Origin and Functional Prediction of Pollen Allergens in Plants

Miaolin Chen2, Jie Xu2, Deborah Devis, Jianxin Shi, Kang Ren, Iain Searle, and Dabing Zhang*

Joint International Research Laboratory of Metabolic and Developmental Sciences, Shanghai Jiao Tong University-University of Adelaide Joint Centre for Agriculture and Health, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China (M.C., J.X., J.S., K.R., D.Z.); School of Agriculture, Food, and Wine, University of Adelaide, Waite Campus, Urrbrae, South Australia 5064, Australia (D.D., I.S., D.Z.); and School of Biological Sciences, University of Adelaide, South Australia 5005, Australia (I.S.)

ORCID IDs: 0000-0002-6050-9947 (D.D.); 0000-0002-7717-0863 (J.S.); 0000-0002-1764-2929 (D.Z.).

Pollen allergies have long been a major pandemic health problem for human. However, the evolutionary events and biological function of pollen allergens in plants remain largely unknown. Here, we report the genome-wide prediction of pollen allergens and their biological function in the dicotyledonous model plant Arabidopsis (Arabidopsis thaliana) and the monocotyledonous model plant rice (Oryza sativa). In total, 145 and 107 pollen allergens were predicted from rice and Arabidopsis, respectively. These pollen allergens are putatively involved in stress responses and metabolic processes such as cell wall metabolism during pollen development. Interestingly, these putative pollen allergen genes were derived from large gene families and became diversified during evolution. Sequence analysis across 25 plant species from green alga to angiosperms suggest that about 40% of putative pollen allergenic proteins existed in both lower and higher plants, while other allergens emerged during evolution. Although a high proportion of gene duplication has been observed among allergen-coding genes, our data show that these genes might have undergone purifying selection during evolution. We also observed that epitopes of an allergen might have a biological function, as revealed by comprehensive analysis of two known allergens, expansin and profilin. This implies a crucial role of conserved amino acid residues in both in planta biological function and allergenicity. Finally, a model explaining how pollen allergens were generated and maintained in plants is proposed. Prediction and systematic analysis of pollen allergens in model plants suggest that pollen allergens were evolved by gene duplication and then functional specification. This study provides insight into the phylogenetic and evolutionary scenario of pollen allergens that will be helpful to future characterization and epitope screening of pollen allergens.

During the past four decades, allergic diseases have become a pandemic health problem. In general, pollen allergens are considered a major risk factor for both seasonal allergic rhinitis and asthma, and studies showed that more than 50% of patients with perennial allergic rhinitis are sensitized to pollen allergens. The sensitization rate of pollen is up to 30%, and the number of people affected by pollen allergy is on the increase worldwide (D’Amato et al., 2007; Pawankar et al., 2013). Unfortunately, pollen allergens are difficult to avoid because of the extremely small size and high prevalence of pollen, and this may contribute to pollen-food and pollen-fruit syndromes by cross-reactivity (Vieths et al., 2002).

Pollens from trees, grasses, and weeds all have been found to elicit allergic reactions in atopic individuals (Emberlin, 2009). To date, 11 groups of grass pollen allergens with the ability to elicit a specific IgE response in atopic individuals have been identified (Hrabina et al., 2008), and these mainly focused on pollen allergens from weeds and trees (Gadermaier et al., 2004; Mothes and Valenta, 2004). Those studies also suggested that these pollen allergens only belong to a few protein families, such as expansins, profilins, and calcium-binding proteins. Profilins are conserved in plants and act as a pan-allergen capable of inducing allergic reaction in various species (Valenta et al., 1992; Radauer and Breiteneder, 2006). Biologically, many pollen allergenic proteins are thought to play important physiological roles in pollen, especially the pollination process (Songmuang, 2013).
Pollen is the microgametophyte of seed plants that produces the male gametes (sperm cells) for subsequent sexual reproduction. The pollen protoplasm is surrounded by a specialized cell wall, the pollen wall, in which the inner pollen wall (also called the intine) is typically a thin multilayer composed of cellulose and pectin. In contrast, the exine refers to the very resistant outer wall that provides robust protection of the pollen grain from disintegration (Shi et al., 2015). Allergenic proteins are usually located within the pollen protoplast and readily released during the rehydration process (Grote, 1999). For example, birch (Betula spp.) pollen allergens Bet v 1 and Bet v 2 (profilin) are located within the pollen cytoplasm in the anhydrous state, in close proximity to ribosome-rich areas. Upon rehydration, birch pollen allergens are released within minutes from apertures and subsequently found on the entire pollen surface (Grote et al., 1993).

Over the past few decades, increasing information about allergens together with the advancement of bioinformatics tools has enabled scientists to predict and compare allergens from different sources (FAO/WHO, 2003; Stadler and Stadler, 2003; Saha and Raghava, 2006; Soeria-Atmadja et al., 2006; Wang et al., 2013c). These advances provided the prerequisites to allow a comparative analysis and a molecular evolution analysis of pollen allergens. Radauer and Breiteneder (2007) first introduced the evolutionary scope of the origin of plant allergens and proposed two scenarios for allergen evolution. One was that allergenicity could be an intrinsic property of the ancestral members of certain protein families still present in present-day allergens, and the other was that allergenicity emerged randomly in certain proteins and was inherited by their descendants. Recently, the evolution of major allergen gene families in peanut (Arachis hypogaea) was analyzed and revealed lineage-specific expansion and loss of allergenic genes (Ratnaparkhe et al., 2014). However, little information on the origin and evolution of pollen allergens has been reported.

In this study, we performed genome-wide analysis of potential pollen allergens in two well-studied model plants, the dicot Arabidopsis (Arabidopsis thaliana) and the monocot rice (Oryza sativa ssp. japonica), as well as their homologs in 25 species ranging from basal green alga to angiosperms. While some pollen allergens seemed to be derived from the duplication and diversification of large gene families from lower to higher plants, other allergens seemed to be recently evolved. Importantly, these genes seemed to have undergone purifying selection during evolution, implying that allergenic motifs are associated with the biological function of the allergens. A model is also proposed to explain how plants produced and maintained pollen allergens. This phylogenetic and evolutionary insight into pollen allergens will be useful in future characterization, epitope screening, and medical prevention of pollen allergens.

RESULTS

Prediction and Classification of Pollen Allergens in Arabidopsis and Rice

To identify putative pollen allergens from Arabidopsis and rice, we analyzed 186 and 261 candidates allergenic proteins by comparing the proteomic data of mature pollen from rice and Arabidopsis, respectively (Holmes-Davis et al., 2005; Noir et al., 2005; Dai et al., 2006; Sheoran et al., 2006). Next, using the combination of two methods for allergen prediction (PREAL [Wang et al., 2013c] and a sequence-based approach [FAO/WHO, 2003]), a total of 20 and 31 candidate proteins were identified as allergen proteins from rice and Arabidopsis, respectively. Furthermore, by analyzing transcriptomic data from mature pollen, 140 rice proteins and 94 Arabidopsis proteins were identified as putative allergens (Qin et al., 2009; Wei et al., 2010; Fig. 1, A and B). Together, we obtained 145 and 107 putative pollen allergens from rice and Arabidopsis, respectively (Fig. 1, A and B; Table I). Among the 145 rice candidates, five proteins were present only in the proteomic data, 15 in both the proteomic and transcriptomic data, and the remaining 125 only in the transcriptomic data. Similarly, of the 107 putative pollen allergens in Arabidopsis, 13 proteins were identified only from the proteomic data, 18 from both the proteomic and transcriptomic data, and the remaining 76 only in the transcriptomic data. The observation that most putative allergens were predicted from transcriptomic data sets is explained by the relatively low sensitivity of proteomic analysis.

The AllFam database of allergen families (Radauer and Breiteneder, 2006) contains over 2,500 protein families present in seed plants, and of these 2,500 families, about 59 plant protein families are inhalation allergens. In our analysis, we identified 254 putative pollen allergens (145 in rice and 107 in Arabidopsis) that were classified into 81 protein families, including most of the known allergenic pollen protein families present in the AllFam database (Radauer et al., 2008; Table I; Supplemental Fig. S1). Of these 81 families, 10 of the 13 known allergens were identified, except for three known allergens (Ara t GLP, Ara t 3, and Ory s 23) from Arabidopsis and rice (Supplemental Table S1), demonstrating the reliability of our prediction method. The absence of three known allergens is possibly due to their low expression levels in pollen and the absence of probes in microarrays (Qin et al., 2009; Wei et al., 2010).

Expression Analysis and Functional Prediction of Candidate Pollen Allergens

To understand the biological functions of these identified putative pollen allergens from Arabidopsis and rice, we performed Gene Ontology analysis and observed that these putative pollen allergens have key housekeeping biological functions, such as metabolic and cellular activities, stress response, and cellular...
component formation (Fig. 2A). For instance: Bet v 1, PR10 proteins, are associated with stress responses; profilins regulate actin polymerization by sequestering or releasing actin monomer during pollen growth; and polcalcins are involved in calcium signaling to help guide pollen tube growth (Supplemental Table S1).

To further characterize the functions of these putative pollen allergens, we performed in silico expression analysis. Our previous clustering analysis demonstrated that these candidates were present in pollen proteomic and transcriptomic data, but they also displayed distinct temporal expression patterns. In Figure 1, C and D, genes present in the green cluster (60 of 143 genes in rice and 33 of 107 genes in Arabidopsis) displayed ubiquitous expression that was associated mainly with stress responses, oxygen species metabolism, and glycolysis. The genes in the red cluster (31 of 143 genes in rice and 26 of 107 genes in Arabidopsis) exhibited high expression specifically in tricellular and mature pollens, and these genes were largely related to cell wall metabolism and organization (Figs. 1, C and D, and 2B). The main allergens specifically expressed in pollen include polygalacturonases, pectate lyases, and expansins that participate in the metabolism of carbohydrates and pollen tube wall formation during germination (Barral et al., 2005). These analyses imply that pollen-specific allergens were functionally restricted in pollen to be involved in cell wall metabolic activities, while the ubiquitous expressed putative allergens were associated mainly with stress responses (Supplemental Fig. S2).

Phylogenetic Analysis of Putative Pollen Allergens among 25 Plant Species

To understand the evolutionary events that gave rise to pollen allergens in plants, we identified the closest homologs (present in protein families) of these putative pollen allergens from rice and Arabidopsis in 25 sequenced plant species ranging from lower plants (green alga) to higher plants (angiosperms; Fig. 3). During angiosperm evolution, multiple rounds of polyploidy occurred (Bowers et al., 2003; Adams and Wendel, 2005); therefore, we proposed that pollen allergens might have expanded via gene duplication. In our analysis, a total of 1,797 and 1,302 close homologs of pollen allergens in rice and Arabidopsis were identified.
Table I. Gene information and family classification of putative pollen allergens

<table>
<thead>
<tr>
<th>Gene Identifier</th>
<th>Gene Family</th>
<th>Interpro</th>
<th>Specific Expression</th>
<th>Transcriptome</th>
<th>Proteome</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G23800</td>
<td>Aldehyde dehydrogenase</td>
<td>IPR015590</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G25780</td>
<td>Allergen V5/Tpx-1 related</td>
<td>IPR001283</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G09590</td>
<td>Allergen V5/Tpx-1 related</td>
<td>IPR001283</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G01310</td>
<td>Allergen V5/Tpx-1 related</td>
<td>IPR001283</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G66400</td>
<td>EF hand</td>
<td>IPR002048</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G17480</td>
<td>EF hand</td>
<td>IPR002048</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G03430</td>
<td>EF hand</td>
<td>IPR002048</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G03290</td>
<td>EF hand</td>
<td>IPR002048</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G77840</td>
<td>elf4-ylf5/elf2-ε</td>
<td>IPR003307</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G36530</td>
<td>Gelsolin</td>
<td>IPR007122</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G57320</td>
<td>Gelsolin</td>
<td>IPR007122</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G41740</td>
<td>Gelsolin</td>
<td>IPR007122</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G27120</td>
<td>Gene family candidate, 0016872</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G05620</td>
<td>Gene family candidate, 0020947</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G18310</td>
<td>Gene family candidate, 0025808</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G14040</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT3G07850</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT5G48140</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G02790</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G48100</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G07820</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT3G07840</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT3G07830</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR000743</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT1G55120</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR001362</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G62660</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR001362</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G26000</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR001362</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G36190</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR001362</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G22240</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR001362</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G09080</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT3G09440</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT5G02490</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT5G09590</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT3G42020</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT5G28540</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT3G37910</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT1G12580</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G11660</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G02500</td>
<td>Heat shock protein70</td>
<td>IPR013126</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G03030</td>
<td>Heat shock protein DnaJ, N terminal</td>
<td>IPR001623</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G24190</td>
<td>Heat shock protein90</td>
<td>IPR001404</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G56000</td>
<td>Heat shock protein90</td>
<td>IPR001404</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G03380</td>
<td>Heavy metal transport/detoxification protein</td>
<td>IPR006121</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G15020</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR010997</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G47520</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR010997</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G35240</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR010997</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G10920</td>
<td>Manganese and iron superoxide dismutase</td>
<td>IPR001189</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G08580</td>
<td>Microfibrillar-associated1, C terminal</td>
<td>IPR009730</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G63140</td>
<td>NAD-dependent epimerase/dehydratase</td>
<td>IPR001309</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G04500</td>
<td>Nucleotide-binding, α-β plait</td>
<td>IPR012677</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G14420</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G15110</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G24670</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G07010</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G1270</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G02720</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT1G69940</td>
<td>Pectinesterase, catalytic</td>
<td>IPR000070</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT5G07430</td>
<td>Pectinesterase, catalytic</td>
<td>IPR000070</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G07420</td>
<td>Pectinesterase, catalytic</td>
<td>IPR000070</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G06830</td>
<td>Pectinesterase, catalytic</td>
<td>IPR000070</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Table continues on following page.)
<table>
<thead>
<tr>
<th>Gene Identifier</th>
<th>Gene Family</th>
<th>Interpro</th>
<th>Specific Expression*</th>
<th>Transcriptome</th>
<th>Proteome</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT5G43060</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G09850</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G39090</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G47128</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G63470</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT3G52000</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G05850</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G12480</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT3G45010</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT3G10410</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT4G00230</td>
<td>Peptidase S8, subtilisin related</td>
<td>IPR015500</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G14067</td>
<td>Peptidase S8, subtilisin related</td>
<td>IPR015500</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G26330</td>
<td>Peptidase S8, subtilisin related</td>
<td>IPR015500</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G34870</td>
<td>Peptidyl-prolyl cis-trans-isomerase, cyclophilin type</td>
<td>IPR002130</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G16600</td>
<td>Peptidyl-prolyl cis-trans-isomerase, cyclophilin type</td>
<td>IPR002130</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G38740</td>
<td>Peptidyl-prolyl cis-trans-isomerase, cyclophilin type</td>
<td>IPR002130</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G21130</td>
<td>Peptidyl-prolyl cis-trans-isomerase, cyclophilin type</td>
<td>IPR002130</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G60570</td>
<td>Pollen allergen, N terminal</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G39700</td>
<td>Pollen allergen, N terminal</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT1G29140</td>
<td>Pollen Ole e 1 allergen and extensin</td>
<td>IPR006041</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT4G08685</td>
<td>Pollen Ole e 1 allergen and extensin</td>
<td>IPR006041</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G19760</td>
<td>Profilin, plant</td>
<td>IPR005455</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G29340</td>
<td>Profilin, plant</td>
<td>IPR005455</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT2G19770</td>
<td>Profilin, plant</td>
<td>IPR005455</td>
<td>S</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>AT3G44590</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G01100</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G00810</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G27710</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G74000</td>
<td>Stricotsidone synthase</td>
<td>IPR004141</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT2G28190</td>
<td>Superoxide dismutase, copper/zinc binding</td>
<td>IPR001424</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G08830</td>
<td>Superoxide dismutase, copper/zinc binding</td>
<td>IPR001424</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G75040</td>
<td>Thauatin, pathogenesis related</td>
<td>IPR001938</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G75050</td>
<td>Thauatin, pathogenesis related</td>
<td>IPR001938</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G45145</td>
<td>Thioredoxin fold</td>
<td>IPR012335</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G08710</td>
<td>Thioredoxin fold</td>
<td>IPR012335</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G42980</td>
<td>Thioredoxin fold</td>
<td>IPR012335</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G21750</td>
<td>Thioredoxin like</td>
<td>IPR017936</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G47470</td>
<td>Thioredoxin like</td>
<td>IPR017936</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G77510</td>
<td>Thioredoxin like</td>
<td>IPR017936</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G19510</td>
<td>Translation elongation factor EF1B, β- and δ-chains, guanine nucleotide exchange</td>
<td>IPR014038</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G68185</td>
<td>Ubiquitin</td>
<td>IPR000626</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G80040</td>
<td>Ubiquitin system component Cue</td>
<td>IPR003892</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G39900</td>
<td>Zinc finger, LIM type</td>
<td>IPR001781</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G55770</td>
<td>Zinc finger, LIM type</td>
<td>IPR001781</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G64920</td>
<td>Zinc finger, RING type</td>
<td>IPR001841</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G54680</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>AT1G30750</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

**Table continued on following page.**
<table>
<thead>
<tr>
<th>Gene Identifier</th>
<th>Gene Family</th>
<th>Interpro</th>
<th>Specific Expression</th>
<th>Transcriptome</th>
<th>Proteome</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_Os09g15770</td>
<td>elf4-γ/elf5/elf2-ε</td>
<td>IPR003307</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os10g083550</td>
<td>Enolase</td>
<td>IPR000941</td>
<td>T</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g14450</td>
<td>Enolase</td>
<td>IPR000941</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g20820</td>
<td>Enolase</td>
<td>IPR000941</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g25150</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os04g25160</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os04g25190</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os06g45200</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os06g444470</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os06g45160</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os06g45180</td>
<td>Expansin, cellulose-binding-like domain</td>
<td>IPR007117</td>
<td>T</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os10g40090</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g01610</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os12g36040</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os03g01640</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os01g60770</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g51040</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g01630</td>
<td>Expansin/pollen allergen, DPBB domain</td>
<td>IPR007112</td>
<td>T</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g47780</td>
<td>FAS1 domain</td>
<td>IPR000782</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g57570</td>
<td>Flavodoxin/nitric oxide synthase</td>
<td>IPR008254</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g42190</td>
<td>Flavodoxin/nitric oxide synthase</td>
<td>IPR008254</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g51440</td>
<td>Gelsoin</td>
<td>IPR007122</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g24220</td>
<td>Gelsoin</td>
<td>IPR007122</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g44890</td>
<td>Gelsoin</td>
<td>IPR007122</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g46080</td>
<td>Gene family candidate, 0005331</td>
<td>IPR001087</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g32190</td>
<td>Gene family candidate, 0014938</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g12620</td>
<td>Gene family candidate, 0015182</td>
<td>/</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g21120</td>
<td>Gene family candidate, 0015338</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g09180</td>
<td>Gene family candidate, 0018686</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g17140</td>
<td>Gene family candidate, 0020281</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g38490</td>
<td>Gene family candidate, 0021210</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g13980</td>
<td>Glycoside hydrolase, family 16</td>
<td>IPR013320</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g39060</td>
<td>Glycoside hydrolase, family 17</td>
<td>IPR000490</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g41610</td>
<td>Glycoside hydrolase, family 17</td>
<td>IPR000490</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g71810</td>
<td>Glycoside hydrolase, family 17</td>
<td>IPR000490</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g36280</td>
<td>Glycoside hydrolase, family 19, catalytic</td>
<td>IPR000726</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g41680</td>
<td>Glycoside hydrolase, family 19, catalytic</td>
<td>IPR000726</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g39330</td>
<td>Glycoside hydrolase, family 19, catalytic</td>
<td>IPR000726</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g41100</td>
<td>Glycoside hydrolase, family 19, catalytic</td>
<td>IPR000726</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g10300</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os01g33300</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os06g35320</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os08g23790</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>S</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os06g40890</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g46510</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g61800</td>
<td>Glycoside hydrolase, family 28</td>
<td>IPR012334</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g01590</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR013148</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g45290</td>
<td>Glycoside hydrolase, family 32</td>
<td>IPR013148</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os11g08440</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g16920</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g16880</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os11g47760</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g16860</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g02410</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g53420</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g02260</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g60620</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td>P</td>
</tr>
<tr>
<td>LOC_Os09g31486</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g38530</td>
<td>Heat shock protein70</td>
<td>IPR001023</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g01740</td>
<td>Heat shock protein70</td>
<td>IPR001404</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g39140</td>
<td>Heat shock protein70</td>
<td>IPR001404</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene Identifier</td>
<td>Gene Family</td>
<td>Interpro</td>
<td>Specific Expression*</td>
<td>Transcriptome</td>
<td>Proteome</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------</td>
<td>----------</td>
<td>----------------------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>LOC_Os09g30412</td>
<td>Heat shock protein90</td>
<td>IPR001404</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g29840</td>
<td>Heat shock protein90</td>
<td>IPR001404</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g51140</td>
<td>Helix-loop-helix DNA binding</td>
<td>IPR011598</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g04390</td>
<td>Helix-loop-helix DNA binding</td>
<td>IPR011598</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os10g25420</td>
<td>Lipase, GDSL</td>
<td>IPR01087</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g25040</td>
<td>Lipase, GDSL</td>
<td>IPR01087</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g49880</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR01097</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g46070</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR01097</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g33720</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR01097</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g61380</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR01097</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g56280</td>
<td>Malate dehydrogenase, type 1</td>
<td>IPR01097</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g25850</td>
<td>Manganese and iron superoxide dismutase</td>
<td>IPR001189</td>
<td>T, P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g43710</td>
<td>Man-6-P receptor, binding</td>
<td>IPR009011</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os07g06590</td>
<td>MD-Z-related lipid recognition</td>
<td>IPR003172</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g27760</td>
<td>Met sulfoxide reductase B</td>
<td>IPR002579</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g28060</td>
<td>MORN motif</td>
<td>IPR003409</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g05660</td>
<td>Nucleosome assembly protein</td>
<td>IPR002164</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g05209</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g12300</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g38510</td>
<td>Pectate lyase/Amb allergen</td>
<td>IPR002022</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os11g45730</td>
<td>Pectinesterase, catalytic</td>
<td>IPR012334</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g43010</td>
<td>Peptidase C13, legumain</td>
<td>IPR001096</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g24600</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g27030</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g01810</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g73980</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os11g14900</td>
<td>Peptidase C1A, papain</td>
<td>IPR013128</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g09190</td>
<td>Peptidase S10, Ser carboxypeptidase</td>
<td>IPR001563</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g47150</td>
<td>Peptidase S8, subtilisin related</td>
<td>IPR015500</td>
<td>S, T, P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g44590</td>
<td>Peptidase S8, subtilisin related</td>
<td>IPR015500</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g49480</td>
<td>Peptidyl-prolyl cis-trans-isomerase</td>
<td>IPR002130</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g39780</td>
<td>Peptidyl-prolyl cis-trans-isomerase</td>
<td>IPR002130</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g01270</td>
<td>Peptidyl-prolyl cis-trans-isomerase</td>
<td>IPR002130</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g02890</td>
<td>Peptidyl-prolyl cis-trans-isomerase</td>
<td>IPR002130</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g40010</td>
<td>Plant lipid transfer protein</td>
<td>IPR000528</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g39950</td>
<td>Pollen Ole e 1 allergen and extensin</td>
<td>IPR006041</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g51950</td>
<td>Pollen Ole e 1 allergen and extensin</td>
<td>IPR006041</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os10g17660</td>
<td>Profilin, plant</td>
<td>IPR002097</td>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g05880</td>
<td>Profilin, plant</td>
<td>IPR002097</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g55710</td>
<td>Proteasome maturation factor UMP1</td>
<td>IPR008012</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g57810</td>
<td>Protein of unknown function DUF689</td>
<td>IPR007785</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g27490</td>
<td>Protein of unknown function DUF702</td>
<td>IPR007818</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g05630</td>
<td>Protein phosphatase 2C, N terminal</td>
<td>IPR015655</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g16430</td>
<td>Proteinase inhibitor I25, cystatin</td>
<td>IPR000010</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g25540</td>
<td>Rapid alkalization factor</td>
<td>IPR008801</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g48780</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g13080</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g37330</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g02340</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g09510</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g32760</td>
<td>Ribosomal protein 60S</td>
<td>IPR001813</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g22810</td>
<td>Superoxide dismutase, copper/zinc binding</td>
<td>IPR001424</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os07g46090</td>
<td>Superoxide dismutase, copper/zinc binding</td>
<td>IPR001424</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g11960</td>
<td>Superoxide dismutase, copper/zinc binding</td>
<td>IPR001424</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os11g36340</td>
<td>Targeting for Xklp2</td>
<td>IPR009675</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g24090</td>
<td>Tetratricopeptide region</td>
<td>IPR013026</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g44830</td>
<td>Thioredoxin, core</td>
<td>IPR013766</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g06430</td>
<td>Thioredoxin like</td>
<td>IPR013766</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g27830</td>
<td>Thioredoxin like</td>
<td>IPR013766</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g23740</td>
<td>Thioredoxin like</td>
<td>IPR013766</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os06g42000</td>
<td>Thioredoxin-like fold</td>
<td>IPR012336</td>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Table continues on following page.)
from the genomes of the 25 plant species, and in most families, the number of homologs increased from green algae to angiosperms. Notably, some putative allergic protein families displayed multiple sequences with high similarity in one species. For example, two rice expansins had 12 close sequence homologs in higher plants. Allergenic polygalacturonases such as Ara t 3, Zea m 14, and Tri a 14 may have the function of transferring lipids and fatty acids through cell membranes (Thoma et al., 1999). The pollen allergen in some grass and cypress species, only has homologs in higher plants. Allergenic polygalacturonases from the Japanese cypress Chamaecyparis obtusa (Mori et al., 1999) and timothy grass (Phleum pratense; Suck et al., 2000) play roles in pollen maturation and pollen tube growth (Supplemental Table S2). Another allergen found only in higher plants, pollen Ole e 1 allergen (Jimenez-Lopez et al., 2012), accumulates in pollen tube cell walls and may have a role in pollen germination and pollen tube growth (Supplemental Table S2). Interestingly, Arabidopsis pectinesterase, another pollen allergen (Mahler et al., 2001), only has homologs in dicots, which have sequence variation with that of rice counterparts. Pectinesterase from olive (Olea europaea) was reported to affect cell wall stability during pollen germination and pollen tube growth through the deesterification of pectin into pectate and methanol (Salamanca et al., 2010; Esteve et al., 2012; Jimenez-Lopez et al., 2012). Altogether, our observations on the putative allergens of 33 other families suggest that they may have evolved in parallel in either monocots or dicots with diversified biological functions.

Table 1. (Continued from previous page.)

<table>
<thead>
<tr>
<th>Gene Identifier</th>
<th>Gene Family</th>
<th>Interpro</th>
<th>Specific Expression</th>
<th>Transcriptome</th>
<th>Proteome</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_Os10g25290</td>
<td>Tify</td>
<td>IPR010399</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os07g46750</td>
<td>Translation elongation factor EF1B, (\beta)- and (\delta)-chains, guanine nucleotide exchange</td>
<td>IPR014038</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os07g42300</td>
<td>Translation elongation factor EF1B, (\beta)- and (\delta)-chains, guanine nucleotide exchange</td>
<td>IPR014038</td>
<td>T</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g37310</td>
<td>Uncharacterized protein family UPF0029, N terminal</td>
<td>IPR001498</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g50490</td>
<td>X8 domain</td>
<td>IPR012946</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g36820</td>
<td>Zinc finger, Mcm10/DnaG type</td>
<td>IPR015408</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g72480</td>
<td>Zinc finger, RING type</td>
<td>IPR001841</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g30940</td>
<td>/</td>
<td>/</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g29740</td>
<td>/</td>
<td>/</td>
<td>S</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>LOC_Os09g31080</td>
<td>/</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os08g25080</td>
<td>/</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g57440</td>
<td>/</td>
<td>/</td>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Genes that belong to the pollen specifically expressed gene cluster are marked S.

Chen et al.
Evolutionary Events in Generating and Maintaining Pollen Allergens

Gene duplication events that produce functionally redundant genes have been considered a major driver underlying gene evolution (Nei, 1969; Lynch and Conery, 2000; Cui et al., 2015). Therefore, we asked whether sequence variation within these duplicated genes affects the allergenicity of proteins. Pollen allergens seemed to be produced by duplication events. The proportion of duplicated genes (including tandem repeat and block repeat) in pollen-expressed genes was about 40% in Arabidopsis and 30% in rice. However, the percentage of duplicated genes in putative pollen allergens increased markedly, 60% in Arabidopsis and 49% in rice (Fig. 4, A and B). In genetics, Ka/Ks represents the ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site, and the value of Ka/Ks can be used as an indicator of selective pressure on a protein-coding gene. A gene with Ka/Ks $>1$ is usually regarded as having evolved under positive selection, while Ka/Ks $<1$ is usually regarded as an indicator of genes having undergone purifying selection (Hurst, 2002). Although many putative pollen allergen genes seemed to be produced by duplication, the Ka/Ks values of these genes were low, which means a low ratio of nonsynonymous substitutions of these genes.
suggesting that these pollen allergenic proteins evolved under purifying selection (Fig. 4, C and D). In rice, the Ka/Ks rate of allergen genes was around 0.25, which is similar to that of Arabidopsis (below 0.2), indicating that pollen allergens generated from duplication events have been maintained by purifying selection.

**Profilins Represent the Ancient Allergenic Families**

To further investigate the evolution and the relationship between allergenicity and the biological function of pollen allergens, two major allergenic families, profilins and expansins, were analyzed further. Profilin is an actin-binding protein involved in the dynamic turnover and restructuring of the actin cytoskeleton. Plant profilins share many of the same biochemical properties and are structurally similar to nonplant profilins (Thorn et al., 1997). Profilin is a common pan-allergen in plants and is present in many plant organs, thereby leading to various routes of exposure depending on the plant species (Valenta et al., 1992).
allergenic profilins that have the same route of exposure tend to be in another clade. For example, pollen profilins were seen in the grass family, fruit profilins in family Rosaceae, and seed profilins in family Leguminosae (Fig. 5A). LOC_Os10g17660 and LOC_Os10g17680 are tandem duplicated genes in rice, and both were highly expressed in late anther developmental stages, while tandem duplicated gene pairs (AtProfilin1/AtProfilin5 and AtProfilin2/AtProfilin4) showed totally different expression patterns in Arabidopsis. AtProfilin1 and AtProfilin2 were expressed in many tissues, while AtProfilin4 and AtProfilin5 were expressed specifically and highly in pollen (Fig. 5B). AtProfilin4 and AtProfilin5 redundantly regulate polarized pollen tube growth (Liu et al., 2015). Obviously, proteins like AtProfilin4 and AtProfilin5 have a higher probability to be pollen allergens. The sequence and structure of Ara h 5 in peanut have been studied extensively, and eight surface-exposed amino acids have been highlighted as epitopes (Fig. 5C). Within these eight epitopes, amino acids in common with Ara h 5 of allergenic profilins are colored in yellow, while amino acids different from Ara h 5 common epitopes are colored in blue. Crucial residues of known biological function and structural role are marked by red stars (Thorn et al., 1997). Ara h, Arachis hypogaea; Mal d, Malus domestica; Api g, Apium graveolens; Zea m, Zea mays; Os, Oryza sativa; CR, Chlamydomonas reinhardtii; VC, Volvox carteri; MRCC, Micromonas sp. RCC. D, Three-dimensional (3D) models of Ara h 4 (left) and AtProfilin1 (right), with putative surface epitopes of Ara h 5 and the corresponding position of AtProfilin1 shown. Different epitopes (#1–#7) were mapped on the surface in different colors.
epitopes were identified (Radauer et al., 2006; Cabanos et al., 2010) and are were mapped in Figure 5C for comparison. These epitopes included some crucial amino acid residues required for biological function and structural roles in profilin; for example, epitope 1 includes two pyridoxal-5’-phosphate-binding residues (Thorn et al., 1997; Fig. 5C; Supplemental Fig. S5). Most known allergenic profilins, such as Zeam 12, Mal d 4, and Api g 4, displayed almost no variation in epitope sequence, while profilins in rice and lower plants exhibited more variation (Fig. 5C; Supplemental Fig. S4). These results indicate that the allergenicity of the profilin family was changed possibly through evolution. Furthermore, variations in epitope position caused structural changes in the proteins of Ara h 5 and AtProfilin1 (Fig. 5D).

Variations in epitopes also were found among members of the profilin family of Arabidopsis.

Allergenicity Evolved with the Functional Specification of Expansins in Grass

Expansins are proteins that promote cell wall loosening and extension (Cosgrove, 2000). In pollen, expansins may facilitate cell wall deposition in pollen grains and are involved in pollen germination (Choi et al., 2006). Even though expansins have numerous family members present in both dicots and monocots, only members in the EXPB-I (for β-expansin I) clade of β-expansins in grass are allergenic. In grasses, the
EXPB-I clade was separated into two groups (conservative EXPB-I and divergent EXPB-I) by the sigma whole-genome duplication, while known allergenic β-expansins gathered in subbranches of divergent EXPB-I (Tang et al., 2010). The divergent EXPB-I might have evolved to act on highly substituted xylans that were the interstitial material of primary walls in grasses (Sampedro et al., 2015). Phylogenetic analysis showed that expansins in rice clustered into two main branches (conservative EXPB-I and divergent EXPB-I), and all expansins in Arabidopsis belonged to the conserved EXPB-I clade (Fig. 6A). Ory s 1 allergens, which include OsEXPB1, OsEXPB10, and OsEXPB13, were highly expressed in later developmental stages of anther (microspore/pollen) and inflorescence development (Xu et al., 1995; Hirano et al., 2013; Fig. 6B).

Sequence alignment of β-expansins demonstrated that the identified epitopes of allergenic expansins differed from those of their nonallergic expansin orthologs present in lower plants (Selaginella moellendorfii and Physcomitrella patens), dicots, and monocots (Fig. 6C; Supplemental Fig. S5). These epitopes also included important residues: the epitope SITE-A identified by Esch and Klapper (1989) contained a short binding pocket, and SITE-D identified by Hiller et al. (1997) covered part of the long conserved binding surface with the motif TWYG (Yennawar et al., 2006). Sequence variation among these expansins may lead to diverse functions and allergenicity of each expansin. Rice Ory s 1 is homologous to the maize allergen Zea m 1 and two other pollen allergens, Lol p 1 and Phl p 1 from rye-grass (Lolium perenne) and timothy grass, respectively (Petersen et al., 1995; Cosgrove et al., 1997; Yennawar et al., 2006). Zea m 1 was suggested to be involved in cell wall loosening of the stigma and style, aiding in pollen tube invasion of maternal tissue (microspore/pollen) and inflorescence development (Xu et al., 1995; Hirano et al., 2013; Fig. 6B).

DISCUSSION

Pollen grain-caused allergen is one of the most intractable problems in allergy research. Large numbers of pollen allergens have been characterized, but little is known about their evolution and taxonomic distribution patterns. To provide answers to these questions, we performed genome-wide allergen prediction of transcriptomic and proteomic data sets in the model monocot rice and dicot Arabidopsis and performed phylogenetic analysis of pollen allergens. The taxonomic distribution of putative pollen allergens was investigated using phylogenetic analysis, which showed distinct distribution patterns for some of these allergens. Both the expression pattern and the taxonomic distribution of these putative pollen allergens in model plants are likely to be useful to predict potential allergens in other plant species, especially those species without complete genome sequences. The sequence variation of allergen proteins among species, especially between lower and higher plants, indicated that allergenicity might change along with plant evolution.

In many pollen allergens like allergenic expansins and profilins, epitopes usually include important functional amino acid residues. We observed low Ka/Ks values and higher gene duplication ratios in putative pollen allergens, which importantly also indicated a relationship between allergenicity and the evolution of protein functions. Therefore, we suggest that allergenicity might be a by-product of gene duplication and functional specification.

Conserved epitope sequences in allergens have been proposed to result in desensitization in humans after long-term exposure (Radauer et al., 2012). Gene

![Model of the origination and evolution of pollen allergen genes in plants.](image)
duplication promotes neofunctionalization by variation of protein sequence, thereby promoting the opportunity for new allergen formation or changing the allergenicity of previous allergens. We observed significantly higher gene duplication rates of putative pollen allergens in both rice and Arabidopsis. Allergenicity emerged from gene duplication events in some cases. For example, the EXPB-I clade of this family was separated into two groups by gene duplication: a divergent group containing allergenic β-expansins and a conservative group (Sampedro et al., 2015). The lack of divergent EXPB-I genes in eudicots or in the recently sequenced genomes of banana (Musa spp.), date palm (Phoenix dactylifera), and oil palm (Elaeis guineensis) also supports a recent split (Tang et al., 2010). In rice, divergent and conservative EXPB-I groups were inferred to have evolved from the sigma whole-genome duplication in grasses, and changes in tissue expression of divergent EXPB-I permitted pollen-specific β-expansins (OsEXPB1, OsEXPB10, OsEXPB13, and OsEXPB9). OsEXPB1, OsEXPB10, and OsEXPB13 were produced by tandem duplication events, and OsEXPB9 was produced by the rho whole-genome duplication (Tang et al., 2010). In addition, features of the expansin family demonstrated the way that gene duplication led to function specification and allergenicity. Divergent EXPB-I proteins may have evolved to act on a preferred substrate, highly substituted xylans in grasses (Sampedro et al., 2015).

Allergens have stringent structural and epitope requirements (Burks et al., 1999); however, variation within the epitope may create new allergens or disrupt the allergenicity. One good example is the peanut allergen Ara h 3 gene family, which arose by segmental and tandem duplications and evolved in a conservative manner (Ratnaparkhe et al., 2014). Low Ka/Ks rates of putative pollen allergens in rice and Arabidopsis indicate that these allergens might have experienced purifying selection (Fig. 4, C and D). The limited ratio of nonsynonymous mutations implied that these allergens might have evolved to have unique functions in pollen. The molecular function of a protein requires a stable structure, and so do existing allergens. Our data suggest that epitopes might be located in conserved functional sites of putative allergenic proteins, as we observed a limited ratio of nonsynonymous mutation in putative pollen allergens. As mentioned previously, pollen allergens tended to be involved in cell wall (pollen wall) metabolic processes and stress responses (Supplemental Table S2), which indicated that they underwent a strict purifying selection through pollen competition or other stresses to perform the function. That also may be the reason for the phenomenon that putative pollen allergens showed both higher gene duplication rates and lower Ka/Ks values. Allergenic β-expansins are good examples influencing the outcome of pollen competition by affecting pollen tube growth (Valdivia et al., 2007).

CONCLUSION

In summary, this work predicted 145 and 107 pollen allergens from rice and Arabidopsis, respectively and these pollen allergens are associated with stress responses and metabolic events during pollen development. Interestingly, sequence analysis across 25 plant species from low plants to high plants suggests that some pollen allergens belong to large gene families generated by gene duplication, purifying selection, and functional diversification during evolution. During this process, two selection processes were evident: the fixation of duplication (maintaining the allergenicity) and the fixation of allergen-determining residues (retaining allergenic epitopes). Stress, pollen competition, and functional selection (like cell wall metabolic processes) could be involved in the fixation processes (Fig. 7). Our analysis of putative pollen allergens from model plants is helpful to predict pollen allergens in other species and future medical treatment of pollen allergenicity. Our model of pollen allergen evolution could provide an insight into the mechanisms underlying how allergenicity evolved and help in the identification of epitopes.

MATERIALS AND METHODS

Identification of Allergenic Genes

Gene sequences for the prediction of allergens present in mature pollen grains of Arabidopsis (Arabidopsis thaliana) and rice (Oriza sativa) were collected from the literature (Holmes-Davis et al., 2005; Noir et al., 2005; Dai et al., 2006; Sheoran et al., 2006). Gene expression data sets of pollen in Arabidopsis and rice (Qin et al., 2009; Wei et al., 2010), Gsm692545, Gsm692546, Gsm69254, Gsm436364, Gsm436365, Gsm436366, and Gsm436367, were downloaded from the Gene Expression Omnibus at the National Center for Biotechnology Information. Allergens from rice and Arabidopsis, respectively and allergens from rice and Arabidopsis were downloaded from the Allergome database (http://www.allergome.org/; Mari et al., 2006) and the World Health Organization/IUIS Allergen Nomenclature official database (http://www.allergen.org/). Protein sequence data in FASTA format were downloaded from the Universal Protein Resource database release 2014_03 (http://www.uniprot.org/; UniProt Consortium, 2014).

Sequence-based and maximum relevance minimum redundancy feature selection methods were used to detect potential allergens in mature pollen using two published prediction tools, proAP (Wang et al., 2013b) and PREAL (Wang et al., 2013c), on our server. The sequence-based approach was proposed by the Food and Agriculture Organization of the United Nations and the World Health Organization (FAO/WHO, 2003), and the number of exact matches in a stretch of consecutive identical amino acids was set to more than eight (rule 1).

Sequence-based and maximum relevance minimum redundancy feature selection methods were used to detect potential allergens in mature pollen using two published prediction tools, proAP (Wang et al., 2013b) and PREAL (Wang et al., 2013c), on our server. The sequence-based approach was proposed by the Food and Agriculture Organization of the United Nations and the World Health Organization (FAO/WHO, 2003), and the number of exact matches in a stretch of consecutive identical amino acids was set to more than eight (rule 1). Proteins predicted by both methods were retained as potential pollen allergens to ensure accuracy (Wang et al., 2013c). The prediction results and information on putative pollen allergens are shown in Supplemental Data Set S1.

Gene Expression Profile Analysis

The expression data of genes corresponding to potential allergens in Arabidopsis and rice were downloaded from the Bio-Analytic Resource for Plant Biology (http://bar.utoronto.ca/; Toufighi et al., 2005) or the Rice Oligonucleotide Array Database (http://www.ricearray.org/; Cao et al., 2012), respectively. Information
about the expression data, such as growth stage, tissues, and samples, is listed in Supplemental Data Set S2. To avoid batch effects, Combat (Johnson et al., 2007), an R package, was used to adjust expression data from different experiments. To examine expression patterns and the specificity of target genes, the data were clustered by a hypergeometric distribution test.

MapMan and GO Analysis

The PLAZA database version 2.5 (http://bioinformatics.psb.ugent.be/plaza/; Van Bel et al., 2012) and the PANTHER classification system (http://pantherdb.org/; Mi et al., 2013) were used to perform GO classification and enrichment analysis. To investigate the metabolic processes involved, MapMan was used to check the metabolic overview of potential allergens (Thimm et al., 2004), and significance was tested by a hypergeometric distribution test.

Protein Family and Taxonomic Distribution Analysis

Genes were classified into protein families using the Pfam protein families database version 27.0 (http://pfam.xfam.org/; Finn et al., 2014) and the Plant Gene Family Database (http://green.dna.affrc.go.jp/PGF-DB/). Homologs including in-paralogs (i.e. BLAST hit of genes in the same species having higher bit scores than the best hit from any other species) were obtained from 25 plants (including Arabidopsis and rice) after BLAST at the PLAZA database version 2.5 (E value threshold of 1e-05).

Construction of the Phylogenetic Tree, Sequence Analysis, and 3D Modeling

The Clustal Omega (Sievers and Higgins, 2014) server at the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalo/) was used to compare protein sequences downloaded from the UniProt database. Results of sequence alignments are shown with known secondary structure information from the Protein Data Bank (Berman et al., 2000) by the Web-based tool Easy Sequencing in PostScript (Robert and Gouet, 2014). Unrooted phylogenetic trees were reconstructed by MEGA6 (Tamura et al., 2013) using neighbor-joining and maximum likelihood methods. The 3D structures of Ara h 5 and AtProfilin1 were obtained from the Protein Data Bank under accession numbers 4ESP (Wang et al., 2013d) and 1AOK (Thom et al., 1997), and the 3D models were visualized by UCSF Chimera (Petterson et al., 2004).

Gene Duplication Analysis and Genome-Level Ka/Ks Estimation

Gene duplication data were obtained from the PLAZA database version 2.5 including tandem duplication and block duplication. These duplication events were identified through collinearity information using i-ADHoRe version 3.0 (Proost et al., 2012). To estimate selective pressure acting on genes, four closely related species to Arabidopsis and rice were chosen to calculate Ka/Ks rates in each species. Homolog gene pairs of Arabidopsis and rice were identified with the method of paralogous pairs. These duplication events were reconstructed by MEGAl4 (Tamura et al., 2013) using neighbor-joining and maximum likelihood methods. The 3D structures of Ara h 5 and AtProfilin1 were obtained from the Protein Data Bank under accession numbers 4ESP and 1AOK (Thom et al., 1997), and the 3D models were visualized by UCSF Chimera (Petterson et al., 2004).

Gene Duplication Analysis and Genome-Level Ka/Ks Estimation

Gene duplication data were obtained from the PLAZA database version 2.5 including tandem duplication and block duplication. These duplication events were identified through collinearity information using i-ADHoRe version 3.0 (Proost et al., 2012). To estimate selective pressure acting on genes, four closely related species to Arabidopsis and rice were chosen to calculate Ka/Ks rates in each species. Homolog gene pairs of Arabidopsis and rice were identified with the method of paralogous pairs. These duplication events were reconstructed by MEGAl4 (Tamura et al., 2013) using neighbor-joining and maximum likelihood methods. The 3D structures of Ara h 5 and AtProfilin1 were obtained from the Protein Data Bank under accession numbers 4ESP and 1AOK (Thom et al., 1997), and the 3D models were visualized by UCSF Chimera (Petterson et al., 2004).

Accession Numbers

Accession numbers for the genes in this article are as follows: OsEXPB1a (LOC_Os03g01610), OsEXPB1e (LOC_Os03g01650), OsEXPB2 (LOC_Os10g40710), OsEXPB2 (LOC_Os10g40710), OsEXPB3 (LOC_Os10g40720), OsEXPB4 (LOC_Os10g40730), OsEXPB5 (LOC_Os04g46650), OsEXPB6 (LOC_Os10g40700), OsEXPB7 (LOC_Os03g01270), OsEXPB8 (LOC_Os03g01260), OsEXPB9 (LOC_Os10g40990), OsEXPB10 (LOC_Os03g01640), OsEXPB11 (LOC_Os02g44108), OsEXPB12 (LOC_Os03g44290), OsEXPB13 (LOC_Os05g1630), OsEXPB14 (LOC_Os02g44106), OsEXP15 (LOC_Os04g46630), OsEXPB16 (LOC_Os02g42650), OsEXPB17 (LOC_Os04g47480), OsEXPB18 (LOC_Os05g15960), AtEXPB1 (AT2G20750), AtEXPB2 (AT1G65680), AtEXPB3 (AT4G28520), AtEXPB4 (AT2G45110), AtEXPB5 (AT3G06570), AtProfilin1 (AT2G19760), AtProfilin2 (AT4G29350), AtProfilin3 (AT3G56600), AtProfilin4 (AT4G29340), and AtProfilin5 (AT2G19770).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Analysis of the putative pollen allergen protein families in rice and Arabidopsis.

Supplemental Figure S2. Summary of the MapMan classification of the candidate allergen genes in rice and Arabidopsis.

Supplemental Figure S3. Sequence alignment of the expansin family.

Supplemental Figure S4. Taxonomic distribution of expansin homologs.

Supplemental Figure S5. Sequence alignment of the profilin family.

Supplemental Table S1. Published allergens found in pollen genes predicted in rice and Arabidopsis.

Supplemental Table S2. Summary of gene family functions of putative allergen homologs.

Supplemental Data Set S1. Candidate pollen allergens identified in proteome and transcriptome.

Supplemental Data Set S2. Expression profiles of putative pollen allergens in rice and Arabidopsis.

Supplemental Data Set S3. Ka/Ks values of putative pollen allergens in rice and Arabidopsis.

Received April 21, 2016; accepted July 17, 2016; published July 19, 2016.

LITERATURE CITED


Evolution and Functional Prediction of Pollen Allergens

Chen et al.


Sampedro J, Gutman M, Li LC, Cosgrove DJ (2015) Evolutionary divergence of β-expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits. Plant J 81: 108–120


Evolution and Functional Prediction of Pollen Allergens


Copyright © 2016 American Society of Plant Biologists. All rights reserved.