techniques led to the development of capillary zone electrophoresis (CZE) in free solution (Tsuji, 1991), which has been widely used as a powerful tool for the separation of ethanol-soluble seed proteins of cereal crops and tobacco. The broad application of CZE results from its higher resolution and shorter analysis time than that offered by high-performance liquid chromatography (HPLC) and gel electrophoresis (Deng, Lu, & Yan, 1998; Lookhart & Bean, 1995). Since the limited protein solubility can be overcome by the addition of ethanol and organic solvents to focusing buffer, isoelectric focusing (IEF) has been used to separate storage proteins from several cereals (Bean & Lookhart, 2000). Görg and coworkers developed a technique to generate immobilized pH gradients in the first dimension (IPG-Dalt) of 2-D electrophoresis (2-DE) (Görg, Postel, Baumer, & Weiss, 1992), which has been applied in the discrimination of barley cultivars with different malting grades. However, given its technical limitations, such as poor reproducibility, low sensitivity, and long duration, 2-DE is rarely used in the separation of ethanol-soluble proteins and cultivar identification in cereals.

In cereals, ethanol-soluble protein fractions extracted from seeds were generally thought to represent storage proteins. However, in addition to storage proteins, seeds contain two other groups of proteins: structural and metabolic proteins, and protective proteins (Shewry & Halford, 2002). Therefore, other ethanol-soluble proteins, such as structural and metabolic proteins and protective proteins, may also be present in the ethanol-soluble protein fraction. For example, an earlier study demonstrated that besides avenins, the true prolamins, three low-molecular weight proteins, including two α-amylase/trypsin inhibitors and one novel protein, were also found in the ethanol-soluble protein fraction extracted from the oat grains (Pernollet et al., 1989).

With the recent development of proteomic techniques, particularly with the significant improvement in the sensitivity, mass accuracy, and fragmentation in the field of mass spectrometry (MS), proteomic strategies became widely adopted (Walther & Mann, 2010) and accelerated the progress in seed proteomics research. Thus far, synthesis and processing of seed storage proteins have been a major focus of seed proteomics (Miernyk & Hajduch, 2011). Similarly, a significant amount of research has been devoted to the proteomic analysis of the development and germination of rice seeds (Cheng et al., 2017; Deng, Gong, & Wang, 2013; Koller et al., 2002; Komatsu, Konishi, Shen, & Yang, 2003; Wang, Liu, Song, & Møller, 2015; Yang et al., 2007). In addition, 2-DE and MALDI-TOF-MS identified 302 proteins in mature rice seeds (Yang et al., 2013). However, little is known about the proteomic profile of ethanol-soluble protein fraction extracted from rice seeds.

Therefore, the present investigation was focused on the analysis of the proteomic profile of 2-chloroethanol extract of rice seeds using LC-MS/MS. First, the profile of ethanol-soluble proteins from rice seeds extracted with 2-chloroethanol/water (25% v/v) was compared with that obtained with ethanol/water (50% v/v) using SDS-PAGE. Subsequent proteomic analysis revealed that besides seed storage proteins, including prolamins and glutelins, many metabolic and functional proteins, particularly allergens and protease inhibitors, were identified among the extracted proteins. These findings provide new insights into the protein profile of 2-chloroethanol extract of rice seeds.

2. Materials and methods

2.1. Preparation of rice seeds and 2-chloroethanol extraction

Seeds of rice (*O. sativa* cv. Aichi asahi) were supplied by China National Rice Research Institute (Fuyang, Zhejiang province, China). Rice grains were de-hulled and manually ground into dry powder. A sample of 1 g of dry powder was transferred to a 50 ml centrifuge tube and mixed with 50 ml of 25% (v/v) 2-chloroethanol (Shanghai Guoyao Group Chemical Reagent Co., Ltd., China) or 50% (v/v) ethanol (Shanghai Guoyao Group Chemical Reagent Co., Ltd., China). The tube was rocked on a shaker (THZ-D, Jiansu, China) at 150 rpm overnight and then centrifuged for 15 min at 13,000*g*. The supernatant was transferred into another centrifuge tube, and a 12 μl sample was taken to determine the protein concentration using the Bradford method (Coomassie Protein Assay Kit, Bio-Rad, Hercules, CA, USA). Bovine serum albumin was used as the standard.

2.2. SDS-PAGE analysis

A sample of 20μg of ethanol-soluble proteins was loaded onto 12% SDS-PAGE gel. Molecular weight markers (Bio-Rad)were used to measure the size of bands corresponding to ethanol-soluble proteins. Constant voltage was set at 100 V, and the run time was approximately 1 h. The gel was stained overnight in 0.12% (w/v) colloidal Coomassie brilliant blue G-250, 10% (w/v) phosphoric acid, 10% (w/v) ammonium sulfate, and 20% (v/v) methanol. The gel was then destained in distilled water until clear bands appeared, and scanned to acquire gel image using a GS-800 imaging densitometer (Bio-Rad).

2.3. Protein digestion

Protein samples (100 μg) were added to 50 mM NH_4HCO_3 to the final volume of 150 μl. Protein samples were added to 100 mM dithiothreitol (DTT) and a final concentration of DTT was 10 mM, mixed at 600 rpm for 1 min, and incubated at 37 °C for 1 h. Iodoacetic acid (IAA, 0.5 M) was added to reach the final concentration of 50 mM, mixed at 600 rpm for 1 min, and stored at room temperature in the dark for 30 min. Protein mixture was filtered through a 10 kD cut-off membrane and centrifuged at 14,000*g* for 15 min at 4 °C. A total of 150 μl 50 mM NH_4HCO_3 was added, centrifuged at 14,000 ×*g* for 15 min; this step was repeated twice. A 100 μl aliquot of trypsin buffer (2 μg trypsin in 100 μl NH_4HCO_3 buffer) was added, and the sample was digested for 16–18 h at 37 °C. Filtrates were collected after centrifugation at 14,000 ×*g* for 10 min, addition of 100 μl NH_4HCO_3, and repeated centrifugation at 14,000 ×*g* for 10 min. The generated peptides were desalted using a C18 column (Thermo Fisher Scientific, Waltham, MA, USA).

2.4. LC-MS/MS

The mixture of tryptic peptides was desalted and separated by the Ettan™ MDLC system (GE Healthcare, Chicago, IL, USA), utilizing RP trap columns (Zorbax 300 SB C18, Agilent Technologies, Santa Clara, CA, USA) for desalting and RP columns (150 μm i.d., 100 mm length, Column Technology Inc., Fremont, CA, USA) for the separation. Mobile phase A consisted of 0.1% formic acid in HPLC-grade water, and mobile phase B consisted of 0.1% formic acid in acetonitrile. A sample containing 20 μg of tryptic peptide mixture was loaded onto the columns, and the separation was done at a flow rate of 2 μl/min using a linear gradient of 4–50% B for 120 min. A Finnigan™ LTQ™ linear ion trap MS (Thermo Electron, Madison, WI, USA) equipped with an electrospray interface was connected to the LC setup for the detection of eluted peptides. Data-dependent MS/MS spectra were obtained simultaneously. Each scan cycle consisted of one full MS scan in the profile mode followed by five MS/MS scans in the centroid mode with the following Dynamic Exclusion™ settings: repeat count 2, repeat duration 30 s, exclusion duration 90 s. Each sample was analyzed in triplicate.

MS/MS spectra were automatically searched against the non-redundant protein database from the NCBI (version 20100612) using the BioworksBrowser rev. 3.1.The peptides were constrained to be tryptic, and up to two missed cleavages were allowed. Carbamidomethylation of cysteine residues was treated as a fixed modification, whereas oxidation of methionine residues was considered as a variable modification. The mass tolerance allowed for the precursor ions was 2.0 Da and 0.2 Da for fragment ions. The protein identification criteria were based on Delta CN ($\geq$0.1) and cross-correlation scores (Xcorr, one charge$\geq$1.9, two charges $\geq$2.2, three charges $\geq$3.75).

2.5. Bioinformatic analysis of the identified proteins

Transmembrane regions in all identified proteins were predicted using the TMpred program from the ExPASy site (http://www.ch.embnet.org/software/TMPRED_form.html). The algorithm employed is based on the statistical analysis of TM base, a database of naturally occurring transmembrane proteins as described by Hofmann and Stoffel (1993). The Gene Ontology (GO) enrichment analysis of the identified proteins was performed by searching the Rice Oligonucleotide Array Database (ROAD) (http://www.ricearray.org, version 20130122). To determine the confidence of the GO enrichment results, the hyper p-value of < 0.05 was considered statistically significant. The GO classification of these proteins was illustrated using the Sigmaplot software, version 10.0 (Systat Software, San Jose, CA, USA). KEGG enrichment analysis of the identified proteins was performed in DAVID 6.8 (https://david.ncifcrf.gov/).

3. Results

3.1. Protein profile and identification of 2-chloroethanol extract of rice seeds

SDS-PAGE demonstrated that the protein profile of 2-chloroethanol extract of rice seeds was distributed in the molecular weight range of 10–66 kD. Most proteins were in the low molecular weight range, 10–25 kD. The differences in the profile of ethanol-soluble proteins between 2-chloroethanol/water (25% v/v) and ethanol/water (50% v/v) extracts were also compared. As shown in Fig. 1, under the same loading conditions, the 2-chloroethanol/water extract appeared to have a larger amount of ethanol-soluble proteins, particularly of low molecular weight (10–25 kD) than the ethanol/water extract. This finding indicates a different profile of ethanol-soluble proteins between 2-chloroethanol/water and ethanol/water extracts.

Subsequently, proteins obtained by extraction with 2-chloroethanol were digested with trypsin to obtain a mixture of peptides. A total of 1940 unique peptides were detected using LC-MS/MS (Table 1S). By database search and match, 64 unique proteins were identified (Table 1).

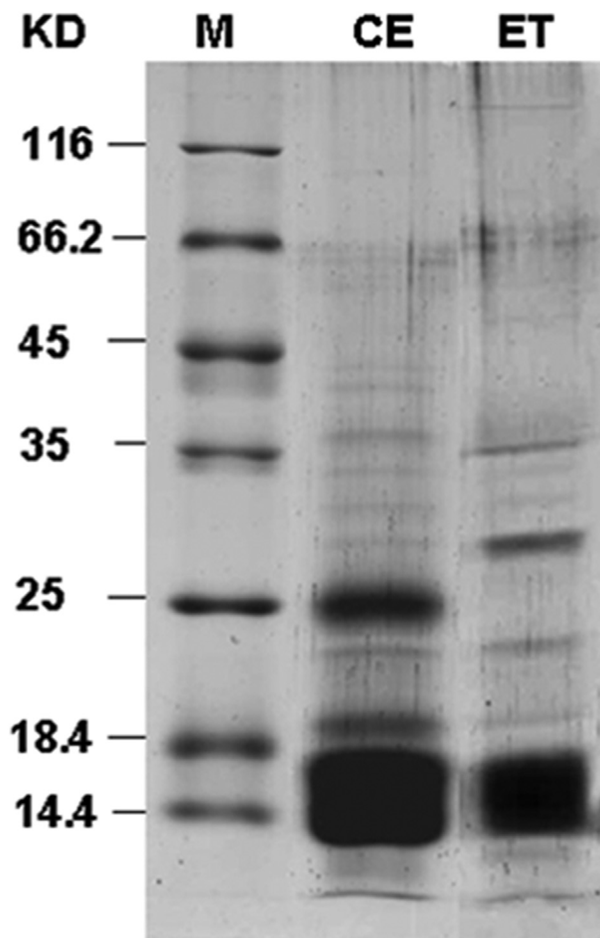


Fig. 1. Protein profile of 2-chloroethanol extract from rice seeds. 20μg of ethanol-soluble proteins2-chloroethanol extract from rice seeds were loaded onto 1-D SDS-PAGE gel (12% gel). Gels were stained overnight with colloidal Coomassie brilliant blue. Numbers on the left represent apparent molecular masses. CE and ET denote 2-chloroethanol/water (25% v/v) and ethanol/water (50% v/v), respectively.

3.2. Biochemical properties of identified proteins

As shown in Table 1, the molecular weights of all identified proteins were distributed in the range from 8.5 to 64.5 kD, with most proteins in the range from 10 to 25 kD (Fig. 2A). This result was consistent with the SDS-PAGE analysis of the 2-chloroethanol extract of rice seeds. Similarly, the pI values of all identified proteins ranged from 4.7 to 10.1, with the majority of them having mid-acidic (5–7) and mid-basic (7–9) pI (Fig. 2B).

Database searches combined with transmembrane domain (TMD) analysis revealed that over 70% of the identified proteins were hydrophobic and had at least one TMD (Fig. 3). Proteins containing only one TMD were the most abundant class, accounting for 46% of all identified proteins. The second most abundant class, 17%, was represented by proteins containing two TMDs. Proteins containing three or more TMDs constituted the smallest class, accounting only for 10% of all proteins.

3.3. KEGG (Kyoto Encyclopedia of Genes and Genomes) functional classification of identified proteins

As shown in Fig. 4, the GO (Gene Ontology) classification and enrichment analysis of the identified proteins was performed in terms of cellular component, molecular function, and biological processes by ROAD searching. The proteins were classified as belonging to 18 types of cell components in terms of GO annotation (Table 2S-1), but only 5 types were significant (hyper p-value of < 0.05) (Table 2S-1). These components included extracellular space, extracellular region, periplasmic space, lipid particle, cell outer membrane, and monolayer-surrounded lipid storage body (Table 2S-1 and Fig. 4). The two largest classes of proteins were located in the extracellular space and extracellular region (Fig. 4).

Similarly, according to the GO annotation, the identified proteins were classified as having 28 types of molecular functions (Table 3S-1), of which 17 types were significant (hyper p-value of < 0.05) (Table 3S-1). These 17 types of molecular functions were related to serine-type endopeptidase inhibitor activity, nutrient reservoir activity, ice binding, metal ion binding, lipid binding, antioxidant activity, superoxide dismutase activity, peptidase inhibitor activity, unfolded protein binding, copper ion binding, α-amylase inhibitor activity, peptide-methionine-(*S*)-S-oxide reductase activity, peroxiredoxin activity, endopeptidase inhibitor activity, acyl-CoA dehydrogenase activity, nucleic acid binding, and protein disulfide oxidoreductase activity. The most frequent category of molecular function among the identified proteins was serine-type endopeptidase inhibitor activity (Fig. 4).

In addition to the cell component and molecular function categories, the GO annotation classified the identified proteins as being involved in 21 types of biological processes. However, only 13 of them were significant (hyper p-value of < 0.05) (Table 4S-1). These processes included type I hypersensitivity, response to freezing, homoiothermy, lipid transport, response to stress, protein folding, superoxide metabolic process, cell redox homeostasis, response to water, response to acidity, response to the abscisic acid stimulus, and embryo development (Table 4S-1 and Fig. 4). The molecular process annotated most frequently to the identified proteins was the type I hypersensitivity (Fig. 4).

KEGG analysis showed that the identified proteins were enriched in the peroxisome pathway (osa04146) only. P93407 (superoxide

Table 1
Identification of ethanol-soluble proteins in rice seeds by LC MS/MS.

No.	Accession no.[a]	RAP loci	Locus ID[b]	MW (kD)[c]	pI[d]	Protein name
1	Q01882	Os07g0214300	LOC_Os07g11380.2	17,865	8.06	RAL4 - Seed allergenic protein RA5/RA14/RA17 precursor, expressed
2	BAF21097.1	Os07g0213800	LOC_Os07g11330.1	17,270	8.72	RAL2 - Seed allergenic protein RA5/RA14/RA17 precursor, expressed
3	P29421.2	Os04g0526600	LOC_Os04g44470.1	21,418	8.66	KUN1 - Kunitz-type trypsin inhibitor precursor, expressed
4	BAF26923.1	Os10g0505900	LOC_Os10g36180.1	45,582	4.74	Expressed protein
5	BAA01998.1	None[e]	None	17,888	8.06	Allergenic protein
6	Q01883	Os07g0214100	LOC_Os07g11360.1	17,568	6.92	RAL3 - Seed allergenic protein RA5/RA14/RA17 precursor, expressed
7	BAF21100.1	Os07g0214600	LOC_Os07g11410.1	16,983	8.36	RAL5 - Seed allergenic protein RA5/RA14/RA17 precursor, expressed
8	BAF08359.1	Os02g0250600	LOC_Os02g15250.1	47,315	6.4	Late embryogenesis abundant domain-containing protein, putative, expressed
9	BAA01996.1	None	None	17,118	8.74	Allergenic protein
10	Q01881	Os07g0215500	LOC_Os07g11510.1	17,293	8.36	RAL6 - Seed allergenic protein RA5/RA14/RA17 precursor, expressed
11	BAA00800.1	None	None	15,081	5.92	Copper/zinc-superoxide dismutase
12	BAF22321.1	Os07g0638300	LOC_Os07g44430.1	24,042	5.97	Peroxiredoxin, putative
13	BAF04641.1	Os01g0276300	LOC_Os01g16920.1	24,485	6.07	Embryonic protein DC-8, putative, expressed
14	BAF25418.1	Os09g0482600	LOC_Os09g30439.1	64,498	5.58	Heat shock protein, putative
15	BAF12746.1	Os03g0663400	LOC_Os03g46060.1	22,766	7.37	Thaumatin family domain containing protein, expressed
16	BAF23472.1	Os08g0327700	LOC_Os08g23870.1	15,163	9.1	Late embryogenesis abundant group 1, putative, expressed
17	BAA07772.1	None	None	12,335	8.74	Allergenic protein
18	P29835.1	Os05g0499100	LOC_Os05g41970.1	21,055	7.48	SSA1 - 2S albumin seed storage family protein precursor, expressed
19	BAF10944.1	Os03g0159600	LOC_Os03g06360.2	27,190	4.95	Late embryogenesis abundant protein D-34, putative, expressed
20	BAF18109.1	Os05g0542500	LOC_Os05g46480.4	20,514	5.89	Late embryogenesis abundant protein, group 3, putative, expressed
21	BAF11884.2	Os03g0322900	LOC_Os03g20680.1	36,791	6.39	Late embryogenesis abundant protein 1, putative, expressed
22	BAF23070.1	Os08g0189100	LOC_Os08g08960.1	23,712	6.4	Cupin domain containing protein, expressed
23	BAF12032.1	Os03g0351500	LOC_Os03g22810.1	15,251	5.71	Copper/zinc superoxide dismutase, putative, expressed
24	BAF12732.2	Os03g0659300	LOC_Os03g45720.1	21,930	8.88	Glyoxalase family protein, putative, expressed
25	EAZ29742.1	None	None	54,002	9.06	Hypothetical protein OsJ_013225
26	EAZ10632.1	None	None	11,998	7.72	Hypothetical protein OsJ_000457
27	P07728.1	Os01g0762500	LOC_Os01g55690.1	56,247	9.09	Glutelin, putative, expressed (Glutelin type I)
28	BAF08683.1	Os02g0453600	LOC_Os02g25640.2	57,120	8.96	Glutelin, putative, expressed
29	BAF22692.1	Os08g0104400	LOC_Os08g01370.1	10,490	5.13	Expressed protein
30	BAA00462.1	None	None	56,320	8.93	Prepro-glutelin
31	BAF12166.1	Os03g0385400	LOC_Os03g26820.1	16,899	8.62	LTPL52 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
32	BAF21096.1	Os07g0213600	LOC_Os07g11310.1	15,891	6.13	LTPL166 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
33	BAF21106.1	Os07g0216700	LOC_Os07g11650.1	16,477	7.48	LTPL164 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
34	BAF24082.1	Os08g0502400	LOC_Os08g39270.1	28,354	8.93	Fasciclin domain containing protein, expressed
35	P14323.1	Os02g0249900	None	56,551	9.26	Glutelin type-B 1 precursor
36	BAF28520.1	Os11g0582400	LOC_Os11g37270.1	47,385	6.04	AMBP1 - Antimicrobial peptide MBP-1 family protein precursor, expressed
37	EAZ26837.1	None	None	61,439	7.18	Hypothetical protein OsJ_010320
38	EAZ12939.1	None	None	8533.1	4.78	Hypothetical protein OsJ_002764
39	BAF21113.1	Os07g0222000	LOC_Os07g12080.1	15,767	7.48	LTPL169 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
40	P30287.1	Os01g0702500	LOC_Os01g50700.1	23,236	6.64	Dehydrin family protein, expressed Dehydrin Rab25 (Drought-resistant protein 1) (rDRP1)
41	P46520.1	Os05g0349800	LOC_Os05g28210.1	10,165	5.57	Embryonic abundant protein 1
42	BAF16634.2	Os05g0160200	LOC_Os05g06770.1	19,946	10.01	40S ribosomal protein S27a, putative, expressed
43	P93407.1	Os08g0561700	LOC_Os08g44770.2	21,301	5.79	Copper/zinc superoxide dismutase, putative, expressed
44	Q0JR25	Os01g0124401	LOC_Os01g03360.1	27,790	5.38	BBTI5 - Bowman-Birk type bran trypsin inhibitor precursor, expressed
45	EAZ13254.1	None	None	22,132	9.14	Hypothetical protein OsJ_003079
46	BAF10787.1	Os03g0133100	LOC_Os03g04080.1	12,631	5.2	Expressed protein
47	Q6ERU3	Os02g0268100	LOC_Os02g16820.1	56,808	9	Glutelin, putative, expressed
48	P55142	Os04g0508300	LOC_Os04g42930.2	11,774	5.77	OsGrx_C2.2 - glutaredoxin subgroup I, expressed
49	BAF22828.1	Os08g0129200	LOC_Os08g03520.1	18,694	6.28	Retrotransposon protein, putative, Ty1-copia subclass, expressed
50	EAZ38338.1	None	None	33,765	7.37	Hypothetical protein OsJ_021821
51	BAF12898.1	Os03g0699000	LOC_Os03g49190.1	17,222	10.1	Oleosin, putative, expressed
52	BAF21105.1	Os07g0216600	LOC_Os07g11630.1	16,406	7.48	LTPL163 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
53	BAF22842.1	Os08g0131200	LOC_Os08g03690.1	12,048	8.1	LTPL24 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
54	BAF29006.1	Os12g0115100	LOC_Os12g02320.2	11,345	9.58	LTPL12 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
55	BAF24481.1	Os09g0109600	LOC_Os09g02180.1	7847.6	5.58	Expressed protein
56	BAF12070.1	Os03g0360700	LOC_Os03g24600.1	14,672	6.81	Peptide methionine sulfoxide reductase msrB, putative, expressed
57	BAF22161.1	Os07g0609000	LOC_Os07g41820.1	26,500	7.11	Stress responsive A/B Barrel domain containing protein, expressed
58	EAZ11000.1	None	None	23,597	4.73	Hypothetical protein OsJ_000825
59	BAF17132.1	Os05g0330600	LOC_Os05g26620.1	16,883	8.79	PROLM14 - Prolamin precursor, putative, expressed
60	BAF11001.1	Os03g0168100	LOC_Os03g07180.3	41,860	8.26	Embryonic protein DC-8, putative, expressed
61	BAF14618.1	Os04g0404400	LOC_Os04g33150.1	32,107	4.91	Desiccation-related protein PCC13-62 precursor, putative, expressed
62	BAF13030.1	Os03g0723400	LOC_Os03g51350.1	18,054	6.44	Expressed protein
63	AAT99435.1	None	None	40,488	4.7	Lysm-domain GPI-anchored protein
64	BAF22821.1	Os08g0127900	LOC_Os08g03410.1	58,046	8.72	Glutelin, putative, expressed

[a] Accession no.: accession number in NCBI database.
[b] Locus ID: locus identification.
[c] MW(kD): match molecular weight.
[d] pI: Isoelectric point values.
[e] None: not available.

dismutase [CuZn], chloroplastic), P28757 (superoxide dismutase [CuZn] 2), and Q0DRV6 (superoxide dismutase [CuZn] 1) represented the only 3 proteins identified as involved in the peroxisome pathway, which functions in the removal of free radicals (Table S5).

4. Discussion

The present investigation compared the difference between the profiles of ethanol-soluble proteins from rice seeds extracted using 2-chloroethanol/

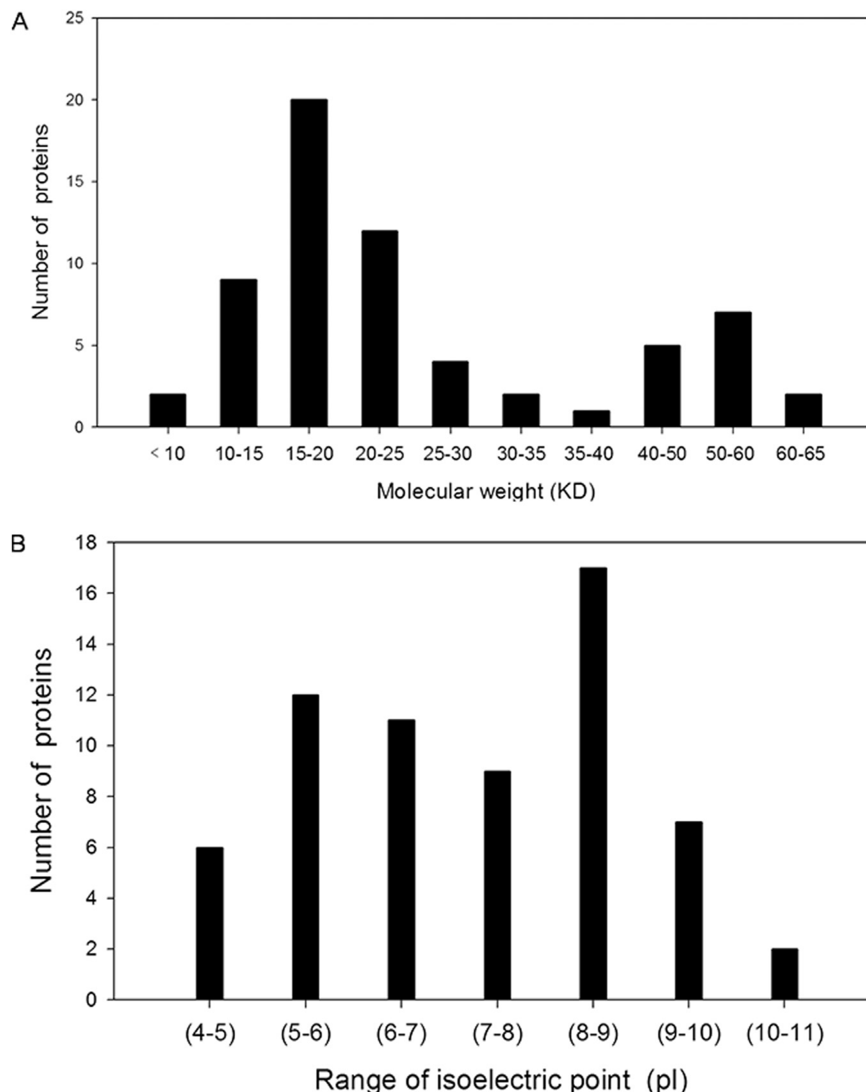


Fig. 2. Distribution of size and isoelectric point all identified proteins from 2-chloroethanol extract. A. Size distribution of all identified proteins is in a range of 8.5–64.5 kD in terms of minimal and maximal protein molecular weight, the majority of which are in a range of 10–25 kD. B. Isoelectric point distribution of all identified proteins is in a range of pH 4.7–10.1 in terms of minimal and maximal protein pI, the majority of which are in mid-acidic (5–7) and mid-basic (7–9) pH ranges. pI represents isoelectric point.

water (25% v/v) and ethanol/water (50% v/v). Although in other cereals, such as wheat and barley, 2-chloroethanol/water is commonly used for the extraction of ethanol-soluble proteins for cultivar identification (Cooke, 1984; Weiss et al., 1991), it is less frequently employed than ethanol/water. The current results demonstrated that a difference exists between profiles of proteins extracted using 2-chloroethanol/water and ethanol/water. This difference is predominantly apparent in low-molecular-weight proteins, ranging in size from 10 to 25 kD (Fig. 1). This range corresponds exactly to the size of ethanol-soluble proteins, including prolamins.

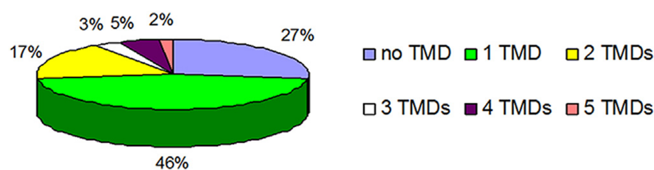


Fig. 3. Analysis of transmembrane domains of all identified proteins from 2-chloroethanol extract. Over 70% of all identified proteins are hydrophobic, which have at least one TMD. Proteins which contain only one TMD are the most abundant class, accounting for 46% of all identified proteins. TMDs represent transmembrane domains.

The current work used the MS-based shotgun proteomics to identify all tryptic peptides obtained from ethanol-soluble proteins. Prior to the MS analysis, these peptides were separated by RP-HPLC. RP-HPLC is commonly used for the separation of ethanol-soluble proteins due to its high resolution and its applicability to hydrophobic molecules (Bietz, 1983; Paulis & Bietz, 1986; Pernollet et al., 1989). Therefore, the MS-based shotgun proteomics can be considered a proper tool for identifying ethanol-soluble proteins. Here, a total of 64 unique proteins were identified using LC-MS/MS (Table 1), and over 70% of them were hydrophobic (Fig. 4). Interestingly, besides PROLM14, a true prolamin, over 40 ethanol-soluble proteins, are also found in these hydrophobic ones, which is similar to the result that besides avenins, three low-*Mr* proteins also appeared in oat ethanol-soluble protein fraction (Pernollet et al., 1989). In addition to prolamin, some glutelins are also present in the ethanol-soluble fraction of proteins from rice grains. It is usually suggested that glutelins represent a type of acid/alkaline-soluble storage proteins in rice seeds. Although the existence of alcohol-soluble glutelins in rice seeds has not been reported yet, they are present in sorghum seeds (Smith & Smith, 1988). Therefore, we raise the possibility that alcohol-soluble glutelins may exist in rice grains as well.

In addition, it should be noted that only 51 proteins could have been used for GO enrichment analysis by ROAD since the locus IDs are not available for the remaining proteins (Table 1). In the GO enrichment analysis, a

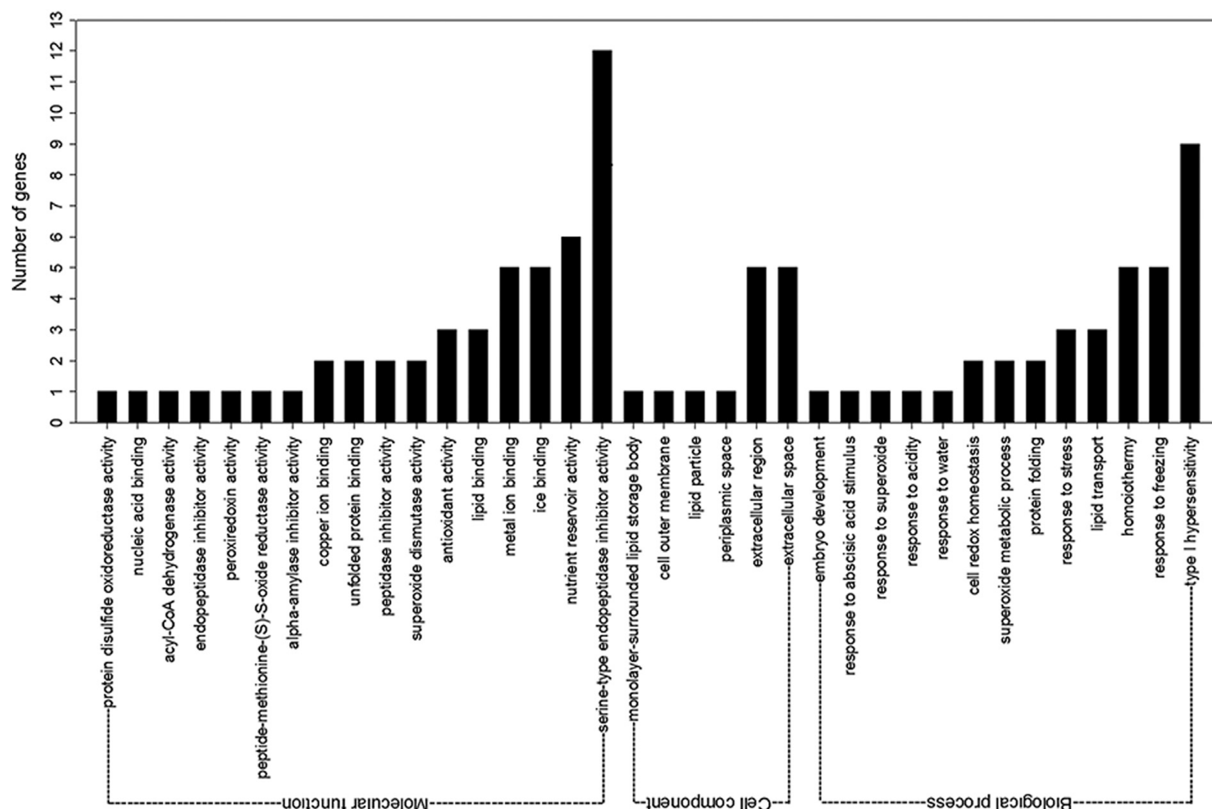


Fig. 4. GO classification and enrichment analysis of the identified proteins from 2-chloroethanol extract. The identified proteins were classified as being involved in13 kinds of biological processes, 5 kinds of cell components, and 17 kinds of molecular functions, which were significantly in accordance with hyper p value which is < 0.05. Of these, type I hypersensitivity, extracellular space and extracellular region, and serine-type endopeptidase inhibitor activity enrich the first largest class of proteins. GO represents gene ontology.

considerable fraction of proteins have been not functionally classified yet due to the absence of GO annotation. The largest number of un-annotated proteins, 32, is in the cell component type (Table 2S-2, 3), the smallest number, 18, is in the molecular function type (Table 3S-2, 3), and an intermediate number, 23, is in the biological process type (Table 4S-2, 3). The presence of un-annotated proteins might result in a certain bias of the entire GO enrichment analysis, but its validity was confirmed in the present work by statistically significant results of the analysis in ROAD.

Recent studies show that ethanol/water extracts of rice seeds have antifungal activity (Bretanha, Kupsk, Garda-Buffon, Badiale-Furlong, & Pagnussatt, 2013; Pagnussatt, Bretanha, Garda-Buffon, & Badiale-Furlong, 2011; Pagnussatt, Bretanha, Meza, Garda-Buffon, & Badiale-Furlong, 2013). This effect is due to the presence of enzyme inhibitors in rice seed extracts, in particular, amylase inhibitors which can effectively inhibit the activity of fungal α-amylase (Bretanha et al., 2013; Pagnussatt et al., 2011; Pagnussatt et al., 2013). The present study identified protease inhibitors, seed allergen proteins, and lipid transfer proteins in the 2-chloroethanol extract of rice seeds. NCBI blast results show that seed allergen proteins belong to a class of α-amylase inhibitors. In addition, several antimicrobial proteins, such as thaumatin family domain-containing protein, Cupin domain-containing protein, and AMBP1 (Antimicrobial peptide MBP-1 (maize basic peptide 1) family protein precursor) were also detected in the 2-chloroethanol extract. Together, the accumulated data provide valuable insights into the properties of ethanol extracts of rice seeds.

5. Conclusions

The study documented the differential profile of proteins extracted from rice seeds by 2-chloroethanol and ethanol. Subsequent proteomic analysis revealed that besides storage proteins, including prolamins and glutelins, the extracted proteins contained some allergens and protease inhibitors. These proteins may provide antimicrobial and insect resistance and their identification may help discover candidate genes for controlling plant disease and pest insects, contributing to the improvement of rice breeding.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2020.100002.

Declaration of competing interest

The authors have declared no conflict of interest.

Acknowledgments

We thank Prof. Xing-Zhong Xue (Zhejiang University) for critical reading of the manuscript. This work was supported by the Grants from the National Key Research and Development Program of China (No.2017YFD0200900), the Active Design Project of Hangzhou Agricultural and Social Development Scientific Research (No. 20172015A02), the Project of Hangzhou Great Science and Technology Innovation (20131812A02) and Science and Technology Innovation and Demonstration extension Fund of Hangzhou Academy of Agricultural Sciences (No. 2019HNCT-28).

References

Baxter, G., Blanchard, C., & Zhao, J. (2004). Effects of prolamin on textural and pasting properties of rice flour and starch. *Journal of Cereal Science, 40*, 205–211.

Bean, S. R., & Lookhart, G. L. (2000). Electrophoresis of cereal storage proteins. *Journal of Chromatography A, 881*, 23–36.

Bietz, J. A. (1983). Separation of cereal proteins by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A, 255*, 219–238.

Bretanha, C. C., Kupsk, I. L., Garda-Buffon, J., Badiale-Furlong, E., & Pagnussatt, F. A. (2013). Promising antifungal effect of rice (*Oryza sativa* L.), oat (*Avena sativa* L.) and wheat (*Triticum aestivum* L.) extracts. *Journal of Applied Biotechnology, 1*, 37–44.

Cheng, J., Wang, L., Zeng, P., He, Y., Zhou, R., Zhang, H., & Wang, Z. (2017). Identification of genes involved in rice seed priming in the early imbibition stage. *Plant Biology*, *19*, 61–69.

Cooke, R. J. (1984). The characterisation and identification of crop cultivars by electrophoresis. *Electrophoresis*, *5*, 59–72.

Deng, B., Lu, J., & Yan, J. (1998). Separation of alcohol-soluble proteins from tobacco seeds by capillary zone electrophoresis. *Journal of Chromatography A*, *803*, 321–327.

Deng, Z. Y., Gong, C. Y., & Wang, T. (2013). Use of proteomics to understand seed development in rice. *Proteomics*, *13*, 1784–1800.

Fei, Y. Y., Yang, J., Wang, F. Q., Fan, F. J., Li, W. Q., Wang, J., Xu, Y., Zhu, J. Y., & Zhong, W. G. (2019). Production of two elite glutinous rice varieties by editing *Wx* gene. *Rice Science*, *26*, 118–124.

Görg, A., Postel, W., Baumer, M., & Weiss, W. (1992). Two–dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: Discrimination of cultivars with different malting grades. *Electrophoresis*, *13*, 192–203.

Gorinstein, S., Nue, I. A., & Arruda, P. (1991). Alcohol-soluble and total proteins from Amaranth seeds and their comparison with other cereals. *Journal of Agricultural Food Chemistry*, *39*, 848–850.

Hofmann, K., & Stoffel, W. (1993). TM base—A database of membrane spanning proteins segments. *Biological Chemistry Hoppe-Seyler*, *74*, 166.

Kim, H. J., Lee, J. Y., Yoon, U. H., Lim, S. H., & Kim, Y. M. (2013). Effects of reduced prolamin on seed storage protein composition and the nutritional quality of rice. *International Journal of Molecular Science*, *14*, 17073–17084.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., Yates, JR 3rd. (2002). Proteomic survey of metabolic pathways in rice. PNAS,99:11969–11974.

Komatsu, S., Konishi, H., Shen, S., & Yang, G. (2003). Rice proteomics: A step toward functional analysis of the rice genome. *Molecular & Cellular Proteomics*, *2*, 2–10.

Lookhart, G. L., & Bean, S. R. (1995). Rapid differentiation of oat cultivars and of rice cultivars by capillary zone electrophoresis. *Cereal Chemistry*, *72*, 312–316.

Miernyk, J.A., Hajduch, M. (2011). Seed proteomics. Journal of Proteomics74(4):389–400.

Pagnussatt, F. A., Bretanha, C. C., Garda-Buffon, J., & Badiale-Furlong, E. (2011). Extraction of fungal amylase inhibitors from cereal using response surface methodology. *International Research Journal of Agricultural Science and Soil Science*, *1*, 428–434.

Pagnussatt, F. A., Bretanha, C. C., Meza, J., Garda-Buffon, S. L. R., & Badiale-Furlong, E. (2013). Activity of rice bran proteic extracts against *Fusarium graminearum*. *African Journal of Agricultural Research*, *8*, 6283–6290.

Paulis, J. W., & Bietz, J. A. (1986). Separation of alcohol-soluble maize proteins by reversed-phase high performance liquid chromatography. *Journal of Cereal Science*, *4*, 205–216.

Pernollet, J.C., Potier, B., Galle, A.M., Huet, J.C., Beauvais, F., Sallantin, M. (1989).2D-HPLC separation, electrophoretic characterization and *N*-terminal sequences of oat seed prolamins. Phytochemistry28:2565–2570.

Shewry, P.R., Halford, N.G. (2002). Cereal seed storage proteins: Structures, properties androle in grain utilization. Journal of Experimental Botany, Inorganic Nitrogen Assimilation Special Issue53:947–958.

Smith, J. S., & Smith, O. S. (1988). Variation in kafirin and alcohol-soluble glutelin chromatograms of sorghum inbred lines revealed by reversed-phase high-performance liquid chromatography. *Theoretical Applied Genetics*, *76*, 97–107.

Tsuji, K. (1991). High-performance capillary electrophoresis of proteins: Sodium dodecyl sulphate—Polyacrylamide gel-filled capillary column for the determination of recombinant biotechnology-derived proteins. *Journal of Chromatography A*, *550*, 823–830.

Walther, T. C., & Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. *Journal of Cell Biology*, *190*, 491–500.

Wang, W., Liu, S., Song, S., & Møller, I. M. (2015). Proteomics of seed development, desiccation tolerance, germination and vigor. *Plant Physiology and Biochemistry*, *86*, 1–15.

Wang, Y. L., Wang, L., Zhou, J. X., Hu, S. B., Chen, H. Z., Xiang, J., Zhang, Y. K., Zeng, Y. J., Shi, Q. H., Zhu, D. F., & Zhang, Y. P. (2019). Research progress on heat stress of rice at flowering stage. *Rice Science*, *26*, 1–10.

Weiss, W., Postel, W., & Görg, A. (1991). Barley cultivar discrimination: II. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis and isoelectric focusing with immobilized pH gradients. *Electrophoresis*, *12*, 330–337.

Yamagata, H., & Tanaka, K. (1986). The site of synthesis and accumulation of rice storage proteins. *Plant Cell Physiology*, *27*, 135–145.

Yang, P., Li, X., Wang, X., Chen, H., Chen, F., & Shen, S. (2007). Proteomic analysis of rice (Oryza sativa) seeds during germination. *Proteomics*, *7*, 3358–3368.

Yang, Y., Dai, L., Xia, H., Zhu, K., Liu, H., & Chen, K. (2013). Protein profile of rice (*Oryza sativa*) seeds. *Genetics and Molecular Biology*, *36*, 87–92.