

Genome-wide Identification and Molecular Characterization of Ole_e_I, Allerg_1 and Allerg_2 Domain-containing Pollen-Allergen-like Genes in *Oryza sativa*

Shu-Ye JIANG, Phua Xiao Hua JASMIN, Yeo Yun TING, and Srinivasan RAMACHANDRAN*

Rice Functional Genomics Group, Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604

(Received 18 February 2005; revised 9 May 2005)

Abstract

Pollen allergens play important roles in plant development in addition to their allergenic nature for human. More than 10 groups of pollen allergens have been reported. Among them, Pollen_Ole_e_I (Ole), Pollen_allerg_1 (Allerg1) and Pollen_allerg_2 (Allerg2) domain-containing proteins are the majority of allergens. We have identified 114 pollen-allergen-like genes in rice genome by bioinformatics using public databases. Among them, 45 genes encode Ole domain-containing proteins, 62 with Allerg1 and 7 with Allerg2. They are distributed on 11 of 12 rice chromosomes excluding chromosome 11. Comparison analysis of coding regions from both predicted genes and isolated full-length cDNAs showed that most of predicted genes were correct in the splicing of exons and introns, and only 7 exhibited wrong predictions. The fact suggested the applicability of the prediction programs to identify pollen-allergen genes. Phylogenetic analysis revealed the high diversity within *OsOle* genes and recent evolutionary event in *OsAllerg1* genes, and suggested that some of *OsOle* genes were new members of the family. Expression analysis by RT-PCR showed that most of the genes were expressed in all tested tissues and only eight genes exhibited panicle-specific expression, suggesting that pollen-allergen genes play roles in not only productive but also vegetative development.

Key words: *Oryza sativa*; Pollen_Ole_e_I; Pollen_allerg_1; Pollen_allerg_2; Bioinformatics

1. Introduction

Pollen allergens are specific substances, which cause hypersensitivity reactions of the immune system called allergies.¹ They are proteins with molecular weights from 8 to 50 kD, exhibiting features of solubility and stability. These proteins are responsible for the IgE-mediated allergies in human. To date, more than 10 groups of pollen allergens have been reported.² Among all groups of pollen allergens, Pollen_Ole_e_I (Ole) domain-containing proteins are the major allergens.³ On the other hand, based on the prevalence of IgE antibody recognition among grass pollen-sensitized individuals, members of two groups, groups 1 and 5, have been shown to dominate the immune response to grass pollen extract.² The group 1 is described as Pollen_allerg_1 (Allerg1)

domain-containing proteins and the group 5 allergens contain the domain Pollen_allerg_2 (Allerg2). Therefore, in this study, we will focus on these three groups of pollen allergens.

Ole e 1 was the first purified allergen from *Olea europaea*⁴ and named as Ole e 1 according to the IUIS nomenclature.⁵ The protein consists of a single polypeptide chain of 145 amino acid residues,¹ similar to the polypeptide encoded by *LAT52* gene from tomato and *Zmc13* gene from maize pollens.³ The Ole domain contains six conserved cysteines which may be involved in disulfide bonds (<http://au.expasy.org/>).¹ The domain also contains the consensus sequences E/Q/T-G-X-V-Y-C-D-T/N/P-C-R (X, any residues) (<http://au.expasy.org/>). Some of Ole-containing proteins are proline-rich, encoding extensin-like proteins.⁶ To date, ten Ole domain-containing genes have been isolated and characterized from the olive pollens.⁷ However, little is known about Ole-containing proteins in rice and *Arabidopsis*.

* Communicated by Satoshi Tabata

To whom correspondence should be addressed. Tel. +65-6872-7480; Fax. +65-6872-7007, E-mail: sri@tll.org.sg

Allerg1-containing genes were considered to encode grass group 1 pollen allergens that caused human allergy responses.⁸ All group 1 allergens examined to date have an N-glycosylation site at amino acid position 9 and carry glycans corresponding to 4–5% of their molecular weight.² They contain seven conserved cysteine residues, located mainly in the N-terminal half of the proteins (<http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01357>).² Now Allerg1-containing genes from at least 10 species have been cloned including one gene from rice.^{2,9} However, no report showed genome-wide identification in rice.

The BLAST search of GenBank using the first expansin cDNA sequence revealed the similarity of the sequences to that encoding the group 1 allergens.⁸ Subsequently, Cosgrove et al.¹⁰ reported that the group 1 allergen from maize (*Zea mays*) pollens, called Zea m 1, has wall-loosening activity characteristic of expansins and proposed that these proteins loosen the cell walls of the stigma and style to aid pollen tube penetration. Grass group 1 pollen allergens are now recognized as a subgroup of β -expansins.¹¹ Today, expansins are regarded as proteins that mediate cell wall extension in plants including two families, named α and β expansins (<http://www.bio.psu.edu/expansins/>).^{8,10–12} Although group 1 allergens were first identified in pollens, β -expansin genes are expressed more widely during the growth and development of vegetative tissues.^{13–15}

Allergen2 domain-containing genes encoded grass pollen proteins of group 5 with ribonuclease activity (<http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01620>).^{16,17} The allergens in this family were identified as Lol p 5, Pha a 5, Phl p 5, Phl p 6, Phl p 11 and Poa p 9 with predicted molecular weights ranging from 28.3 to 37.8kD.¹⁸ Compared to group 1 allergens, no N-linked glycosylation has been reported to occur in this group, but they contain several hydroxylated proline residues.^{2,19} Alanine constitutes up to 30% of the amino acid content while the representation of methionine, cysteine, histidine and tryptophan is unusually low.² Although this group of allergens from at least seven species have been cloned,² little is known about this group of genes and their proteins in rice and *Arabidopsis*.

Although the expression of some of Allerg1-containing genes was detected in pollens and other tissues, previous reports showed that Ole- and Allerg2-containing genes were pollen specific. Both *Zmc13* in maize and *LAT52* in tomato were Ole domain-containing genes with pollen-specific expression manner, playing an important role in pollen hydration and/or pollen germination.^{20,21} In rice, *PS1* gene also encoded a protein with Ole domain, showing a pollen-specific expression pattern.²² In addition to this, three Allerg2-containing genes from kentucky bluegrass were also confined to pollen tissues.²³ Therefore, a question raises: do all Ole- or Allerg2-containing genes show pollen-specific expression patterns?

In this report, we have identified 114 pollen-allergen-like genes (referred to pollen-allergen genes hereafter for convenience) in rice genome by bioinformatics using information available in public databases, some of which might be new members of these families. We also demonstrated the applicability of the prediction programs to identify pollen-allergen genes and analyze their expression pattern by RT-PCR.

2. Materials and Methods

2.1. Plant materials and their growth conditions

Japonica rice (*Oryza sativa* cv. Nipponbare) was used for all RNA extractions. The total RNAs were extracted from the following tissues; young leaves from 3-week-old plants, mature leaves from flowering plants, young panicles from non-flowering plants, mature panicles from flowering plants, young and mature roots from 10- to 20-days-old seedlings grown on MS media at 25°C under 16 h light/8 h dark conditions.

2.2. Databases search for pollen-allergen genes

We have used several ways to search and predict pollen-allergen genes. The key pollen-allergen genes were obtained from the protein families database (Pfam, <http://www.sanger.ac.uk/Software/Pfam/>). The representative sequences were selected for BLAST searches based on the phylogenetic analysis. The GenBank accession numbers of the amino acid sequences used for BLAST searches were as follows: CAA02039, Q03211, AAC69131, P33050, CAB78861, O49813, P13447, P19963 and AAA98492 for Ole-containing genes; BAA05476, AAF72986, AAB61710, Q41576, O81999, AAC96078, AAF17571, AAF62183, Q9FL80 and Q9LN94 for Allerg1-containing genes; AAB41585, CAA76556, JT0756, C39098, AAB35985, AAB35986 and AAB35987 for Allerg2-containing genes. The domain amino acid sequences of Ole, Allerg1 and Allerg2 were used to identify corresponding pollen-allergen genes with a high similarity in rice genome using TBLASTN and BLASTP program in RGP (Rice Genome research Program, <http://rgp.dna.affrc.go.jp/>), TIGR (<http://tigrblast.tigr.org/>), Gramene (<http://www.gramene.org/>), Oryzabase (<http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>), NCBI (<http://www.ncbi.nlm.nih.gov>), and DDBJ (<http://www.ddbj.nig.ac.jp>) databases. All predicted pollen-allergen genes were used for similarity searches again in order to confirm predicted genes and to detect new candidates. Alternatively, databases were searched using ‘ole, allerg1 or allerg2’ as keywords to achieve more pollen-allergen genes. The databases used for this approach are as follows: RiceGAAS (<http://ricegaas.dna.affrc.go.jp/rgadb/>), Oryzabase, Gramene and NCBI databases. Several rice

annotation programs such as RiceHMM program (<http://rgp.dna.affrc.go.jp>), RiceGAAS (<http://ricegaas.dna.affrc.go.jp>) and the New GENSCAN Web Server at MIT (<http://genes.mit.edu/GENSCAN.html>) were used for predicting full-length pollen-allergen coding regions and their proteins. The rice full-length cDNA database KOME (Knowledge-based *Oryza* Molecular biological Encyclopedia; <http://cdna01.dna.affrc.go.jp/cDNA/>)²⁴ was used for searching full-length cDNAs of all predicted pollen-allergen genes. The comparison of coding regions between predicted cDNAs and KOME full-length cDNAs was carried out using the programs 'BLAST 2 SEQUENCES' in NCBI database and DNASTAR. In addition, the *Arabidopsis* pollen-allergen genes were obtained by searching TIGR and TAIR (<http://www.arabidopsis.org/>) databases using the program BLASTP and keyword searches.

2.3. Identification of motifs in the predicted pollen-allergen genes

In order to confirm the presence of conserved domains in the predicted genes and use their sequences for phylogenetic analysis, motifs in the predicted pollen-allergen genes were identified using Pfam program (<http://www.sanger.ac.uk/>). BLASTP in NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) and InterPro (<http://www.ebi.ac.uk/interpro/>) databases were also employed to detect conserved domains.

2.4. Sequence alignment and phylogenetic analysis

Preliminary sequence manipulations were performed using DNASTAR program. The sequence alignment was generated using ClustalX (Version 1.8).²⁵ The aligned amino acid sequences formed the basis for the phylogenetic analysis. Maximum-parsimony (MP) and maximum-likelihood (ML) analyses were performed using the program Mac PAUP 4.0b8 (ppc) (<http://www.paup.csit.fsu.edu>). Heuristic search option with MILPARS, TBR (tree-bisection-reconnection) branch swapping, ACCTRAN (accelerated transformation) optimization, and random addition with 100 replicates was used to find the best tree. Bootstrap analysis was used to investigate tree support of MP and ML trees. Bootstrap supports (BS) of specific nodes²⁶ were estimated with 500 replicates for ML analysis and 1000 for MP under default options as implemented in PAUP program. Nodes with BS values >70% were considered supported with 95% probability.²⁷

2.5. Primer selection and RT-PCR

All primers used in this study were designed by Primer Selection in DNASTAR program and then submitted to NCBI database for BLAST search. All non-specific primers shown by the search results were eliminated instead of re-designed primers with specific binding to the corresponding genes. Thus, the results from RT-PCR analysis

might represent the expression pattern of pollen-allergen genes.

Total RNAs from six different tissues (young and mature leaves, panicles and roots) were used as templates. The RNAs were isolated using Qiagen RNeasy Mini Kit. In order to eliminate the possibility of genomic DNA contamination in PCR amplification, total RNAs were treated by RNase-free DNase. RT-PCRs were carried out using Qiagen One-Step RT-PCR kit. PCR amplification were performed in 25 μ l reaction mixtures with 200 ng of total RNA, 400 μ M of each dNTPs, 0.6 μ M of each primer, 5 μ l of Q solution, 4 units of RNase inhibitor and 5 μ l of buffer provided by the supplier. The reactions were performed in PTC-100 thermo-cyclers. The program used for RT-PCR was 50°C for 30 min and 95°C for 15 min followed by 35 cycles at 94°C for 1 min, 53–65°C for 1 min (depending on the T_m value of primers) and 72°C for 1 min; followed by a 10 min extension step at 72°C and 10°C forever to terminate.

3. Results

3.1. Identification of pollen-allergen genes using public sequence databases

More than 230 sequences encoding three types of pollen allergens and/or their homologs were identified by searching all public databases as described in the Materials and Methods. Some of the sequences were redundant as the sequence of a pollen allergen or its homolog overlapped between two YAC or BAC clones. Others showed no homology to three types of domains namely Ole, Allerg1 and Allerg2 by further BLASTP search. After filtering these sequences, 114 pollen-allergen genes were predicted. Among them, 45 genes were identified which encoded Ole domain-containing proteins with 350–6188 bp long in genome sequence. They were named as *Oryza sativa* Ole 1–45 (*OsOle1–45*). Other 62 predicted genes encode Allerg1 domain-containing proteins, and their genomic region ranged from 353 to 6919 bp long named as *Oryza sativa Allerg1-1* to *49* (*OsAllerg1-1* to *62*). The remaining seven members were genes encoding Allerg2-containing proteins and were from 809 to 5164 bp long which were named as *Oryza sativa Allerg1-1* to *7* (*OsAllerg1-1* to *7*). Supplemental Table 1a–c (www.dnaresearch.oxfordjournals.org) lists all the pollen-allergen genes predicted, their locations on rice chromosomes, BAC, YAC clones with accession numbers, genetic map positions and molecular weights of the proteins they encode. Based on this analysis, we concluded that the rice genome encodes at least 114 putative pollen-allergen genes.

3.2. Phylogenetic analysis

In order to classify the rice pollen-allergen genes predicted in this study, we have constructed phylogenetic trees based on the amino acid sequences of the known

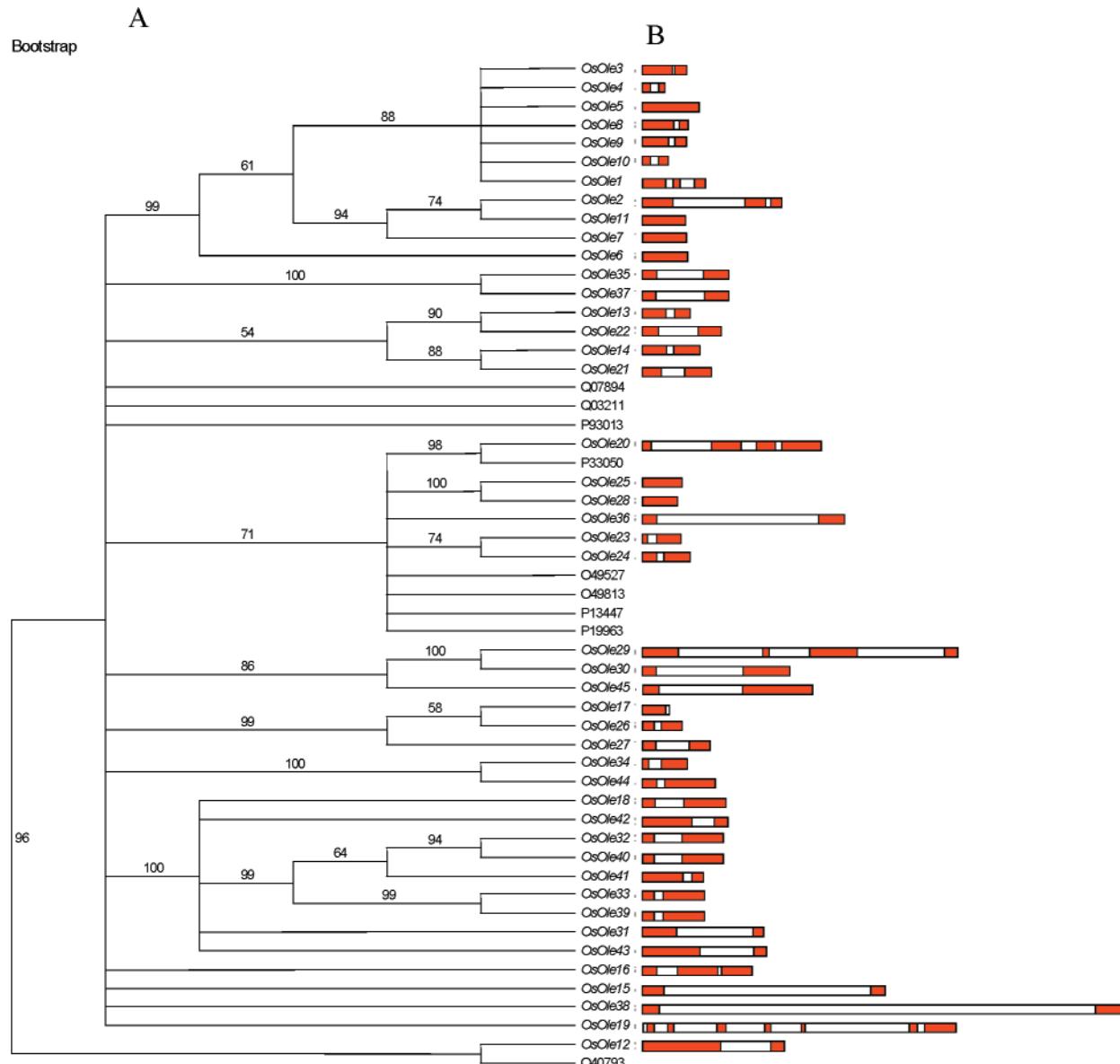


Figure 1. Phylogenetic tree and structure of exons and introns of *OsOle* genes. A, Phylogenetic tree using the bootstrap method with a heuristic search of the PAUP 4.0b8 program. Besides the rice amino acid sequences of Ole domains, the 9 key sequences from other plants were also used and were listed by protein accession numbers. B, Exons and intron structure of rice *OsOle* genes. Red boxes indicate exons and white boxes indicate introns.

Ole, Allerg1 and Allerg2 domains. All domain sequences were obtained by searching databases as mentioned in the Materials and Methods.

Using the Pfam database (<http://www.sanger.ac.uk/>), 9 key Ole-containing genes were identified. Phylogenetic tree was constructed by using these key sequences and all predicted Ole domain sequences (Fig. 1A). Analysis of the tree showed that six genes (*OsOle20*, *23–25*, *28* and *36*) were grouped together with five key Ole-containing genes from other plants with accession numbers P33050, O49527, O49813, P13447 and P19963 (Fig. 1A). The *OsOle12* was in the same cluster of the gene with accession number Q40793. In addition to this, we could not find any

member of rice genes classified with the other three non-rice Ole-containing genes: Q07894, Q03211 and P93013. On the contrary, most of the *OsOle* genes also could not be clustered into any group of non-rice Ole-containing genes, indicating the prediction of new members of Ole-containing genes in rice.

The phylogenetic analysis of the 64 sequences with Allerg1 domain in the Pfam database revealed 10 key Allerg1 domains in plants (data not shown). One member from each group was selected randomly for the tree construction with 62 rice Allerg1 domains to classify into α -, β -expansin or expansin-like members (Fig. 2A). The tree analysis showed that rice *Allerg1* genes could be divided

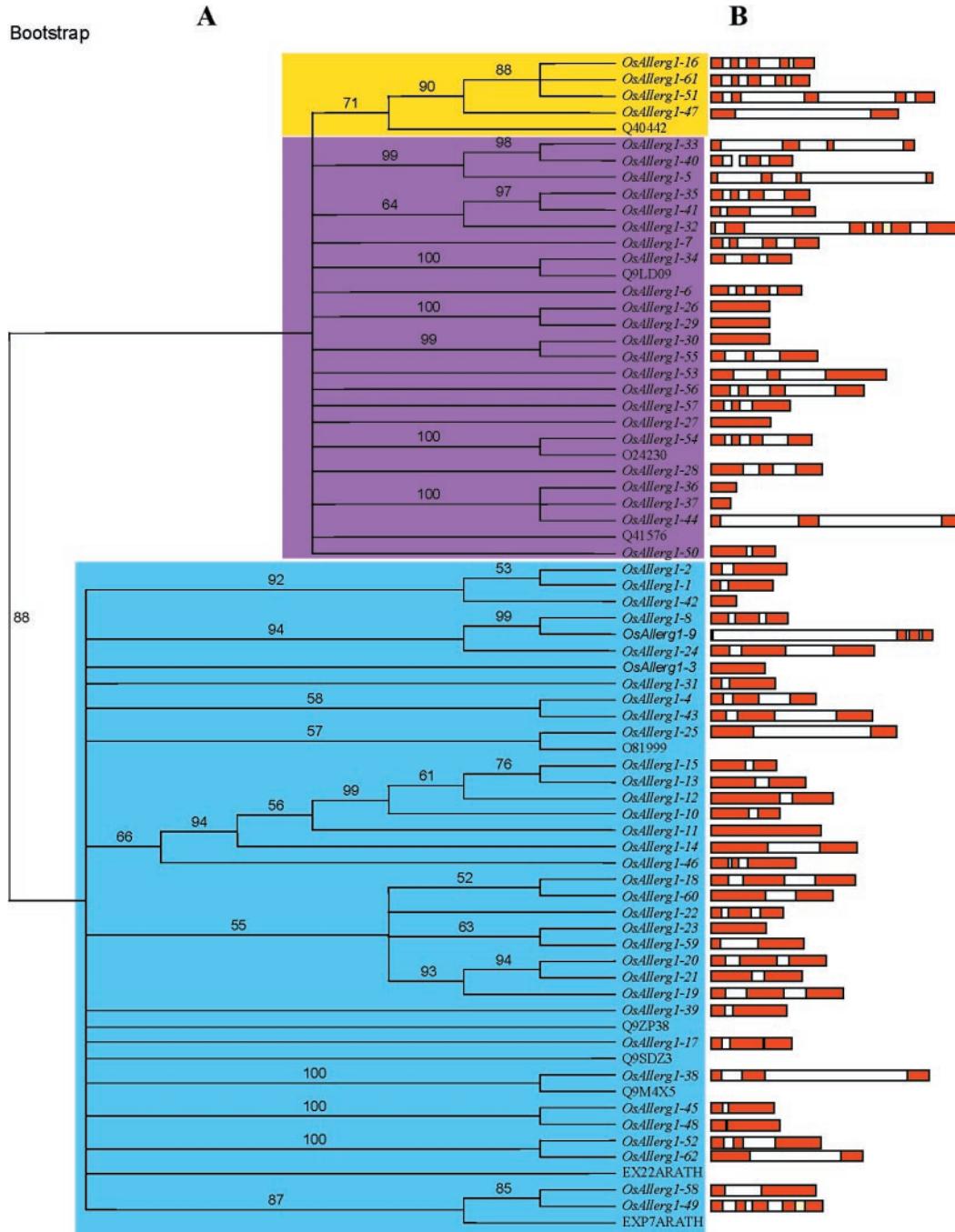


Figure 2. Phylogenetic tree and structures of exons and introns of *OsAllerg1* genes. A, phylogenetic tree was constructed based on the Allerg1 domain sequences using the bootstrap method with a heuristic search of the PAUP 4.0b8 program. Besides the rice amino acid sequences of Allerg1 domains, 10 key sequences from other plants were also used and were listed by protein accession numbers. Yellow shade indicates expansin-like expansins, pink shade β-expansins, and blue shade α-expansins. B, Exons and intron structure of rice *OsAllerg1* genes. Red boxes indicate exons and white boxes indicate introns.

into three subgroups on the basis of domain sequence similarity. The largest subgroup is classed into α-expansins, with 35 members as shown in Fig. 2A with blue shade. The second subgroup is β-expansins, containing 23 members as shown in Fig. 2A with pink shade. The smallest subgroup is expansin-like expansins including four members (Fig. 2A with yellow shade).

3.3. Characterization of putative homologs of pollen-allergen genes in rice

The structural analysis of exons and introns based on the prediction from RiceGAAS revealed that 33 of 45 *OsOle* genes consisted of two exons, six with 3–8 exons and the other six members with no intron (Fig. 1B). On the other hand, among 62 *OsAllerg1* genes 20 possessed

Table 1. Presence of other domains in predicted Ole- and Allerg1-containing genes.

Domain	InterPro ID	Hypothetical function	No. of <i>OsOle</i> genes	No. of <i>OsAllerg1</i> genes
Auxin_inducible	IPR003676	Auxin responsive protein	1	0
Barwin	IPR001153	Protein for defense system	0	1
cytadhesin_P20	IPR009896	Membrane associated protein, important in cytadherence and virulence	1	0
cytadhesin_P30	IPR009896	Membrane associated protein, important in cytadherence and virulence	1	0
DPBB_1	IPR005132	Rare lipoprotein A (RlpA)-like double-psi beta-barrel	0	41
DUF1210	IPR010616	Proline-rich plant protein of unknown function	13	0
Extensin_2	IPR006706	Structural constituent of cell wall	4	0
FYVE	IPR000306	Zinc ion binding	1	0
Glyco_Hydro_45	IPR000334	Cellulase activity	0	4
GRP	IPR010800	Glycine rich protein	0	13
Hydrophobin	IPR001338	Fungal hydrophobin, structural constituent of cell wall	0	1
Lamprin	IPR009437	The major connective tissue component of the fibrillar extracellular matrix	0	1
NapB	IPR005591	The small subunit of a heterodimeric periplasmic nitrate reductase	1	0
Penaeidin	IPR009226	Chitin binding	2	0
PKD	IPR000601	Involving in adhesive protein-protein and protein-carbohydrate interactions	0	1
Ribosomal_S5_C	IPR005324	Structural constituent of ribosome	1	0
rnaseA	IPR001427	Pancreatic ribonuclease activity	0	1
zf_CCCH	IPR000571	Nucleic acid binding	1	0

two exons, 19 had three exons and the 13 consisted of 4–6 exons and 10 had no introns (Fig. 2B). However, the structures of seven Allerg2 genes were highly variable with 1–10 exons (data not shown). Besides three pollen-allergen domains, other domains were also present in some pollen-allergen genes. Table 1 lists 18 other domains detected in Ole and Allerg1 domain-containing genes. For *OsOle* genes, 13 members were detected with the DUF1210 domain, which presents in proline-rich plant proteins. In *OsAllerg1* genes, the domain DPBB_1 was found in 32 members of the predicted genes and other 10 members with the GRP domain detected in glycine-rich proteins.

3.4. Chromosomal distribution

In order to analyze the chromosomal distribution of the predicted pollen-allergen sequences, we performed database searches to localize the map position of corresponding YAC or BAC clones using the data from RGP (Rice Genome Research Program) website (<http://rgp.dna.affrc.go.jp/>). Figure 3A shows the distribution of the *OsOle1–45* genes on different chromosomes in the rice genome. All of predicted *OsOle* sequences were distributed within 10 chromosomes namely chromosome 1–7, 9, 10 and 12, and no homologs were located on chromosomes 8 and 11. Although 10 chromosomes had homologs, each one differed in the number of *OsOle* genes it contained; chromosomes 6 and 12 contained one *OsOle* gene each; chromosomes 2, 3, 4 and 9

contained two each; chromosomes 5 and 7 had three each; chromosome 1 had 5 and chromosome 10 had 14 *OsOle* genes. The *OsOle1–11*, *16–18* and *38–44* were located on the short arms of the rice chromosomes and the rest were located on the long arms. Moreover, *OsOle23* and *OsOle36* were mapped near the centromere and *OsOle 26, 27* and *28* were mapped on the end of the long arms (<4 cM to the end).

Figure 3B shows the chromosomal distributions of Allerg1-containing genes. Similar to *OsOle* genes, all the Allerg1-containing homologs were distributed within 10 chromosomes. They were mapped on chromosome 1–7, and chromosome 8 instead of 9 in *OsOle* genes as well as 10 and 12, no homologs were located on chromosomes 9 and 11. Chromosomes 8 and 12 contained one homolog each; chromosome 7 had two; chromosome 5 had three and chromosomes 1, 3, 4, 6 and 10 had four or more homologs each. Twenty-two genes were located on the short arms of rice chromosomes, which were *OsAllerg1-1–3, 10–21, 26–30, 39* and *49*, and the rest were located on the long arms. Three genes *OsAllerg1-46, 47* and *52* were mapped on the long end of the long arms and seven other sequences *OsAllerg1-23, 36, 37, 39* and *41–43* were mapped near the centromere.

On the other hand, although only seven members of *OsAllerg2* genes were detected in rice genome, they were distributed on six different chromosomes namely chromosome 1–4, 7 and 10 as shown in Supplemental Table 1c (www.dnaresearch.oxfordjournals.org). Only chromosome 4 contained two members.

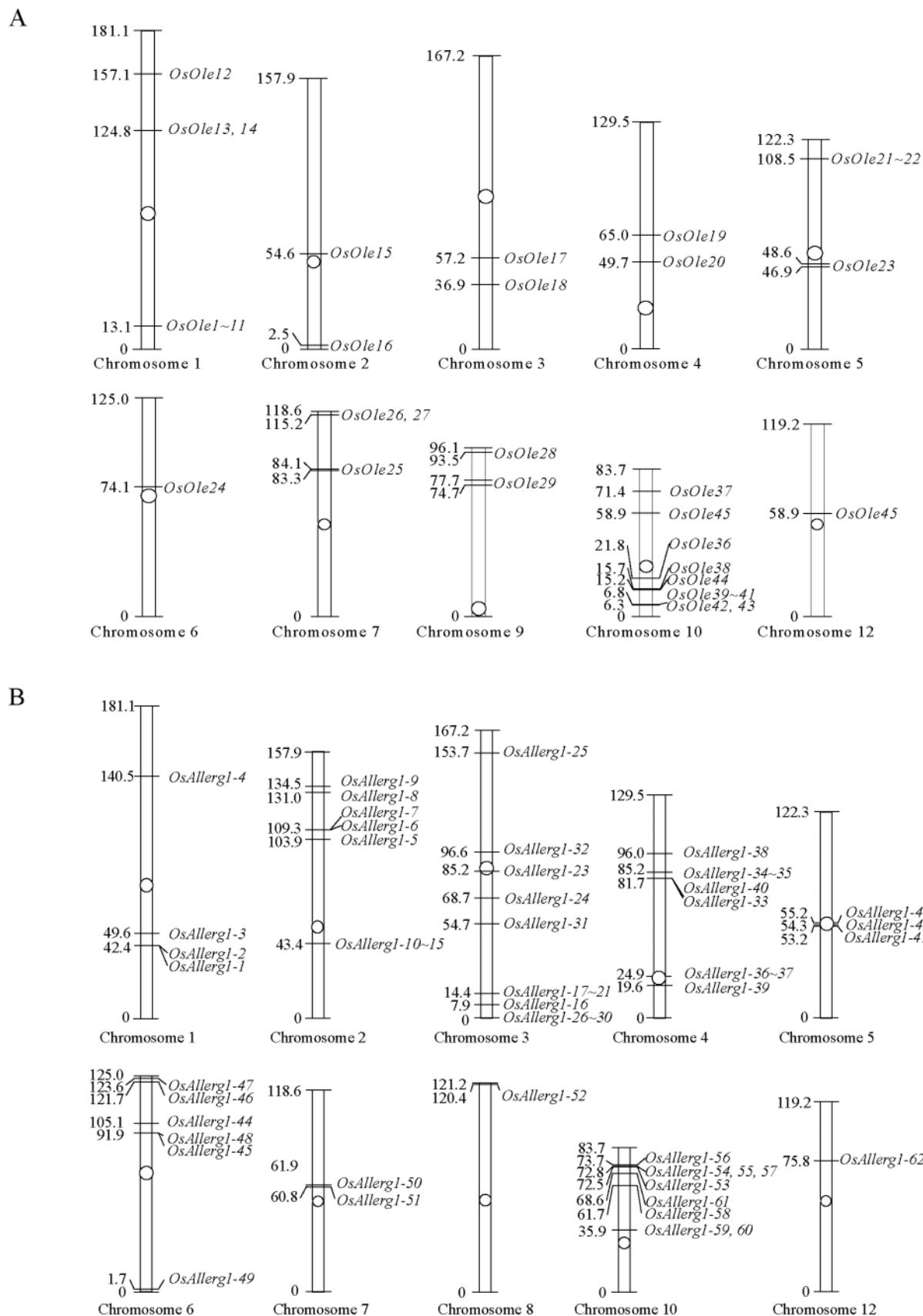


Figure 3. Chromosomal mapping of pollen-allergen genes predicted from the completed rice genome sequence. A, Map of *OsOle* genes. B, Map of *OsAllerg1* genes. Gene names were marked on the left side of chromosomes and genetic positions were on the right of chromosomes.

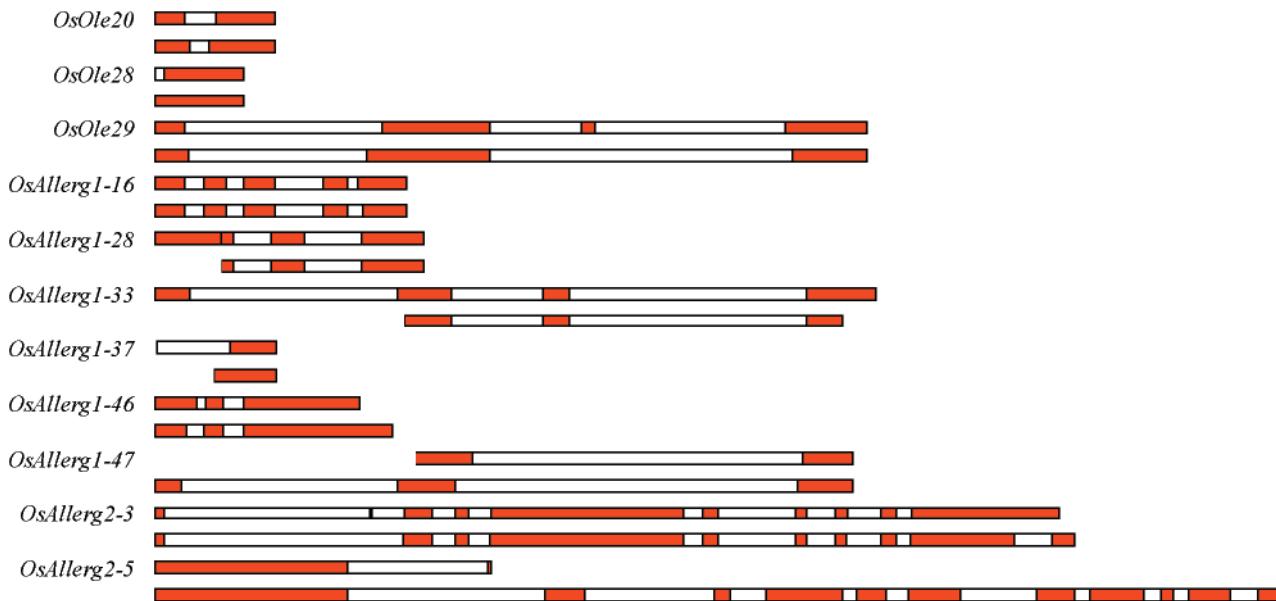


Figure 4. Comparison between both 11 predicted cDNAs and their RIKEN full-length cDNAs. Top row, predicted cDNA; bottom row, cDNA from RIKEN full-length cDNA database for each gene.

3.5. Comparison of predicted genes and isolated full-length cDNAs

In order to validate the gene prediction in the organization of exons and introns, we obtained predicted cDNAs sequences from public databases that are mentioned in Materials and Methods and their corresponding RIKEN full-length cDNA sequences, from KOME database.²⁴ These two sequences were used for BLAST search using 'BLAST 2 SEQUENCES' to detect the differences between two cDNA coding regions. The results showed that most of predicted cDNAs exhibited the same exon and intron structure as shown in RIKEN cDNAs. However, we still found 11 pollen-allergen genes with different splicing of exon and intron between predicted and RIKEN cDNAs (Fig. 4). Among them, three from Ole-containing genes (*OsOle20*, *OsOle28* and *OsOle29*), six from Allerg1-containing genes (*OsAllerg1-16*, *28*, *33*, *37*, *46* and *47*) and two from Allerg2 genes (*OsAllerg2-3* and *OsAllerg2-5*). Five genes exhibited the differences only in the length of exons or introns, which were *OsOle20*, *28*, *OsAllerg1-16*, *28* and *46*. The other six genes showed differences not only in the length of exon and intron but also in the numbers of exon and intron. In some cases, predicted genes contained one or more exons (*OsOle38*, *OsAllerg1-33*, *37* and *OsAllerg2-3*); and sometimes showed lesser number of exons (*OsAllerg1-47* and *OsAlerg2-5*) (Fig. 4).

3.6. Expression pattern of pollen-allergen genes

In order to examine whether these predicted pollen-allergen genes are expressed in rice and to study their expression patterns, the total RNAs isolated from various tissues were subjected to reverse transcription PCR (RT-PCR) analysis (Materials and Methods). To avoid

contamination from genomic DNA, total RNAs were treated with RNase-free DNase prior to RT-PCR. Primers were designed based on coding regions of pollen-allergen genes and are shown in Supplemental Table 2 (www.dnaresearch.oxfordjournals.org). The RT-PCR results showed that not all predicted genes were expressed in tissues tested under normal growth conditions. Thirteen (*OsOle1*, *2*, *6*, *11*, *19*, *22*, *26*, *30*, *35*, *36*, *42*, *44* and *45*) of 45 *OsOle* genes could not be detected even if we modified the RT-PCR conditions. Among the other 32 *OsOle* genes, 12 members showed expression in all six tissues tested: young and mature leaves, roots and panicles with similar expression intensity (Fig. 5A). These genes were as follows: *OsOle3-5*, *9*, *10*, *14*, *15*, *20*, *25*, *28*, *31* and *37*. The other 20 genes exhibited leaf, root or panicle-specific expression pattern, among which *OsOle29*, *31*, *33* and *34* showed high expression in the panicles but were weakly expressed in leaves and roots, and only six genes (*OsOle16*, *18*, *32*, *39*, *40* and *41*) were panicle-specific, might be putative pollen-specific allergen genes.

Among 62 *OsAllerg1* genes, only 49 pairs of primers were designed for detecting the expression pattern of corresponding genes because the other 13 genes shared high homologous sequences with similar primer binding sites. The RT-PCR results showed that 39 *OsAllerg1* genes could be detected among 49 RT-PCR reactions (Fig. 5B). We could not detect any signal in the other 10 genes despite modifying RT-PCR conditions. The analysis showed that 16 among 39 *OsAllerg1* genes were expressed in all tested tissues. These genes were *OsAllerg1-3*, *4*, *6*, *7*, *16*, *25*, *26*, *29*, *36*, *38*, *42*, *49*, *52*, *54*, *58* and *61*, including six members of β -expansin genes, eight α -expansins and two expansin-like genes. Only two members showed panicle-specific expression, which were

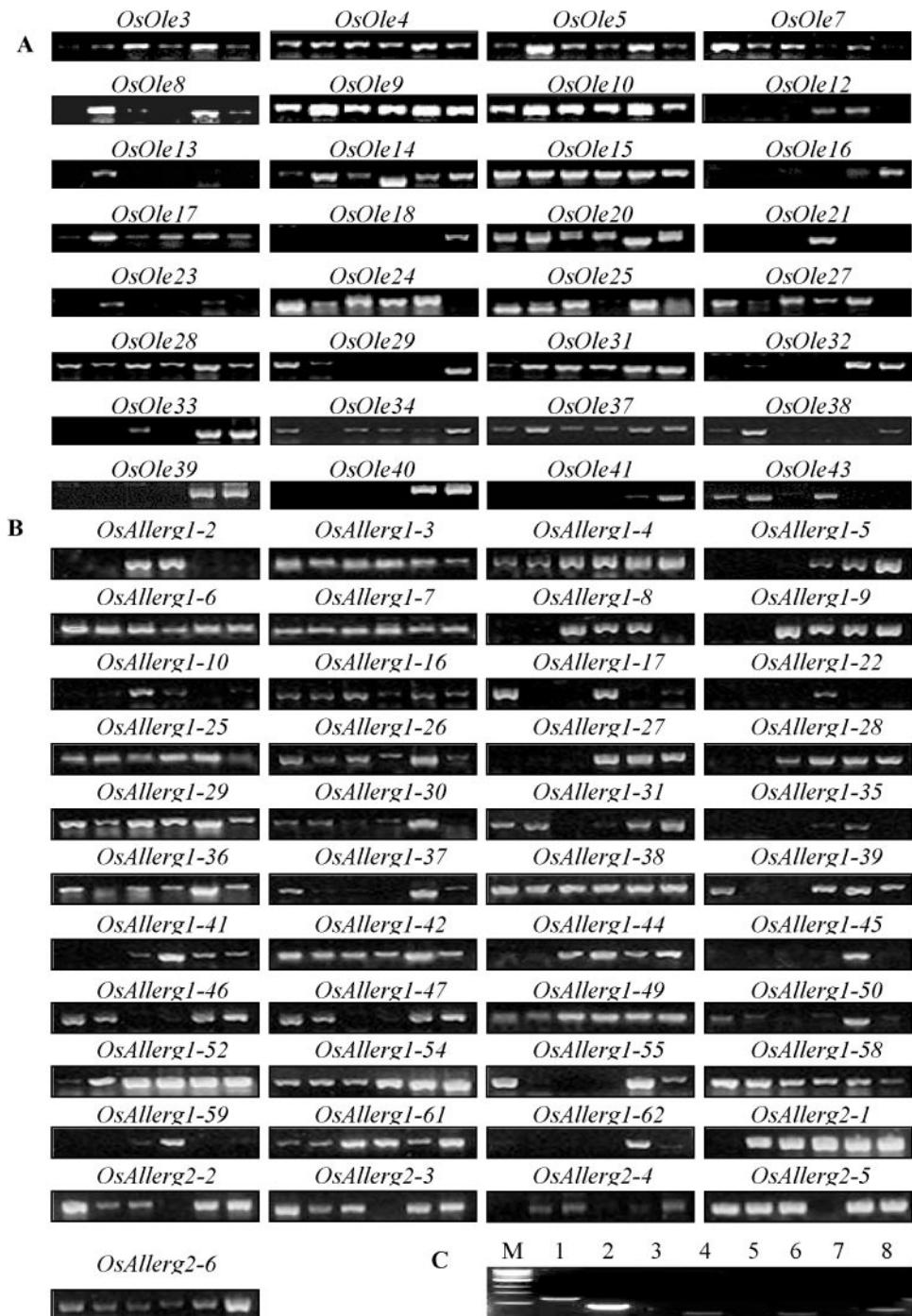


Figure 5. RT-PCR analysis of pollen-allergen genes. Total RNAs from young and mature leaves, young and mature roots and young and mature panicles were used for templates. A, *OsOle* genes; B, *OsAllerg1* and *OsAllerg2* genes. For each predicted gene, bands from left to right represented products from mature leaves, young leaves, mature roots, young roots, mature panicles and young panicles, respectively. C, the expression analysis of panicle-specific genes in mature anthers. M, 1 Kb DNA molecular marker; 1-8, products from *OsOle16*, *OsOle18*, *OsOle32*, *OsOle39*, *OsOle40*, *OsOle41*, *OsAllerg1-34*, *OsAllerg1-49*, respectively.

α -expansin genes *OsAllerg1-45* and *62*. Other two α -expansin genes exhibited root-specific expression including *OsAllerg1-2* and *59*. The remaining 19 genes either showed difference in expression intensity or differentially expressed in one or two among six tested tissues. For example, *OsAllerg1-26* showed the highest expression

in mature panicles and *OsAllerg1-52* showed the lowest expression in mature leaves; four genes (*OsAllerg1-9*, *28*, *41* and *44*) showed no expression only in the leaves.

As for seven predicted *OsAllerg2* genes, six genes showed expression. Among them, three genes (*OsAllerg2-2*, *3* and *5*) showed expression in all tested

Table 2. Pollen-allergen genes in *Oryza sativa* and *Arabidopsis*.

Chromosome	<i>OsOle</i> genes		<i>OsAllerg1</i> genes		<i>OsAllerg2</i> genes	
	<i>O. sativa</i>	<i>Arabidopsis</i> ^a	<i>O. sativa</i>	<i>Arabidopsis</i> ^b	<i>O. sativa</i>	<i>Arabidopsis</i>
1	14	4	4	6	1	Not yet identified
2	2	4	11	7	1	
3	2	1	17	8	1	
4	2	3	8	5	2	
5	3	5	3	9	0	
6	1	-	6	-	0	
7	3	-	2	-	1	
8	0	-	1	-	0	
9	3	-	0	-	0	
10	14	-	9	-	1	
11	0	-	0	-	0	
12	1	-	1	-	0	
Total	45	17	62	35	7	

^a Accession numbers listed as follows: At1g28290, At1g29140, At1g54970, At1g78040, At2g16630, At2g33790, At2g34700, At2g47540, At3g09925, At4g02270, At4g08685, At4g18596, At5g05500, At5g10130, At5g15780, At5g45880, At5g54855;

^b At1g12560, At1g20190, At1g26770, At1g62980, At1g65680, At1g69530, At2g03090, At2g20750, At2g28950, At2g37640, At2g39700, At2g40610, At2g45110, At3g03220, At3g15370, At3g29030, At3g29365, At3g45960, At3g45970, At3g55500, At3g60570, At4g01630, At4g17030, At4g28250, At4g38210, At4g38400, At5g02260, At5g05290, At5g39260, At5g39270, At5g39280, At5g39290, At5g39300, At5g39310, At5g56320.

tissues except in young roots. The *OsAllerg2-1* gene exhibited no expression only in mature leaves. No expression was detected in both mature leaves and young roots, and weak expression in mature panicles for *OsAllerg2-4*. On the other hand, *OsAllerg2-6* showed difference in expression intensity with the strongest expression in young panicles.

Since eight genes showed panicle-specific expression, the total RNA samples from mature anthers were isolated and then used for RT-PCR analysis. The results showed that all panicle-specific genes were expressed in mature anthers with the highest expression for the gene *OsOle18*, the best candidate of pollen-specific genes (Fig. 5C).

4. Discussion

Studies on pollen-specific genes would contribute to understanding the mechanism of male gametophyte development. Since both pollen-specific genes *PS1* in rice and *LAT52* in tomato encoded Ole domain-containing proteins,^{22,28} we were interested to know whether all Ole domain-containing genes were pollen specific. In order to find out, primarily, we predicted all *OsOle* genes by searching public databases based on whole rice genome sequences. Because the Ole domain-containing genes encoded proteins responsible for pollen allergen, we also predicted other two gene families with other two pollen-allergen domains: Allerg1 and Allerg2 in order to know whether these proteins are also pollen specific. Our data showed at least 114 genes encoding

proteins belonging to three groups of pollen-allergen families in rice genome.

To our knowledge, only one *OsOle* gene (*PS1*) was identified in rice genome,²² other 44 were first identified and partially characterized. In *Arabidopsis* genome, only 17 Ole domain-containing genes were predicted when we performed homology searches (Table 2). Ole domain-containing genes were originally characterized from tomato and maize as pollen-specific genes.^{20,21} Subsequently, the domain-containing proteins from olive were identified as the major allergen.³ Now 77 members have been identified from various plants (the Pfam database). However, only few genes were functionally characterized including pollen-specific genes *Zmc13*²⁰ and *LAT52*²¹ as well as pistil-specific genes.⁶ On the other hand, seven *OsAllerg2* genes were also first predicted, all of which were from the RiceGAAS annotation database and not from other databases. Despite the availability of 45 Allerg2 domain-containing sequences in the Pfam database, none of them were from rice. No homologs were found in rice genome when subjected to BLAST search using the 45 sequences, suggesting that rice Allerg2-containing genes might be different from other plants. Despite the identification of 45 Allerg2 domain-containing genes from 7 species (Pfam database), the functional and biological roles of this family is still unknown.

A total of 55 Allerg1 domain-containing genes have been predicted in rice genome (http://www.bio.psu.edu/expansins/other_species.htm), of which one member did not contain the Allerg1 domain. Among the remaining

54 genes, 40 have been characterized as the members of expansin family (26 α - and 14 β -expansin genes).^{13,29} We predicted 62 genes including the 54 genes previously identified. The additional eight predicted genes included four α -expansin genes (*OsAllerg1-4, 8, 15* and *45*) and four β -expansin genes (*OsAllerg1-33, 36, 37* and *44*). On the other hand, 38 expansin genes were reported in *Arabidopsis* genome.³⁰ However, detailed analysis showed that two of them were pseudogenes and another one showed incomplete gene structure. Thus, only 35 members were identified in the genome (Table 2). Besides these, 420 members of Allerg1-containing genes have been identified based on the Pfam database. The comparative analysis suggested that the Allerg1-containing genes should be the major pollen-allergen genes in rice and *Arabidopsis*.

Besides the three pollen-allergen domains, most of pollen-allergen genes contained other domains (Table 1), suggesting the multiple functions of pollen-allergen genes in rice. In *OsOle* genes, 13 contain the domain DUF1210 that presents in proline-rich plant proteins. The function of this domain is yet to be uncovered. In *OsAllerg1* genes, 41 members possessed a domain named DPBB_1, whose functions were also not well understood. Another domain that was detected in 13 *OsAllerg1* genes is GRP, which was found in several glycine-rich and stress-induced proteins.

The analysis of phylogenetic tree indicated that rice genome had at least 16 groups of *OsOle* genes, half of which consisted of only one member (Fig. 1A), indicating the high diversity in rice Ole-containing genes. Phylogenetic analysis also revealed that most of *OsOle* genes might be new members of the gene family as they were clustered into groups different from the genes from other plants. The presence of new members suggested the diversity of biological functions of rice *OsOle* allergen genes. On the other hand, low diversity was observed in Allerg1 domain-containing gene family as they were classified into three subgroups like other plant expansin families.¹² However, phylogenetic analysis revealed that the α -expansin-like subgroup was in the same group as β -expansin subgroup. Therefore, these four members might be regarded as members of β -expansin sub-family although they were identified as expansin-like genes in other studies (http://www.bio.psu.edu/expansins/other_species.htm). Additionally, the member *OsAllerg1-50*, which was identified as a β -expansin-like gene (http://www.bio.psu.edu/expansins/other_species.htm), was grouped together with β -expansin family. Therefore, the gene might be regarded as a member of β -expansin family. The results suggested that there might exist only two sub-families α and β -expansin instead of three sub-families in rice genome.

Rice pollen-allergen genes were distributed into 11 of 12 chromosomes excluding chromosome 11. They were

located on different positions of chromosomes. Some of them were mapped on long-arm, others on short-arm; some of them on centremere regions and others on the ends of chromosomes. However, they were not distributed randomly on chromosomes. Some of them were clustered on chromosomes. For example, 11 *OsOle* genes (*OsOle1-11*) were predicted from a 29 788 bp long fragment of BAC clone B1189A09 located on chromosome 1 (Fig. 3A). These genes shared high homology in their DNA and amino acid sequences, and were grouped together as an independent cluster in phylogenetic tree (Fig. 1A). Interestingly, some of pollen-allergen genes in a cluster exhibited same genomic DNA sequences but located on different positions. For example, both *OsAllerg1-26* and *29* had same sequences but they were located on different position of BAC clone OSJNBA0009C08 as shown in Supplemental Table 1b (www.dnaresearch.oxfordjournals.org). This specific distribution and high similarity in sequences of these genes suggested that the recent evolutionary event have happened within a short DNA fragment in pollen-allergen genes. Similarly, *OsOle39, 40* and *41* clustered together on the clone OSJNBA0031A07 located on chromosome 10 as shown in Supplemental Table 1a (www.dnaresearch.oxfordjournals.org). Similar distribution was also observed in *OsAllerg1* genes on chromosome 3 and 10 as shown in Supplemental Table 1b (www.dnaresearch.oxfordjournals.org), indicating the unique distribution of pollen-allergen genes on rice chromosomes.

Expression analysis was performed by RT-PCR. The analysis showed no expression of some pollen-allergen genes under normal growth conditions and these could be pseudogenes or stress-induced genes. In addition to this, most of pollen-allergen genes were expressed not only in panicles but also in other tissues, and thereby were not pollen-specific genes. However, our data showed six panicle-specific *OsOle* and two *OsAllerg1* genes that were best candidates of pollen-specific genes. Further analysis showed that the eight genes were anther-specific (Fig. 5C). Since only two pollen-specific genes were isolated and characterized in rice, more genes should be identified for further understanding of the mechanism of male gametophyte development. Our data might contribute to the identification of new pollen-specific genes. As for *OsAllerg1* genes, most of efforts focused on the expression analysis of non-panicle organs under various stresses and treatments such as submergence and auxin,^{13,31} and no data focused on the expression analysis in rice panicles. Since the expression of *OsAllerg1* genes in rice reproductive organs as shown in our data, those genes might play important roles in not only vegetative but also reproductive development.

It will become possible and important to identify and characterize a gene family in the whole genome by bioinformatics based on public database after the completion of whole rice genome sequences, and the release of reliable and powerful annotation system. To validate the

reliability and applicability of predicted pollen-allergen genes, we collected all full-length cDNA of pollen-allergen genes from RIKEN full-length cDNA database²⁴ and compared with all predicted coding regions. A total of 56 full-length cDNA were found to encode 56 corresponding pollen-allergen genes as shown in Supplemental Table 1a–c (www.dnaresearch.oxfordjournals.org). Among them, 23 match *OsOle* genes, 29 for *OsAllerg1* genes and the remaining 4 for *OsAllerg2* genes. Comparison analysis revealed that most of pollen-allergen genes were correctly predicted with the same exon and intron organization in encoding regions as shown in full-length cDNAs, suggesting the applicability of the prediction program to identify pollen-allergen genes. Although some of predicted genes were not correct in exon and intron organization as shown in Fig. 4, in most of cases, a researcher could obtain full-length coding region of a pollen-allergen gene by RT-PCR using primers designed from predicted sequences for further functional analysis.

In conclusion, we have identified 114 pollen-allergen genes in the rice genome and classified them. Also, we have partially characterized them for their expression in major tissues, intron/exons structure and other features. This is a preliminary analysis performed with the available sequence information and the gene prediction tools to predict the rice genes. In future, more pollen allergens may be uncovered in the rice genome when better gene analysis tools are available.

Supplementary Material: Supplementary material is available online at www.dnaresearch.oxfordjournals.org

References

- Villalba, M., Batanero, E., Lopez-Otin, C., Sanchez, L. M., Monsalve, R. I., Gonzalez de la Pena, M. A., Lahoz, C., and Rodriguez, R. 1993, The amino acid sequence of Ole e I, the major allergen from olive tree (*Olea europaea*) pollen, *Eur. J. Biochem.*, **216**, 863–869.
- Andersson, K. and Lidholm, J. 2003, Characteristics and immunobiology of grass pollen allergens, *Int. Arch. Allergy Immunol.*, **130**, 87–107.
- de Dios Alche, J., M'rani-Alaoui, M., Castro, A. J., and Rodrigues-Garcia, M. I. 2004, Ole e 1, the major allergen from Olive (*Olea europaea* L.) pollen, increases its expression and is released to the culture medium during *in vitro* germination, *Plant Cell Physiol.*, **45**, 1149–1157.
- Lauzurica, P., Gurbindo, C., Maruri, N., et al. 1988, Olive (*Olea europaea*) pollen allergens—I. Immunochemical characterization by immunoblotting, CRIE and immunodetection by a monoclonal antibody, *Mol. Immunol.*, **25**, 329–335.
- Marsh, D. G., Goodfriend, L., King, T. P., Lowenstein, H., and Platts-Mills, T. A. 1986, Allergen nomenclature, *Bull. World Health Organ.*, **64**, 767–774.
- Goldman, M. H., Pezzotti, M., Seurinck, J., and Mariani, C. 1992, Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins, *Plant Cell*, **4**, 1041–1051.
- Rodriguez, R., Villalba, M., Batanero, E., et al. 2002, Allergenic diversity of the olive pollen, *Allergy*, **57**, 6–16.
- Shcherban, T. Y., Shi, J., Durachko, D. M., et al. 1995, Molecular cloning and sequence analysis of expansins—a highly conserved, multigene family of proteins that mediate cell wall extension in plants, *Proc. Natl Acad. Sci. USA*, **92**, 9245–9249.
- Xu, H., Theerakulpisut, P., Goulding, N., Suphioglu, C., and Singh, M. B. 1995, Cloning, expression and immunological characterization of Ory s 1, the major allergen of rice pollen, *Gene*, **164**, 255–259.
- Cosgrove, D. J., Bedinger, P. A., and Durachko, D. M. 1997, Group I allergens of grass pollen as cell wall-loosening agents, *Proc. Natl Acad. Sci. USA*, **94**, 6559–6564.
- Cosgrove, D. J. 2000, Loosening of plant cell walls by expansins, *Nature*, **407**, 321–326.
- Li, Y., Darley, C. P., Ongaro, V., et al. 2002, Plant expansins are a complex multigene family with an ancient evolutionary origin, *Plant Physiol.*, **128**, 854–864.
- Lee, Y. and Kende, H. 2001, Expression of α -expansins is correlated with internodal elongation in deepwater rice, *Plant Physiol.*, **127**, 645–654.
- Schipper, O., Schaefer, D., Reski, R., and Fleming, A. 2002, Expansins in the bryophyte *Physcomitrella patens*, *Plant Mol. Biol.*, **50**, 789–802.
- Wu, Y.-J., Meeley, R. B., and Cosgrove, D. J. 2001, Analysis and expression of the α -expansin and β -expansin gene families in maize, *Plant Physiol.*, **126**, 222–232.
- Bufe, A., Schramm, G., Keown, M. B., Schlaak, M., and Becker, W. M. 1995, Major allergen Phl p Vb in timothy grass is a novel pollen RNase, *FEBS Lett.*, **363**, 6–12.
- Gehlhar, K., Petersen, A., Schramm, G., Becker, W. M., Schlaak, M., and Bufo, A. 1997, Investigation of different recombinant isoforms of grass group-V allergens (timothy grass pollen) isolated by low-stringency cDNA hybridization antibody binding capacity and allergenic activity, *Eur. J. Biochem.*, **247**, 217–223.
- Suphioglu, C. and Singh, M. B. 1995, Cloning, sequencing and expression in *Escherichia coli* of Pha a 1 and four isoforms of Pha a 5, the major allergens of canary grass pollen, *Clin. Exp. Allergy.*, **25**, 853–865.
- Matthiesen, F. and Lowenstein, H. 1991, Group V allergens in grass pollens. I. Purification and characterization of the group V allergen from *Phleum pratense* pollen, Phl p V, *Clin. Exp. Allergy*, **21**, 297–307.
- Hanson, D. D., Hamilton, D. A., Travis, J. L., Bashe, D. M., and Mascarenhas, J. P. 1989, Characterization of a pollen-specific cDNA clone from *Zea mays* and its expression, *Plant Cell*, **1**, 173–179.
- Muschietti, J., Dircks, L., Vancanneyt, G., and McCormick, S. 1994, LAT52 protein is essential for tomato pollen development: pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilization, *Plant J.*, **6**, 321–338.
- Zou, J. T., Zhan, X. Y., Wu, H. M., Wang, H., and Cheung, A. Y. 1994, Characterization of a rice pollen-specific gene and its expression, *Am. J. Bot.*, **81**, 552–561.
- Silvanovich, A., Astwood, J., Zhang, L., et al. 1991, Nucleotide sequence analysis of three cDNAs coding for Poa pIX isoallergens of kentucky bluegrass pollen, *J. Biol. Chem.*, **266**, 1204–1210.

24. Kikuchi, S., Satoh, K., Nagata, T., et al. 2003, Collection, mapping, and annotation of over 28,000 cDNA clones from Japonica rice, *Science*, **301**, 376–379.
25. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.*, **25**, 4876–4882.
26. Felsenstein, J. 1985, Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, **39**, 783–791.
27. Hills, D. M. and Bull, J. J. 1993, An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis, *Syst. Biol.*, **42**, 182–192.
28. Twell, D., Wing, R., Yamaguchi, J., and McCormick, S. 1989, Isolation and expression of an anther-specific gene from tomato, *Mol. Gen. Genet.*, **217**, 240–245.
29. Lee, Y., Choi, D., and Kende, H. 2001, Expansins: ever-expanding numbers and functions, *Curr. Opin. Plant Biol.*, **4**, 527–532.
30. Li, Y., Jones, L., and McQueen-Mason, S. 2003, Expansins and cell growth, *Curr. Opin. Plant Biol.*, **6**, 603–610.
31. Lee, Y. and Kende, H. 2002, Expression of α -expansins and expansin-like genes in deepwater rice, *Plant Physiol.*, **130**, 1396–1405.