PLOS | ONE

RESEARCH ARTICLE

Comparative and Evolutionary Analysis of Grass Pollen Allergens Using *Brachypodium distachyon* as a Model System

Akanksha Sharma[1], Niharika Sharma[2], Prem Bhalla[1], Mohan Singh[1]*

1 Plant Molecular Biology and Biotechnology Laboratory, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Melbourne, VIC, Australia, **2** Australian Centre for Plant Functional Genomics (ACPFG), School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Plant Genomics Centre, Hartley Grove, Urrbrae, SA, Australia

* mohan@unimelb.edu.au

Check for updates

OPEN ACCESS

Citation: Sharma A, Sharma N, Bhalla P, Singh M (2017) Comparative and Evolutionary Analysis of Grass Pollen Allergens Using *Brachypodium distachyon* as a Model System. PLoS ONE 12(1): e0169686. doi:10.1371/journal.pone.0169686

Editor: Ulrich Melcher, Oklahoma State University, UNITED STATES

Received: August 22, 2016

Accepted: December 19, 2016

Published: January 19, 2017

Copyright: © 2017 Sharma et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are made available as supplementary files.

Funding: The research was funded by the University of Melbourne as a postgraduate scholarship for Akanksha Sharma. Financial support from the Australian Research Council (ARC DPO988972) is also gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Comparative genomics have facilitated the mining of biological information from a genome sequence, through the detection of similarities and differences with genomes of closely or more distantly related species. By using such comparative approaches, knowledge can be transferred from the model to non-model organisms and insights can be gained in the structural and evolutionary patterns of specific genes. In the absence of sequenced genomes for allergenic grasses, this study was aimed at understanding the structure, organisation and expression profiles of grass pollen allergens using the genomic data from *Brachypodium distachyon* as it is phylogenetically related to the allergenic grasses. Combining genomic data with the anther RNA-Seq dataset revealed 24 pollen allergen genes belonging to eight allergen groups mapping on the five chromosomes in *B. distachyon*. High levels of anther-specific expression profiles were observed for the 24 identified putative allergen-encoding genes in Brachypodium. The genomic evidence suggests that gene encoding the group 5 allergen, the most potent trigger of hay fever and allergic asthma originated as a pollen specific orphan gene in a common grass ancestor of Brachypodium and Triticiae clades. Gene structure analysis showed that the putative allergen-encoding genes in Brachypodium either lack or contain reduced number of introns. Promoter analysis of the identified Brachypodium genes revealed the presence of specific *cis*-regulatory sequences likely responsible for high anther/pollen-specific expression. With the identification of putative allergen-encoding genes in Brachypodium, this study has also described some important plant gene families (e.g. expansin superfamily, EF-Hand family, profilins etc) for the first time in the model plant Brachypodium. Altogether, the present study provides new insights into structural characterization and evolution of pollen allergens and will further serve as a base for their functional characterization in related grass species.

Introduction

Grass pollens constitute a major worldwide source of airborne allergens and are known to be responsible for triggering allergic rhinitis and exacerbating asthma [1]. Grasses belong to the family Poaceae (Gramineae) and are grouped with the sedges, rushes, and other monocots belonging to the order Poales. The family Poaceae is the fourth largest family of flowering plants with more than 600 genera and 10,000 species. The three important subfamilies of Poaceae, namely the subtropical Chloridoideae and Panicoideae, and the temperate Pooideae are known to constitute the major bulk of allergenic pollen [2, 3]. *Cynodon dactylon* (Bermuda grass), a popular southern turf grass, is recognized as the chief allergenic grass species of Chloridoideae [2–4]. Likewise, the Panicoideae subfamily mainly includes *Sorghum halepense* (Johnson grass) and *Paspalum notatum* (Bahia grass) as significant allergenic members [2–5]. Notably, the temperate Pooideae subfamily represents the majority and pivotal allergenic grasses such as *Lolium perenne* (ryegrass), *Phleum pratense* (timothy grass), *Holcus lanatus* (velvet grass), *Phalaris aquatica* (canary grass), *Poa pratensis* (bluegrass) and *Dactylis glomerata* (orchard grass) [2, 3]. In the recent years, a member of the same subfamily (Pooideae), *Brachypodium distachyon* (Brachypodium) has been recognised as a model plant for structural and functional genomics research in grasses. It is the first member to be sequenced within the Pooideae subfamily and shares a high degree of synteny with other grasses.

Grass pollen allergenicity can be attributed to certain proteins which are rapidly released from the pollen grain upon hydration [1, 6]. These proteins possess multiple IgE-reactive epitopes that can bind cross-link adjacent IgE molecules leading to degranulation of mast cells leading to clinical symptoms of allergy. Allergens (proteins) identified in one species often have homologues in other evolutionary related species, exhibiting similar physicochemical and immunological properties [1, 7]. The conserved IgE epitopes on these homologous proteins lead to allergenic cross-reactivity, implying that allergic individuals sensitised to an allergen can show allergenic reactivity to related allergen from a different species.

Pollen allergens found in different grass species which exhibit structural similarities and allergic cross-reactivities have been classified as group 1, 2, 3, 4, 5, 6, 7, 11, 12, 13, 22, 24 allergens (Table 1) [8–19]. Out of these, group 1 and group 5 allergens were found to be the most clinically relevant, since 95 and 65–85% of patients, respectively, are reported to be sensitised to them [1, 4]. Together group 1 and five allergens account for more than 80% of total IgE reactivity of grass pollen proteins.

The distribution of allergens across the different grasses varies: some allergens are shared across the entire plant kingdom (e.g. group 12) or between pollen plants (e.g. group 7). Somes allergens are restricted to Poaceae (group 1) or Pooideae (group 5). Group 4, 7, 11 and 12 allergens are not grass-specific while group 12 allergen (profilin) is present in all eukaryotic cells. However, no comprehensive study focussing on origin and evolution of pollen allergen families in the genome of allergenic grass species is available to date. Availability of fully sequenced genome for pooid grasses such as ryegrass and timothy grass can significantly facilitate addressing questions such as allergen sequence homology and evolutionary diversity of the isoforms.

The recent availability of the genomic sequence of *Brachypodium distachyon* [20–24] has provided a powerful genomic resource to further our understanding grass pollen allergens repertoire. The availability of protocols for *Brachypodium* transformation will further allow addressing of allergen-encoding gene function through RNAi or gene editing technologies (CRISPR). Also, because of the phylogenetic relatedness and extensive synteny shared among grass genomes, the genetic information from Brachypodium would be beneficial for understanding the structure, organisation and evolution of pollen allergens in clinically significant grass genera.

Table 1. Overview of characteristics of grass pollen allergens [8–19].

Allergen	Allergen nature and characteristics
Group 1 (31–35 kDa)	Acidic glycoproteins, exist in multiple isoforms distinguished by their respective isoelectric points (pIs), apparent molecular weights and amino acid sequences, belong to a subfamily of structurally related proteins called β-expansins, which are cell-wall loosening proteins.
Group 2/3 (10–12 kDa)	Group 2 (acidic) and group 3 (basic) are non-glycosylated proteins, share a high degree of sequence homology (40% identity, 60–70% similarity) with the C-terminal part of group 1 allergens that suggest a common origin by gene duplication.
Group 4 (50–60 kDa)	High molecular weight basic glycoproteins, sequence similarity to the berberine bridge enzymes (BBE), a member of a flavoprotein oxidoreductase superfamily.
Group 5 (27–33 kDa)	A heterogenous group of non-glycosylated proteins with multiple isoforms varying in pIs and primary sequences, examined in only Pooideae subfamily of grasses and antibodies raised against group 5 allergens failed to detect cross-reactive antigens outside of the subfamily.
Group 6 (11 kDa)	Acidic, non-glycosylated proteins, so far been detected in Timothy grass pollen extracts only.
Group 7 (8–12 kDa)	Small proteins belonging to the polcalcin family of calcium-binding proteins with homologues present in pollen of all monocots and dicots.
Group 11 (16–20 kDa)	Glycoproteins, belonging to the soybean-trypsin-inhibitor-like family, structurally similar to pollen allergens from the olive tree (Ole e 1).
Group 12 (14 kDa)	Members of the conserved profilin protein family present in all eukaryotic cells.
Group 13 (50–60 kDa)	Members of a polygalacturonase like protein family.
Group 22 (9 kDa)	Enolase, so far only identified in Bermuda grass pollen extract.
Group 24 (21 kDa)	Pathogenesis-related protein (PR-1), so far only identified in Bermuda grass pollen extract.

doi:10.1371/journal.pone.0169686.t001

Thus in the present study, an extensive genome-wide analysis was performed to (i) identify putative allergen-encoding genes in *B. distachyon* genome, (ii) analyze gene structures and protein domains, (iii) examine the phylogenetic and evolutionary relationships between *B. distachyon* putative allergens and other related grasses, (iv) identify potential promoters targeting pollen allergen genes and (v) investigate the expression patterns of putative allergen-encoding genes in Brachypodium using available RNA-Seq data.

Materials and Methods

Sequence retrieval and identification of conserved domains in different groups of pollen allergens

A complete set of allergens sequences has been reported from ryegrass (*Lolium perenne* L.) and Timothy grass (*Phleum pratense* L.) [1, 4]. Therefore, protein sequences from these two grass species were selected as representative members and were later used as seed sequences to search against *B. distachyon* proteome database. Lol p 1 (P14946) for group 1, Lol p 2 (P14947) for group 2/3, Lol p 4 (Q5TIW3) for group 4, Lol p 5 (Q40240) for group 5, Phl p 7 (O82040) for group 7, Lol p 11 (Q7M1X5) for group 11, Phl p 12 (P35079) for group 12 and Phl p 13 (CAB42886) for group 13 were chosen as representative sequences. All the protein sequences were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) were listed in S1 Table. Conserved domains were identified in each pollen allergen group using SDAP (Structural database of allergenic proteins; http://fermi.utmb.edu/SDAP) and further verified using SMART (Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de) and Pfam (http://pfam.sanger.ac.uk/). Schematic diagrams of pollen allergens belonging to different groups

with conserved domains were constructed using DOG 2.0 (Domain Graph, version 2.0) (http://dog.biocuckoo.org/) [25].

Identification of putative allergen-encoding genes in *B. distachyon*

Representative pollen allergen sequences of ryegrass (*Lolium perenne* L.) and Timothy grass (*Phleum pratense* L.) were searched against *B. distachyon* proteome database (v2.1 available at Phytozome (v10.1, http://www.phytozome.net/)) using Basic Local Alignment Search Tool algorithms (BLASTP) under default parameters. The resultant top hits from BLASTP searches of *B. distachyon* were analysed for the presence of exemplar conserved domains using SMART and Pfam. The presence of signal peptides was predicted by Signal P4.1 (http://www.cbs.dtu.dk/services/SignalP/) and N-glycosylation sites by NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/).

Multiple sequence alignment and phylogenetic analysis

Protein sequences of *B. distachyon*, *Lolium perenne* and *Phleum pratense* retrieved above were aligned using the multiple sequence alignment (MSA) tool CLUSTALX2.0, using Gonnet Protein Weight Matrix with default parameters and phylogenetic trees were constructed by using neighbour-joining (NJ) method. The stability of nodes was tested by bootstrap analysis with 100 replicates and trees were viewed using MEGA v6.0 [26].

Chromosomal location and gene structure

The information about the chromosomal location of allergen-encoding genes, i.e. chromosome number, start position, end position and gene orientation in Brachypodium was retrieved from the *B. distachyon* genome information file (v2.1). This information was used to locate the genes on a schematic map. Further, the exon-intron organisation of these genes was deduced using GSDS (Gene structure display server; http://gsds.cbi.pku.edu.cn/) via alignment of the cDNAs with their corresponding genomic DNA sequences [27].

Promoter analysis and homology modelling

The upstream DNA sequences of putative allergen-encoding genes were obtained from the *B. distachyon* genome dataset (v2.1) for locating known *cis*-regulatory elements (CREs) using SIGNALSCAN program available in Plant *cis*-Regulatory DNA elements (PLACE) (http://www.dnd.affrc.go.jp/PLACE/) database [28]. Further, protein structure determination by homology modelling was performed using Phyre2 (Protein Homology/AnalogY Recognition Engine V 2.0; http://www.sbg.bio.ic.uk/phyre2) under 'intensive' mode [29].

Expression profiling using published RNA-Seq dataset

To elucidate the expression profiles of putative allergen-encoding genes in *B. distachyon*, Illumina RNAseq reads from 9 distinct tissues including immature leaves, pre-and post-emergence flowers, anther, pistil, whole seed at 5 DAP (days after pollination), whole seed at 10 DAP, embryo at 25 DAP and endosperm at 25 DAP, were downloaded from the Sequence Read Archive (SRA) at the National Centre for Biotechnology Information (SRP008505) [30]. RNA-Seq reads in the fastq format were then used for mapping against the latest version of *B. distachyon* transcripts (v2.1) using RSEM v.1.2.19 and Bowtie v0.12.8 under default settings [31]. RSEM outputs abundance estimates and transcripts expression levels were given as FPKM (fragments per kilobase of exon per million mapped fragments) values. The FPKM values of each transcript from replicate samples of immature leaves and embryo at 25 DAP were

averaged to obtain the final value and transcripts with FPKM greater than or equal to one were considered as expressed. Heat maps were generated to represent the gene expression ratio in percentage with FPKM value shown in the most highly expressed sample. To visualise the gene expression clusters among homologues and /or paralogs, heat maps were plotted against the phylogenetic tree constructed using *B*. *distachyon* protein sequences.

Results

Identification of conserved domains in pollen allergen groups

Ryegrass and Timothy grass protein sequences belonging to eight different allergen groups were chosen as representative members and the conserved domains within each group of pollen allergens were identified. Fig 1 shows the presence of different conserved domains within each group of pollen allergens. This information was utilised for identifying allergen homologs in *B*. *distachyon* by presence of conserved domains. Group 1 allergens are characterised with a Rare lipoprotein-A (RlpA)-like double-psi beta-barrel (DPBB_1) domain and a pollen allergen domain while group 2/3 allergens contain a single pollen allergen domain. Group 4 allergens include both FAD binding and BBE domains. Group 5 allergens encode two ribonuclease (pollen allergen) domains. Grass pollen proteins of group 7 are characterised by two EF-Hand domains and group 11 with an Ole e one domain respectively. Similarly, Group 12 allergens include a profilin domain while group 13 allergens contain glycosyl hydrolases family 28 domain.

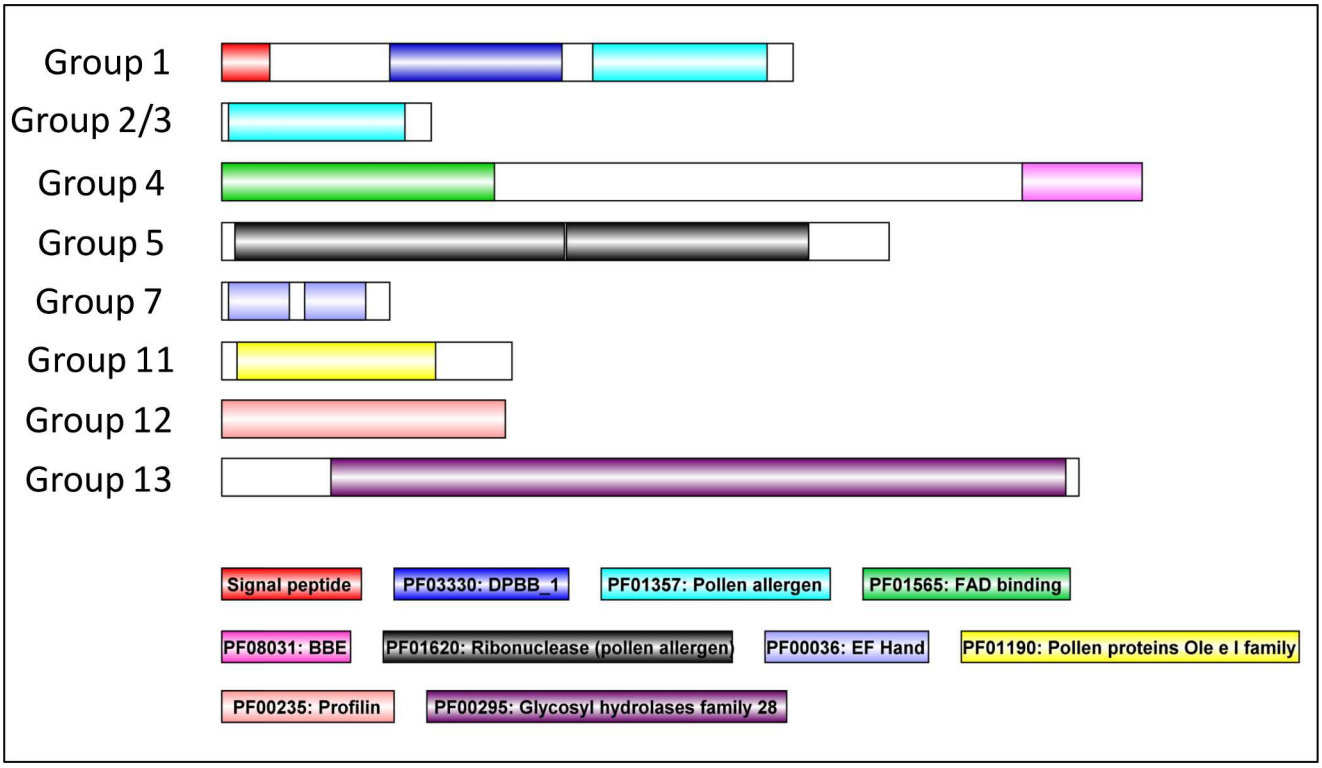


Fig 1. Protein domain architecture of representative allergen proteins from each group. The protein domains in pollen allergen groups are identified by SDAP (Structural Database of Allergenic Proteins) database and verified using the SMART (http://smart.embl-heidelberg.de) and PFAM (http://pfam.sanger.ac.uk/). Lol p 1 (P14946) for group 1, Lol p 2 (P14947) for group 2/3, Lol p 4 (Q5TIW3) for group 4, Lol p 5 (Q40240) for group 5, Phl p 7 (O82040) for group 7, Lol p 11 (Q7M1X5) for group 11, Phl p 12 (P35079) for group 12 and Phl p 13 (CAB42886) for group 13 were chosen as representative members respectively.

doi:10.1371/journal.pone.0169686.g001

Identification, characterization and expression profiles of putative homologs of pollen-allergen genes in *B. distachyon*

A total of 257 Brachypodium proteins were identified using ryegrass and timothy grass representative allergen protein sequences as queries against Brachypodium genome dataset. All the identified 257 proteins were then analysed to confirm the presence of conserved protein domains as identified earlier in different groups of pollen allergens. The phylogenetic relatedness of the identified 257 Brachypodium proteins was further examined by MSA and a Neighbour-Joining phylogenetic tree (Fig 2). The proteins were categorised into eight different groups or protein families (Expansins, group 2/3 allergens, FAD-binding and berberine bridge enzymes, group 5 allergens, EF-hand calcium binding proteins, pollen proteins Ole e one like profilins and polygalacturonases) indicating that all the groups of grass pollen allergens are represented in *B. distachyon* genome.

Expansin protein family in *B. distachyon* and Group 1 allergens. In total, 64 proteins were identified as Expansins by BlastP search of Lol p 1 (P14946, ryegrass group 1 allergen) as a query against Brachypodium proteome database. S1 Fig shows phylogenetic relationships, exon/intron structure, domain architecture and expression profile of the expansin protein

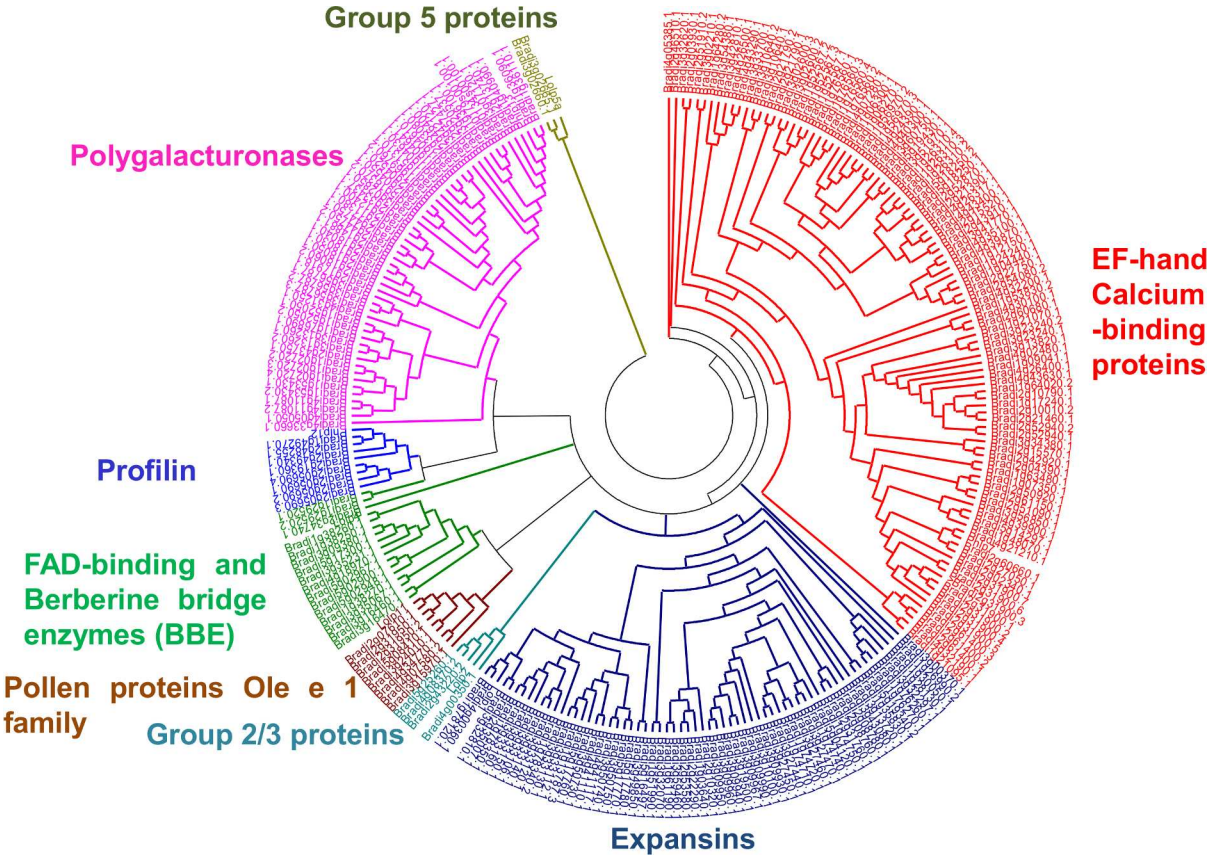


Fig 2. Phylogenetic analysis of pollen allergen protein families in *B. distachyon* and representative member of each group from Ryegrass and Timothy grass. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed. The tree was divided into eight phylogenetic clusters. The members were distinctly coloured to represent respective protein family. Lol p 1 (P14946) for group 1, Lol p 2 (P14947) for group 2/3, Lol p 4 (Q5TIW3) for group 4, Lol p 5a (Q40240) for group 5, Phl p 7 (O82040) for group 7, Lol p 11 (Q7M1X5) for group 11, Phl p 12 (P35079) for group 12 and Phl p 13 (CAB42886) for group 13 were chosen as representative members respectively.

doi:10.1371/journal.pone.0169686.g002

family in *B. distachyon*. The proteins were further categorised into three different sub-classes of expansins (β-expansins, expansin-like and α-expansins) by their functional annotations. The largest subgroup was categorised as α-expansins, with 32 members as shown in clade 3 (green shade). The second largest subgroup was β-expansins, containing 26 members clade 1 (red shade). The smallest subgroup was expansin-like proteins that included six members in clade-2 (blue shade). As group 1 allergens belong to the family of β-expansins, we identified four Brachypodium proteins (Bradi1g78120.1, Bradi4g00360.1, Bradi3g32810.1 and Bradi2g35007.1) as putative group 1 allergens, closely clustered with Lol p 1 and exhibited anther-specific expression as compared to other vegetatively expressed β-expansins. Protein sequence identities of Brachypodium group 1 isoforms with known group 1 allergens from other grasses range between 53% and 89% (S2 Table). Bradi1g78120.1 bears closest sequence relationship in the grasses to a Velvet grass group 1 allergen (Hol l 1) showing 100% sequence coverage and 82% sequence identity. To appropriately identify group 1 allergen isoforms, MSA was carried out for four identified group 1 isoforms in Brachypodium and known group 1 proteins from other allergenic grasses (S2 Fig). Alignment results highlighted the conservation of typical structural characteristics of group 1 allergens [1] in all four isoforms of Brachypodium that included an N-glycosylation site at amino acid position 9, hydroxyproline residues at positions 5 and 8 and seven conserved cysteine residues. Sequence alignments also showed the conservation of IgE-binding residues in Group 1 allergens and Brachypodium homologs (S2 Fig) [32]. Thus, we conclude, four Brachypodium transcripts being transcribed by putative group 1 allergen-encoding genes (S1 Fig).

Group 2 allergens. We identified four proteins (Bradi5g25260.1, Bradi4g08370.2, Bradi2g43220.2 and Bradi4g00350.1) as group 2/3 allergens by protein domain architecture and anther-specific expression profiles of their corresponding transcripts (S3 Fig). S4 Fig illustrates the sequence alignments of the four identified group 2/3 homologs in Brachypodium with other known group 2/3 allergens. A previous study in Timothy grass defined the three-dimensional (3D) structure of a complex between Phl p 2 and a Phl p 2-specific Fab derived from an IgE antibody. The IgE-binding epitope comprised nine critical residues grouped into four sequentially distant segments of the allergen molecule [33]. We observed that all the nine purportedly critical residues were conserved in three Brachypodium proteins (Bradi5g25260.1, Bradi4g08370.2 and Bradi2g43220.2) while Bradi4g00350.1 exhibited seven out of nine epitope critical residues (S4 Fig). Protein sequence identities of Brachypodium group 2/3 isoforms with group 2 allergens range between 53% and 66% (S2 Table).

FAD-binding/Berberine Bridge Enzymes (BBE) protein family in *B. distachyon* and Group 4 allergens. BlastP search with Lol p 4 sequence identified 16 proteins in *B. distachyon* proteome that belong to a family of FAD binding and berberine bridge enzymes (BBE) proteins by the presence of conserved Pfam domains (PF01565 and PF08031) that also included group 4 allergens. S5 Fig shows their phylogenetic relationships, exon/intron structure, domain architecture and their transcript expression profiles. Since group 4 allergens exhibited sequence similarity to FAD binding and/or berberine bridge enzymes (BBE) proteins, we found only one isoform of group 4 in Brachypodium (Bradi1g38260.1) whose corresponding transcript showed anther- specific expression pattern and was grouped closely with Lol p 4 in the phylogenetic tree (S5 Fig). S6 Fig shows MSA of Bradi1g38260.1 and group 4 proteins from other grasses. The results displayed the presence of N-glycosylation sites and conserved residues among group 4 proteins. Bradi1g38260.1 protein sequence showed 99% sequence coverage and 81% sequence identity to Sec c 4 and Tri a 4 allergens respectively (S2 Table).

Group 5 allergens. Using ryegrass group 5 allergen (Lol p 5; Q40240) as a query against Brachypodium proteome dataset, only three Brachypodium proteins (Bradi3g02660.1, Bradi3g02665.1 and Bradi2g48255.1) were revealed to possess conserved group 5 allergen protein

domains (PF01620). Further, expression profiles of the corresponding transcripts also revealed high anther-specific expression pattern (S7 Fig). S8 Fig exhibits MSA of the identified group 5 isoforms in Brachypodium with known group 5 proteins from other allergenic grasses of the Pooideae subfamily. The alignments revealed the presence of conserved IgE-binding residues [34] among different group 5 proteins and Brachypodium homologs (S7 Fig). Protein sequence identities of Brachypodium group 5 isoforms with other known group 5 allergens vary greatly (from 43% to 77%) as typical of sequence variation in group 5 allergens (S2 Table).

Calcium-binding EF-hand family proteins and group 7 allergens. BlastP search of Timothy grass group 7 allergen (Phl p 7; O82040) against *B. distachyon* proteome identified 100 proteins that contained at least one EF-hand domain (PF00036). (S9 Fig). Interestingly, only one protein (Bradi4g21210.1) clustered closely with Phl p 7 in the first clade of the phylogenetic tree and thus, identified as group 7 isoform in Brachypodium. Further, the expression profile of Bradi4g21210.1 confirmed anther specificity of this group 7 homolog. (S9 Fig). S10 Fig shows the sequence alignments of the identified group 7 homolog in Brachypodium with Timothy grass (Phl p 7; O82040) and Bermuda grass (Cyn d 7; P94092) group 7 proteins. The alignment results highlighted the presence of two conserved EF-hand domains in all the protein sequences. Bradi4g21210.1 showed 100% protein sequence coverage and 94% sequence identity with a Cyn d 7 pollen allergen (S2 Table).

Pollen proteins ole e 1 protein family and group 11 allergens. We identified 11 proteins belonging to Ole e 1 family based on the presence of protein domain (PF01190) as observed in ryegrass group 11protein (Lol p 11; Q7M1X5) (Fig 1). S11 Fig shows phylogenetic relationships, exon/intron structure, domain architecture and expression profile of Ole e 1 protein family in *B. distachyon*. Four proteins (Bradi2g07610.1, Bradi2g31950.1, Bradi2g31950.2 and Bradi5g08410.1), comprising the first clade of phylogenetic tree were grouped closely with Lol p 11 and their corresponding transcripts also exhibited anther-specific expression pattern. The four homologs thus are identified as group 11 proteins in Brachypodium. Sequence similarities of group 11 isoforms in *B. distachyon* with ryegrass (Lol p 11; Q7M1X5) and Timothy grass (Phl p 11; AAN32987) ranges between 45% and 88% (S2 Table). MSA of Lol p 11, Phl p 11 and Brachypodium group 11 isoforms marked the presence of conserved amino acid residues in group 11 allergens with a single N-linked glycosylation site at residue 24 and six conserved cysteine residues (S12 Fig).

Profilin protein family in *B. distachyon* and group 12 allergens. Group 12 allergens are the members of profilin protein family and BlastP result of Timothy grass group 12 allergen (Phl p 12; P35079) as queried against Brachypodium proteome identified seven proteins belonging to this family. We categorised two homologs (Bradi2g19360.1 and Bradi2g49340.1) that are grouped together with Phl p 12 in the first clade of the phylogenetic tree as group 12 proteins in *B. distachyon* by the clear demarcation of the anther-specific expression profile of the two corresponding transcripts (S13 Fig). S2 Table shows protein sequence coverage and identity of group 12 homologs from Brachypodium and other grasses. Sequence similarities of Brachypodium group 12 homologs with known group 12 proteins were observed with 100% coverage and 84–92% identity. S14 Fig shows the MSA of two identified group 12 homologs in Brachypodium with group 12 proteins from other grasses. The alignment highlighted the presence of cysteine residues at two conserved positions.

Polygalacturonases/Glycosyl hydrolases family 28 and group 13 allergens. Group 13 allergens belong to the family of pectin-degrading enzymes (i.e., polygalacturonases). BlastP analysis with Timothy grass group 13 allergen (Phl p 13; CAB42886) sequence as a query against Brachypodium proteome identified 52 homologs that contain the conserved Glycosyl hydrolases family 28 domain (PF00295) as observed in Phl p 13 (Fig 1). S15 Fig shows phylogenetic relationships, exon/intron structure, domain architecture and expression profile of

polygalacturonases family in *B. distachyon*. On the basis of the anther-specific expression pattern of their corresponding transcripts, five candidate genes (Bradi1g36110.1, Bradi1g36090.1, Bradi1g40990.1, Bradi2g13740.1 and Bradi3g07120.1) were identified as encoding group 13 allergens in Brachypodium. MSA of five Brachypodium homologs and group 13 proteins from Bermuda grass and maize (Phl p 13, Zea m 13) suggested the conservation of polygalacturonase domains, 14 cysteine residues at conserved positions and N-glycosylation sites (S16 Fig). S2 Table shows protein sequence identities of Brachypodium group 13 homologs with Phl p 13 and Zea m 13 that range between 58% and 85%.

Comparative analysis of pollen allergen homologs in *B. distachyon* and other allergenic grasses

In total, we have identified 24 homologs of pollen allergens in *B. distachyon* belonging to eight different groups. To investigate the molecular evolution of pollen allergens from various allergenic grasses and their homologs in Brachypodium, we constructed a phylogenetic tree using the amino acid sequences (Fig 3). The sequences of all the proteins used are given in S1 Table. Analysis of the tree showed that group 4 and group 12 have clustered together but possessed different domains. A similar observation was made with group 11 and group 13 that were branched into a single clade, but both groups encoded different domains. This observation suggests a greater sequence divergence in some of the allergen groups during evolution.

Expression profile of pollen allergen-encoding transcripts of *B. distachyon*

A complete gene expression catalogue of *B. distachyon* in 9 different tissues is given in S3 Table. Fig 4 illustrates the transcript levels of 24 putative allergen-encodings genes in anther tissues of *B. distachyon*. All these 24 genes show strict anther-specificity with some signal in pistil tissues that can be ascribed to RNA contribution from pollen in stigma/styler tissues. The putative isoforms of group 5 (Bradi3g02660.1, Bradi3g02665.1) proteins identified in Brachypodium showed the highest expression values (FPKMs-3863, 3737 respectively).

Chromosomal distribution and homology modeling

The precise locations of putative allergen-encoding genes on five chromosomes of *B. distachyon* have been illustrated in Fig 5 and their exact positions on *B. distachyon* chromosomes are given in S4 Table. Among all, chromosome 2 contained the highest number of allergen-encoding genes (nine) that were unevenly distributed, followed by chromosome 1 and chromosome 3, each containing five genes. Chromosome 4 carried four genes while two genes were located on chromosome 5 (Fig 5). Interestingly, homologs of group 12 were located on chromosome 2 while other groups' homologs were found to be distributed unevenly among the five chromosomes.

To better understand structural characteristics of the allergen proteins, the tertiary structures of *B. distachyon* putative pollen allergen proteins were predicted by the Phyre2 server in intensive mode. Fig 6 shows the protein structures that are modelled at >90% confidence. The protein structures clearly differentiate various pollen allergen groups. The secondary structures of putative allergen homologs in Brachypodium predominantly comprised of β-sheets and coils with lesser number of α-helices except for group 5 proteins. As inferred from the predicted tertiary structures (Fig 6), group 1 homologs were shown to form dimers while group 2 homologs contain anti parallel β-strands, possibly forming a β-sandwich. Group 5 proteins were seen as mainly α-helical with each monomer containing a 4-helix bundle. The predicted

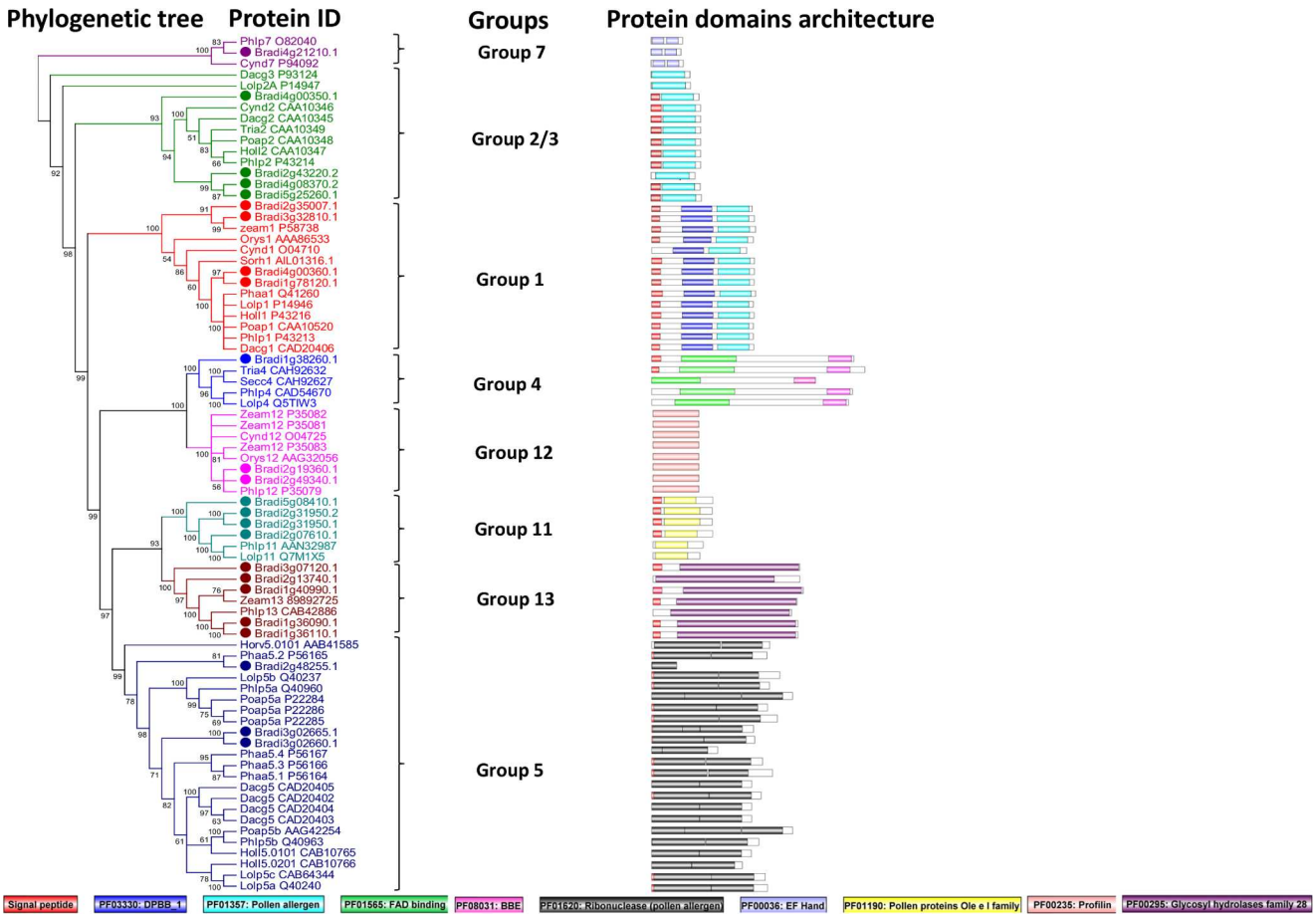


Fig 3. Phylogenetic tree and domain architecture of pollen allergens. Neighbour-Joining phylogenetic tree is constructed based on protein alignments of putative allergen-encoding proteins of *B. distachyon* and known pollen allergens from other grasses using ClustalX2. Bootstrap values greater than 50% are shown at the nodes. The schematic diagrams show the domain organization of these proteins as obtained from SMART and PFAM searches. Different domains are indicated by the use of different colours as shown at the bottom of the figure. The accession numbers for the sequences used in the alignment are also listed in the figure.

doi:10.1371/journal.pone.0169686.g003

structures of group 12 proteins (profilins) in Brachypodium were shown to form a main fold comprising of central anti parallel β-sheets and α-helices on both sides. Group 13 proteins were shown to form a central β barrel having several β-strands, with the α-helix situated on each side of the barrel.

Cis-regulatory elements (CREs) of putative allergen-encoding genes of *B. distachyon*

A key feature of the 24 genes identified as encoding pollen allergens in the Brachypodium genome is their anther/pollen specific expression. The availability of genomic sequences corresponding to these allergen genes presents an opportunity to identify conserved motifs in their upstream sequences. Evaluation of genomic sequences upstream of the initiation codon of the 24 Brachypodium putative allergen-encoding genes resulted in the identification of 198 types of putative CREs (S5 Table). Fourteen of the CREs were found in all of the 24 genes: ARR1AT (5′-NGATT-3'), CAATBOX1 (5′-CAAT-3′), CACTFTPPCA1 (5′-YACT-3′), DOF-COREZM (5′-AAAG-3′), EBOXBNNAPA (5′-CANNTG-3′), EECCRCAH1 (5′-GANTTNC-3′), GATABOX (5′-GATA-3′), GT1CONSENSUS (5′-GRWAAW-3′),

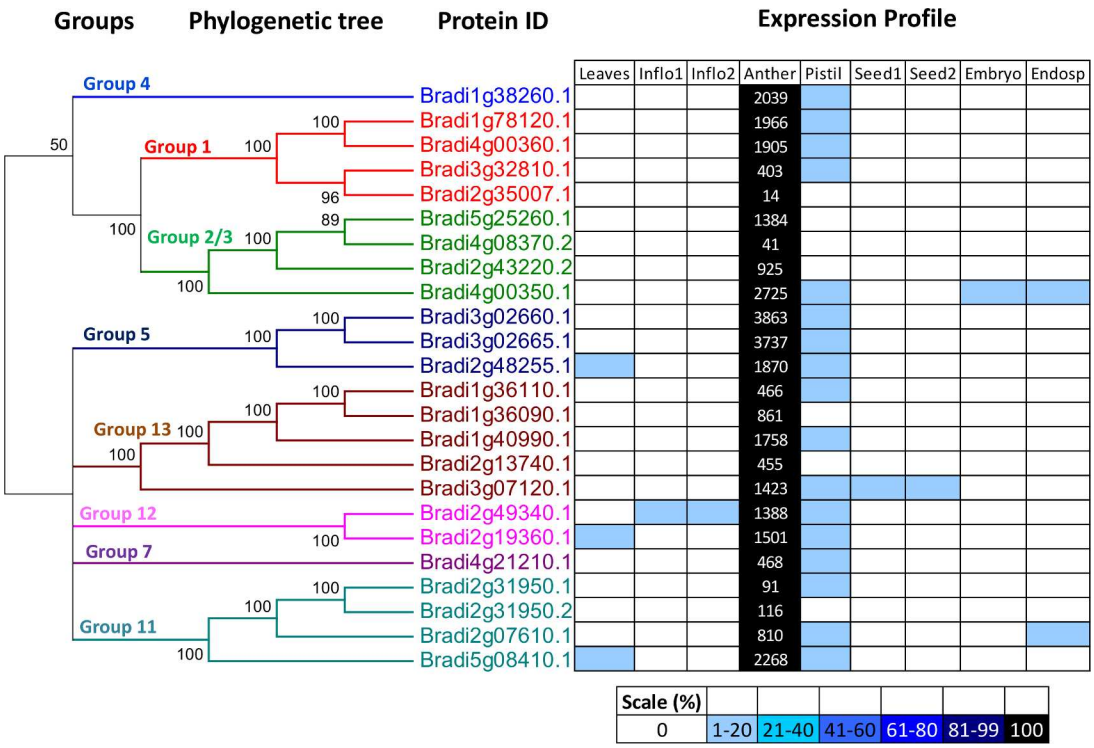


Fig 4. Expression profile and evolutionary pattern of pollen allergens. The heat map shows the relative abundances of putative allergen-encoding genes identified in *B. distachyon*. The level of expression for a gene across different samples are represented as percentage of the maximum expression level in colour code from 0% (white) to 100% (black). Heat map was plotted against the phylogenetic tree constructed using *B. distachyon* protein sequences. The diagram also shows relative expression values in anther tissue of *B. distachyon* based on RNA-Seq data analysis obtained from nine developmental stages. Inflo1 is for pre-emergence flowers stage, Inflo2 is for post-emergence flowers stage, Seed1 is for whole seed at 5days after pollination, Seed2 is for whole seed at 10 days after pollination and Endosp is for endosperm respectively.

doi:10.1371/journal.pone.0169686.g004

GTGANTG10 (5′ – GTGA-3′), MYCCONSENSUSAT (5′ – CANNTG-3′), POLLEN1LELAT52 (5′ – AGAAA-3'), RAV1AAT (5′ – CAACA-3′), WBOXNTERF3 (5′ – TGACY-3′) and WRKY71OS (5′ – TGAC-3′). The duplication frequency of these CREs in all 24 genes is depicted in Fig 7, and the numbers are given in S5 Table. Among these fourteen CREs, CACTFT PPCA1 is the most abundant CRE, with duplications in the range of 6 to 20 in the 1-kb upstream region of each putative allergen-encoding promoter, followed by DOFCOREZM, MYCCONSENSUSAT and EBOXBNNAPA respectively. In addition to these, some CREs were present uniquely to specific pollen allergen groups such as QELEMENTZMZM13 (5′ – AGGTCA-3′) in group 1, SPHCOREZMC1 (5′ – TCCATGCAT-3') in group 2/3, PE2FNTRNR1A (5′ – ATTCGCGC-3′) in group 4, REBETALGLHCB21 (5′ – CGGATA-3′) and TATAPVTRNALEU (5′ – TTTATATA-3′) in group 5, T/GBOXATPIN2 (5′ – AACGTG-3′) and QARBNEXTA (5′ – AACGTGT-3′) in group 12 respectively. Our results show conservation of putative regulatory sequences in the upstream regions of genes encoding allergens belonging to diverse groups. This provides useful data for the functional characterization of regulatory elements underlying pollen specific expression of allergen-encoding genes.

Discussion

Grasses have evolved a group or family of allergen encoding genes that appear to be largely restricted to the species of Poaceae. Since genomic data for important allergenic grasses in the

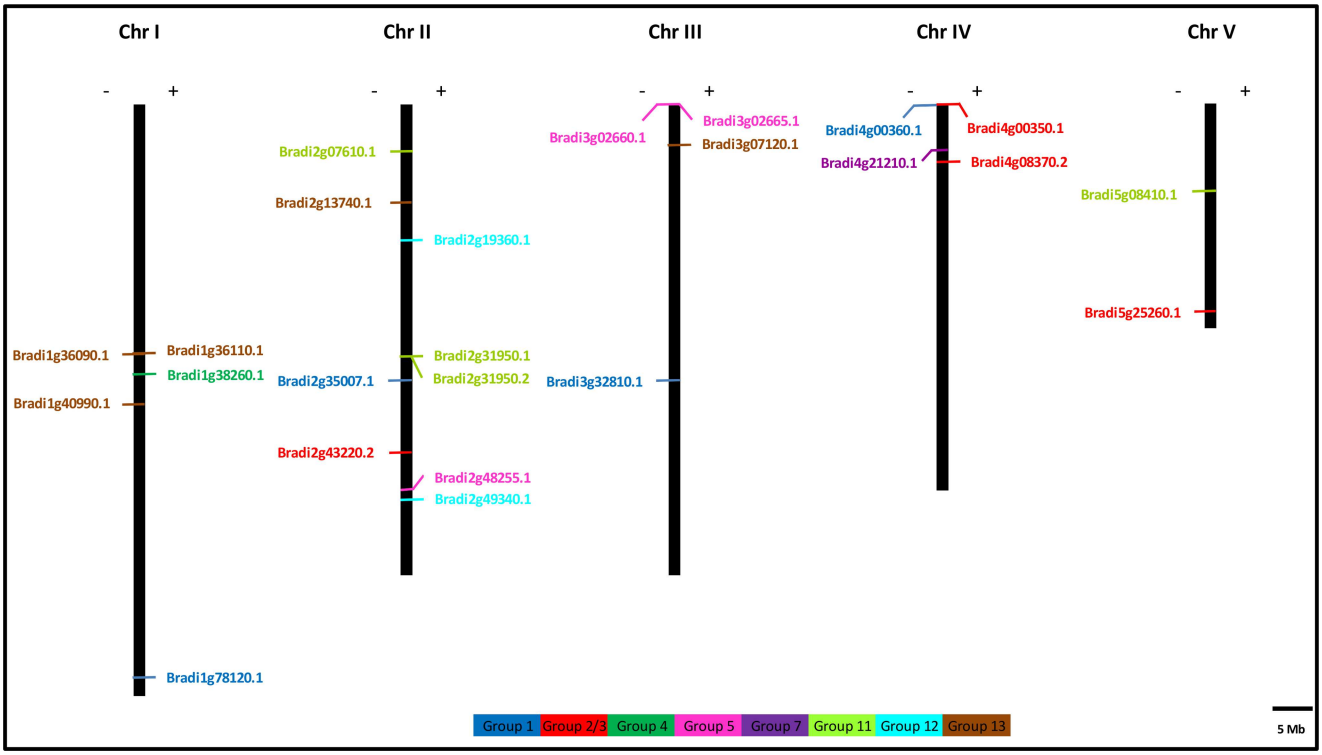


Fig 5. Distribution of 24 putative allergen-encoding genes onto five *B. distachyon* chromosomes. The chromosomal position of each gene was mapped according to the *B. distachyon* genome. The chromosome number is indicated at the top of each chromosome. The scale is 5 Mb. Genes are marked on the left and right hand side of each chromosome based on their orientation. Genes belonging to different groups of pollen allergens are indicated by the use of different colours as shown at the bottom of the figure.

doi:10.1371/journal.pone.0169686.g005

family is still unavailable, the genetic information contained in the sequenced genome of model grass *B. distachyon* provides a valuable approach to disclose pollen allergen-encoding gene repertoire in grasses and trace their expression profiles and evolutionary relationships. However, some reports on transcriptome- and proteome-based examination of putative pollen allergens in important grasses/cereals are present in the literature [5, 18, 35–39]. Rice was the first member of the angiosperm family Poaceae to be sequenced. An investigation of the rice genome and pollen transcriptome revealed the lack of group 5 allergen genes in this species [37] making it unsuitable as a model for investigating allergen gene repertoire of clinically important pooid grasses such as Ryegrass, Fescue and Timothy grass, etc. Though Brachypodium is not a clinically significant grass because of its short stature and limited pollen production, being a pooid grass, its genome and transcriptome provides a first window for revealing allergen gene repertoire investigating evolutionary and structural biology of grass pollen allergens.

In this study, we have successfully identified 24 putative allergen-encoding proteins in *B. distachyon* using genome- and transcriptome-based analysis. To further describe the identified Brachypodium genes, phylogenetic trees and heat maps were constructed which define their evolutionary relationships and depict expression profiles across 9 Brachypodium tissues (Figs 3 and 4). Phylogenetic trees displayed high bootstrap values for internal nodes of different clades and hence, it clearly showed the derivation of statistically reliable pairs of possible homologous proteins sharing similar functions from a common ancestor. The above identified 24 putative allergen-encoding proteins were categorised into eight independent pollen allergen groups

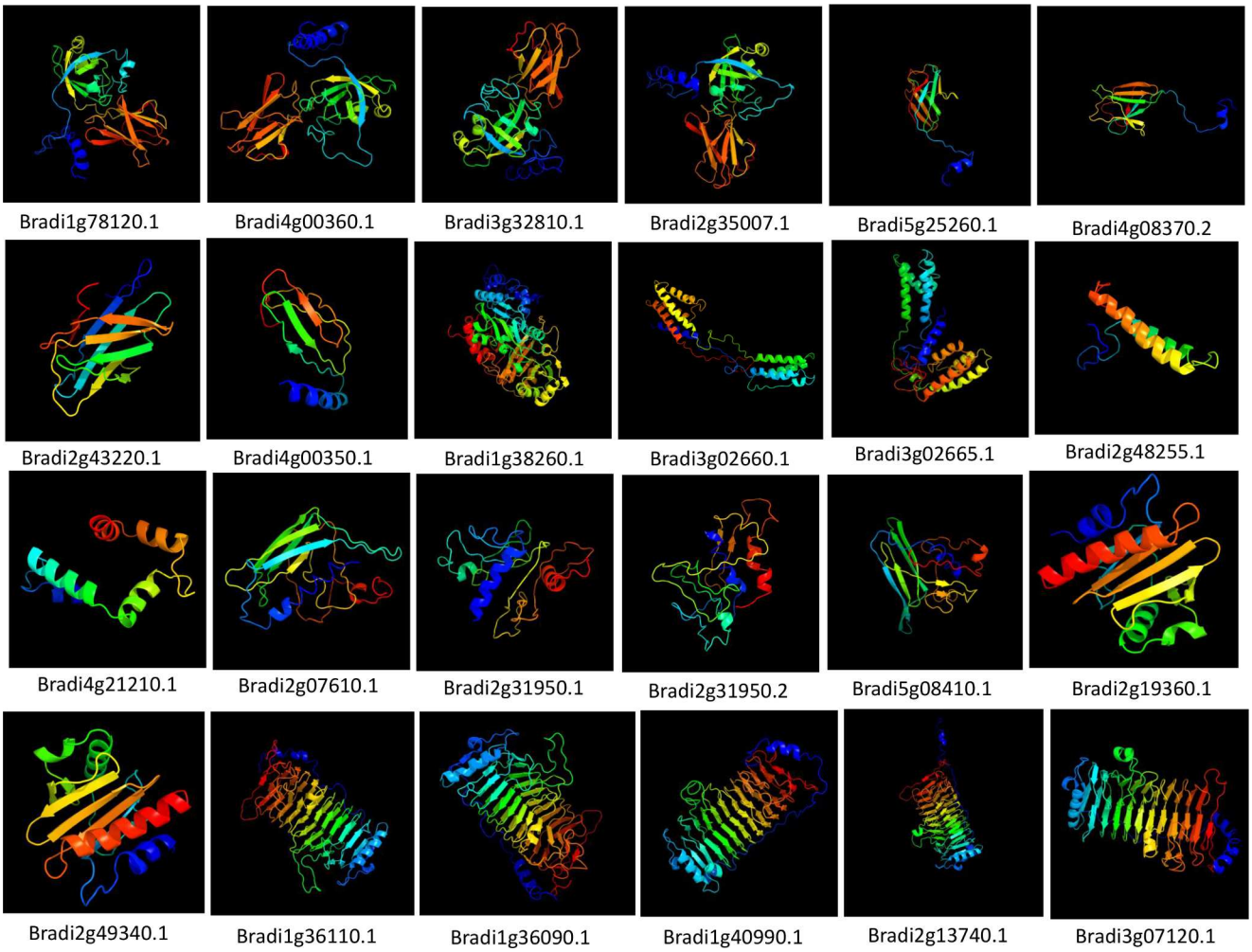


Fig 6. Predicted structures of putative pollen allergen proteins in *B. distachyon*. The structure of 24 putative allergen-encoding proteins with >90% confidence level were shown.

doi:10.1371/journal.pone.0169686.g006

depending on the presence of specific domains in their protein sequences (Fig 1). In addition to the identification of putative allergen-encoding genes in *B. distachyon*, this study also comprehensively illustrates some important plant gene families for the first time in this model plant like the expansin superfamily, the EF-Hand protein family, Ole e 1 protein family and profilins.

Expansin refers to a family of non-enzymatic proteins found in the plant cell walls with important roles in plant cell growth, developmental processes, and providing resistance to stress [40, 41]. Grass group 1 pollen allergens were recognised as a subgroup of the expansins referred to as β-expansins. Group 1 antigens are classified as major allergens since they cause IgE-mediated sensitization in up to 95% of patients with pollen allergy and exist in multiple isoallergenic forms or isoforms distinguished by their respective isoelectric points and amino acid sequences [1]. The isoform complexity of group 1 allergens and their expression patterns [42] suggest that attempts to silence their expression in a similar manner as have been attempted with the group 5 allergen genes [43] will be challenging. The present survey identified four proteins in *B. distachyon* genome belonging to the class of group 1 allergens. Regarding the expression profile, all the four corresponding transcripts showed significantly high expression levels in anther (Fig 4). The expression pattern data will be of particular interest for

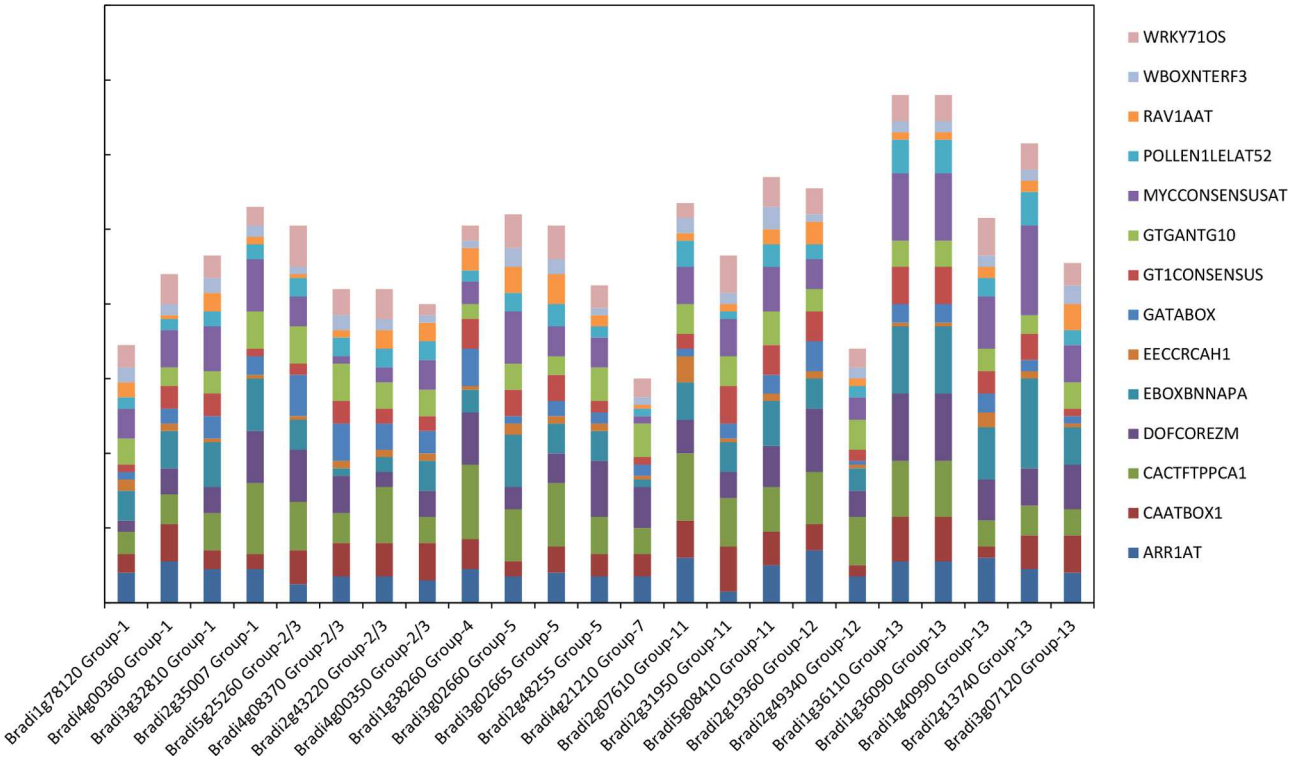


Fig 7. Duplication frequency of 14 most abundant *cis*-regulatory elements (CREs) in putative allergen-encoding genes in *B. distachyon*.

doi:10.1371/journal.pone.0169686.g007

further experiments in selecting the best candidate for gene silencing approaches and functional validation studies.

Group 5 allergens have been reported to trigger IgE-mediated respiratory allergy symptoms of up to 85% pollen-allergic individuals [1]. Group 5 allergens have also been known to exist in multiple isoforms, and so far have been identified only among the subfamily Pooideae. Group 5 allergen was first discovered as an IgE-binding protein in rye grass pollen cDNA expression library. The cDNA sequence revealed flanking transit peptide sequences typical of plastid-targeted proteins and immunolocalization studies confirmed localisation of allergen protein in amyloplasts of grass pollen [44]. Further immunological cross-reactivity studies showed this allergen group to be restricted to grasses of pooid sub-family establishing why pooid grasses such as ryegrass, timothy grass, Kentucky bluegrass, etc. are clinically most significant grasses. This is in contrast to Panicoid and Chlorodoid grasses with relatively lower allergic potential. Our present study suggests that gene encoding group 5 first appeared during the origin of pooid sub-family. Genomic surveys of the sequenced plant genomes have indeed failed to detect any gene with homology to group 5 protein family. This is in contrast to all other grass pollen allergens whose family members could be observed even in unrelated non-allergenic plant genera. The mechanism of *de novo* origin of group 5 allergen gene in Brachypodium or a related sister species remains a mystery. Earlier it has been reported that in the case of *Drosophila*, *de novo* genes had the most pronounced expression, and many of the *de novo* genes are related to male reproductive processes [45, 46]. Intriguingly grass group 5 allergen genes are also expressed at high levels in male gametophyte-specific manner. It has been proposed that evolutionary split between rice containing clade and Brachypodium containing clade happened around 63 mya [47]. Since rice genome and rice pollen transcriptome does not reveal

the presence of group 5 allergen gene but wheat genome shows the presence of group 5 protein, it appears that *de novo* origin of this gene occurred around 50 mya in a common grass ancestor of Brachypodium and Triticiae clades.

Brachypodium pollen allergen-encoding genes were found to be distributed among all five chromosomes following no specific pattern. Some of the genes belonging to the same group of pollen allergen were located on one chromosome suggesting an origin as a result of tandem gene duplication events. For example, homologs of group 12 were solely located on the chromosome 2 (Fig 5). Interestingly, some of the genes exhibited similar gene structures but have different chromosomal locations. For example, both Bradi1g78120.1 and Bradi4g00360.1 (group 1 homologs) had similar exon/intron structures, but they were located on chromosomes 1 and four respectively. This specific distribution suggested that this redistribution may be due to whole grass genome duplication events dating back 96 mya. Similar distribution pattern was also observed in Bradi4g08370.2 and Bradi5g25260.1 (group 2/3 homologs), indicating the unique distribution of pollen allergen genes on Brachypodium chromosomes.

It is a widely accepted that the intron/exon structure of genes is indicative of their evolutionary relationships [48]. Additionally, exon/intron gain/loss is a substantial indication for structural divergence and functional differentiation [49]. All of the transcribed pollen proteins in *B. distachyon* were either encoded by loci lacking introns or containing a reduced number of introns but exhibited a high level of expression in anthers. Since introns in some cases may be recruited in post-transcriptional control [50], elimination of introns could potentially release translational regulation, leading to further enhanced expression.

Three-dimensional protein structure models were constructed for 24 putative pollen-allergen proteins using Phyre2 prediction tool (Fig 6). Phye2 uses the alignment of hidden Markov models via HMM-HMM search [51] to significantly improve the accuracy of alignment and detection rate. The intensive mode of Phyre2 uses the multi-template modelling for higher accuracy. Furthermore, it integrates a new *ab initio* folding simulation termed as Poing [52] to model regions of proteins with no noticeable homology to known structures. The protein structure of all the 24 putative pollen allergens was modelled at >90% confidence and the percentage residue varied from 81 to 100. As observed in our study, tertiary structures of group 1 homologs in Brachypodium were shown to form dimers as these proteins consist of two domains. Similar observation was reported for a timothy grass group 1 allergen (Phl p 1) [53]. Likewise, in a previous study, the three-dimensional structure of Phl p 2 was reported to contain an all-β fold with nine anti parallel β-strands that formed a β sandwich [54]. Homology modelling studies for group 5 proteins in Brachypodium showed that these proteins are composed entirely of α-helices. Similar results have been reported earlier for Phl p 5 and Phl p 6 allergens [53, 55]. Hence, all the predicted protein structures were considered highly reliable, and this offers a preliminary basis for understanding the structures of these Brachypodium proteins.

Different *cis*-regulatory elements (CREs) and transcription factors that bind to these CREs have been reported to lie within anther-/pollen- specific promoters and were responsible for anther-/pollen- specific gene expression [56]. Promoter sequences such as GTGANTG10 (5′-GTGA-3′) and POLLEN1LELAT52 (5′- AGAAA-3′) were earlier reported as regulatory elements responsible for activation of anther-/pollen- specific genes [56]. GTGANTG10 is a GTGA motif found in the promoter of tobacco late pollen gene *g10*. The tobacco gene *g10* is preferentially and highly expressed in mature pollen, shows homology to pectate lyases, and is the putative homologue of the tomato gene *lat56* [57]. POLLEN1LELAT52 is a regulatory element responsible for pollen-specific activation of the tomato*lat52* gene. This CRE has been found in the promoter of tomato endo-beta-mannanase gene during late stages of anther development [58]. We have identified identical CREs in the promoter sequences of all the

identified Brachypodium allergen homologs. In our study, interestingly, QELEMENTZMZ M13, an AGGTCA motif was exclusively found in promoter sequences of putative group 1 allergens. AGGTCA has been reported to be the Quantitative element (Q) responsible for expression enhancing in *Zm13*, a pollen specific gene in maize [59]. Thus, our investigation revealed that the presence of these particular sequences (CREs) in the promoter regions of putative allergen-encoding genes in Brachypodium is responsible for driving high expression in anthers and pollen.

In summary, a total of 24 putative pollen allergen-encoding genes were found to be present in *B.distachyon* genome. Comprehensive phylogenetic analysis and sequence alignments with other known pollen allergens classified them into eight independent groups/families of pollen allergens. The variation in the lengths and structures of gene sequences indicated the level of complexity that has evolved within the allergen families. We also described the structure of 24 putative allergen proteins which could expedite the investigation of its molecular functions and facilitate further investigations in to understanding molecular basis of allergenicity. Hence, the present study is the first report on genome-wide identification, characterization and expression profiling of putative pollen allergens in *B. distachyon*. It is expected that the data obtained from this study would contribute to a better understanding of the complexity of pollen allergens in related grasses and stimulate further bio-prospecting of grass genomes to search for hypoallergenic forms of allergen proteins.

Supporting Information

S1 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of Expansin family in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed. The tree was divided into four clusters. The members were distinctly coloured to represent respective groups.
(TIF)

S2 Fig. Multiple sequence alignment of group 1 homologs in *B. distachyon* with group 1 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0 and conserved residues were highlighted in different colors.
(DOC)

S3 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of Group 2/3 homologs in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed.
(TIF)

S4 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of Group 2/3 homologs in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed.
(DOC)

S5 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of FAD binding and BBE family in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were

collapsed. The members were distinctly coloured to represent respective groups.
(TIF)

S6 Fig. Multiple sequence alignment of group 4 homologue in *B. distachyon* with group 4 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0 and conserved residues were highlighted in different colors.
(DOC)

S7 Fig. Phylogenetic relationships and protein domain architecture of group 5 proteins in *B. distachyon* and other grasses. Gene structures and expression profiles of group 5 homologs in Brachypodium. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed.
(TIF)

S8 Fig. Multiple sequence alignment of group 5 homologs in B. *distachyon* with group 5 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0 and conserved residues were highlighted in different colors.
(DOC)

S9 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of EF-Hand family in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed.
(TIF)

S10 Fig. Multiple sequence alignment of group 7 homologue in *B. distachyon* with group 7 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0 and conserved residues were highlighted in different colors.
(DOC)

S11 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of Ole e one family in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed. Group 11 homologs were represented in blue shade.
(TIF)

S12 Fig. Multiple sequence alignment of group 11 homologs in B. *distachyon* with group 11 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0 and conserved residues were highlighted in different colors.
(DOC)

S13 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of Profilin family in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed. Group 12 homologs were represented in pink shade.
(TIF)

S14 Fig. Multiple sequence alignment of group 12 homologs in *B. distachyon* with group 12 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0

and conserved residues were highlighted in different colors.
(DOC)

S15 Fig. Phylogenetic relationships, gene structures, expression profiles and protein domain architecture of Polygalacturonases family in *B. distachyon*. The protein sequences were aligned by Clustal X2.0 and unrooted phylogenetic tree was constructed by neighbour-joining method with 100 bootstrap replicates. Branches with less than 50% bootstrap support were collapsed. Group 13 homologs were represented in pink shade.
(TIF)

S16 Fig. Multiple sequence alignment of group 13 homologs in *B. distachyon* with group 13 allergen proteins from other grasses. The protein sequences were aligned by Clustal X2.0 and conserved residues were highlighted in different colors.
(DOC)

S1 Table. Accession numbers and protein sequences of pollen allergens.
(XLSX)

S2 Table. Percentage of protein sequence coverage and identity of identified Brachypodium proteins with known pollen allergens from other grasses.
(XLSX)

S3 Table. Gene expression matrix for *Brachypodium distachyon* (mapped against v2.1).
(XLSX)

S4 Table. Chromosomal location co-ordinates of putative allergen-encoding genes in *B. distachyon*.
(XLSX)

S5 Table. Functional motif numbers of identified *cis*-regulatory elements in putative allergen-encoding genes of *Brachypodium distachyon*.
(XLSX)

Acknowledgments

We thank contributors of the *B. distachyon* genome database that was used for searching putative allergen-encoding genes. AS also thanks, Dr. Lim Chee Liew for helping in preparing the figures during manuscript preparation. Financial support from the Australian Research Council (ARC DPO988972) is also gratefully acknowledged.

Author Contributions

Conceptualization: MS AS PB.

Formal analysis: NS.

Funding acquisition: MS PB.

Investigation: MS AS.

Methodology: AS NS.

Project administration: MS PB.

Resources: MS PB.

Software: NS.

Supervision: MS PB.

Validation: AS.

Visualization: AS NS.

Writing – original draft: AS MS.

Writing – review & editing: AS MS PB.

References

1. Andersson K, Lidholm J. Characteristics and Immunobiology of grass pollen allergens. Int Arc Aller A Imm. 2003; 130: 87–107.
2. Davies J M. Grass pollen allergens globally: the contribution of subtropical grasses to burden of allergic respiratory diseases. Clin Exp Allergy. 2014; 44: 790–801. doi: 10.1111/cea.12317 PMID: 24684550
3. Jaeger S. Exposure to grass pollen in Europe. Clin Exp Allergy. 2008; 8: 2–6.
4. Mohapatra SS, Lockey RF, Shirley S. Immunobiology of grass pollen allergens. Curr Allergy Asthma Rep. 2005; 5: 381–387. PMID: 16091211
5. Campbell BC, Gilding EK, Timbrell V, Guru P, Loo D, Zennaro D, et al. Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. J Allergy Clin Immun. 2014; 135: 133–142. doi: 10.1016/j.jaci.2014.06.034 PMID: 25129679
6. Weber RW. Cross-reactivity of pollen allergens. Curr Allergy Asthma Rep. 2004; 4: 401–40. PMID: 15283881
7. Ferreira F, Hawranek T, Gruber P, Wopfner N, Mari A.Allergic cross-reactivity: from gene to the clinic. Allergy. 2004; 59: 243–267. doi: 10.1046/j.1398-9995.2003.00407.x PMID: 14982506
8. Cosgrove DJ, Bedinger P, Durachko DM. Group 1 allergens of grass pollen as cell wall-loosening agents. Proc Natl Acad Sci USA. 1997; 94: 6559–6564. PMID: 9177257
9. Ansari AA, Shenbagamurthi P, Marsh DG. Complete primary structure of a *Lolium perenne* (perennial rye grass) pollen allergen, Lol p III: comparison with known Lol p I and II sequences. Biochem. 1989; 28: 8665–8670.
10. Matthiesen F, Lowenstein H. Group V allergens in grass pollen. I. Purification and characterization of group V allergen from *Phleum pratense* pollen, Phl p V. Cin Exp Allergy. 1991; 21: 297–307.
11. Ong EK, Griffith IJ, Knox RB, Singh MB. Cloning of a cDNA encoding a group V (group IX) allergen isoform from ryegrass pollen that demonstrates specific antigenic immunoreactivity. Gene. 1993; 134: 235–240. PMID: 8262382
12. Suphioglu C, Mawdsley D, Schäppi G, Gruehn S, de Leon M, Rolland JM, et al. Molecular cloning, expression and immunological characterization of Lol p 5C, a novel allergen isoform of rye grass pollen demonstrating high IgE reactivity. FEBS Lett. 1999; 462: 435–441. PMID: 10622741
13. Petersen A, Bufe A, Schlaak M, Becker WM.Characterization of the allergen group VI in timothy grass pollen (Phl p 6) I. Immunological and biochemical studies. Int Arch Allergy Imm. 1995; 108: 49–54.
14. Petersen A, Bufe A, Schramm G, Schlaak M, Becker WM.Characterization of the allergen group VI in timothy grass pollen (Phl p 6) II. cDNA cloning of Phl p 6 and structural comparison to group V. Int Arch Allergy Imm. 1995; 108: 55–59.
15. Vrtala S, Fischer S, Grote M, Vangelista L, Pastore A, Sperr WR, et al. Molecular, immunological, and structural characterization of Phl p 6, a major allergen and P-particle-associated protein from Timothy grass (Phleum pratense) pollen. J Immunol. 1999; 163: 5489–5496. PMID: 10553075
16. Wopfner N, Dissertori O, Ferreira F, Lackner P. Calcium-binding proteins and their role in allergic disease. Immunol Allergy Clin North Am. 2007; 27: 29–44. doi: 10.1016/j.iac.2006.10.003 PMID: 17276877
17. Sohn RH, Goldschmidt-Clermont PJ. Profilin: At the crossroads of signal transduction and the actin cytoskeleton. Bioessays. 1994; 16: 465–472. doi: 10.1002/bies.950160705 PMID: 7945274
18. Sh Kao, Su SN, Huang SW, Tsai JJ, Chow LP. Sub-proteome analysis of novel IgE-binding proteins from Bermuda grass pollen. Proteomics. 2005; 5: 3805–3813. doi: 10.1002/pmic.200401229 PMID: 16121337
19. Chow LP, Chiu LL, Khoo KH, Peng HJ, Yang SY, Huang SW, et al. Purification and structural analysis of the novel glycoprotein allergen Cyn d 24, a pathogenesis-related protein PR-1, from Bermuda grass pollen. FEBS J. 2005; 272: 6218–6227. doi: 10.1111/j.1742-4658.2005.05000.x PMID: 16336260

20. Draper J, Mur LA, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, et al. *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol. 2001.; 12: 1539–1555.
21. Bevan MW, Garvin DF, Vogel JP. Brachypodium distachyon genomics for sustainable food and fuel production. Curr Opin Biotech. 2010; 21: 211–217. doi: 10.1016/j.copbio.2010.03.006 PMID: 20362425
22. Mur LAJ, Allainguillaume J, Catalán P, Hasterok R, Jenkins G, Lesniewska K, et al. Exploiting the Brachypodium tool box in cereal and grass research. New Phytol. 2011; 191: 334–347. doi: 10.1111/j.1469-8137.2011.03748.x PMID: 21623796
23. Mochida K, Shinozaki K. Unlocking Triticeae genomics to sustainably feed the future. Plant Cell Physiol. 2013; 54: 1931–1950. doi: 10.1093/pcp/pct163 PMID: 24204022
24. Girin T, David LC, Chardin C, Sibout R, Krapp A, Ferrario-Méry S, et al. *Brachypodium*: a promising hub between model species and cereals. J Exp Bot. 2014. 65: 5683–5696. doi: 10.1093/jxb/eru376 PMID: 25262566
25. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. DOG 1.0: Illustrator of protein domain structures. Cell Res. 2009; 19: 271–273. doi: 10.1038/cr.2009.6 PMID: 19153597
26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013; 30: 2725–2729. doi: 10.1093/molbev/mst197 PMID: 24132122
27. Guo AY, Zhu QH, Chen X, Luo JC. GSDS: a gene structure display server. Yi Chuan. 2007; 29: 1023–1026. PMID: 17681935
28. Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 1999; 27: 297–300. PMID: 9847208
29. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. Nat Prot. 2009; 4: 363–371.
30. Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SH, et al. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. Plant J. 2012.; 71: 492–502. doi: 10.1111/j.1365-313X.2012.05005.x PMID: 22443345
31. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinfo. 2011.
32. Madritsch C, Gadermaier E, Roder UW, Lupinek C, Valenta R, Flicker S. High-density IgE recognition of the major grass pollen allergen Phl p 1 revealed with single-chain IgE antibody fragments obtained by combinatorial cloning. J Immunol. 2015.; 194: 2069–2078. doi: 10.4049/jimmunol.1402185 PMID: 25637023
33. Padavattan S, Flicker S, Schirmer T, Madritsch C, Randow S, Reese G, et al. High-affinity IgE recognition of a conformational epitope of the major respiratory allergen Phl p 2 as revealed by x-ray crystallography. J Immunol. 2009; 182: 2141–2151. doi: 10.4049/jimmunol.0803018 PMID: 19201867
34. Suphioglu C, Blaher B, Rolland JM, McCluskey J, Schäppi G, Kenrick J, et al. Molecular basis of IgE-recognition of Lol p 5, a major allergen of rye-grass pollen. Mol Immunol. 1998; 35: 293–305. PMID: 9747889
35. Jiang SH, Jasmin PX, Ting YY, Ramachandran S. Genome-wide identification and molecular characterization of Ole_e_I, Allerg_1 and Allerg_2 domain-containing pollen-allergen-like genes in *Oryza sativa*. DNA Res. 2005; 12: 167–179. doi: 10.1093/dnares/dsi005 PMID: 16303748
36. Petersen A, Dresselhaus T, Grobe K, Becker WM. Proteome analysis of maize pollen for allergy-relevant components. Proteomics. 2006; 6: 6317–6325. doi: 10.1002/pmic.200600173 PMID: 17080481
37. Russel SD, Bhalla PL, Singh MB. Transcriptome-based examination of putative pollen allergens of rice (*Oryza sativa* ssp. Japonica). Mol Plant. 2008; 1: 751–759. doi: 10.1093/mp/ssn036 PMID: 19825578
38. De Canio M, D'Aguanno S, Sacchetti C, Petrucci F, Cavagni G, Nuccetelli M, et al. Novel IgE recognised components of *Lolium perenne* pollen extract: comparative proteomics evaluation of allergic patients sensitization profiles. J Proteome Res. 2009; 8: 4383–4391. doi: 10.1021/pr900315a PMID: 19585971
39. Fonseca C., Planchon S, Renaut J, Oliveira MM, Batista R. Characterization of maize allergens-MON810 vs. Its non-transgenic counterpart. J Proteomics. 2012; 75: 2027–2037. doi: 10.1016/j.jprot.2012.01.005 PMID: 22270010
40. Cosgrove DJ. Growth of the plant cell wall. Nat. Rev. Mol. Cell Biol. 2005; 6: 850–861. doi: 10.1038/nrm1746 PMID: 16261190
41. Tenhaken R. Cell wall remodeling under abiotic stress. Front Plant Sci. 2015.
42. Valdivia ER, Sampedro J, Lamb JC, Chopra S, Cosgrove DJ. Recent proliferation and translocation of pollen group 1 allergen genes in the maize genome. Plant Physiol. 2007; 143: 1269–1281. doi: 10.1104/pp.106.092544 PMID: 17220362

43. Bhalla PL, Swoboda I, Singh MB. Antisense-mediated silencing of a gene encoding a major ryegrass pollen allergen. Proc Natl Acad Sci USA. 1999; 96: 11676–11680. PMID: 10500236

44. Singh MB, Hough T, Theerakulpisut P, Avjioglu A, Davies S, Smith PM, et al. Isolation of cDNA encoding a newly identified major allergenic protein of rye-grass pollen: intracellular targeting to the amyloplast. Pro Natl Acad Sci USA. 1991; 88: 1384–1388.

45. Begun DJ, Lindfors HA, Kern AD, Jones CD. Evidence for *de novo* evolution of testis-expressed genes in the *Drosophila yakuba*/ *Drosophila erecta* clade. Genetics. 2007; 176: 1131–1137. doi: 10.1534/genetics.106.069245 PMID: 17435230

46. Zhao L, Saelao P, Jones CD, Begun DJ. Origin and spread of *de novo* genes in *Drosophila melanogaster* populations. Science. 2014.; 343: 769–772. doi: 10.1126/science.1248286 PMID: 24457212

47. Wang X, Wang J, Jin D, Guo H, Lee TH, Liu T, Paterson AH. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. Molecular plant. 2015; 8: 885–98. doi: 10.1016/j.molp.2015.04.004 PMID: 25896453

48. Hu LF, Liu SQ. Genome-wide identification and phylogenetic analysis of ERF gene family in cucumbers. Genet Mol Biol. 2011; 34: 624–633. doi: 10.1590/S1415-47572011005000054 PMID: 22215967

49. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. Pro Natl Acad Sci USA. 2012; 109: 1187–1192.

50. Piriyapongsa J, Jordan IK. Dual coding of SiRNAs and miRNAs by plant transposable elements. RNA. 2008; 14: 814–821. doi: 10.1261/rna.916708 PMID: 18367716

51. Sding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005; 21: 951–960. doi: 10.1093/bioinformatics/bti125 PMID: 15531603

52. Jefferys BR, Kelley LA, Sternberg MJE. Protein folding requires crowd control in a simulated cell. J Mol Biol. 2010; 397: 1329–1338. doi: 10.1016/j.jmb.2010.01.074 PMID: 20149797

53. Antonia FD, Pavkov-Keller T, Zangger K, Keller W. Structure of allergens and structure based epitope predictions. Methods. 2014; 66: 3–21. doi: 10.1016/j.ymeth.2013.07.024 PMID: 23891546

54. Marino S De, Morelli MC, Fraternali F, Tamborini E, Musco G, Vrtala S, et al. An immunoglobulin-like fold in a major plant allergen: the solution structure of Phl p 2 from timothy grass pollen. Structure. 1999; 7: 943–952. PMID: 10467147

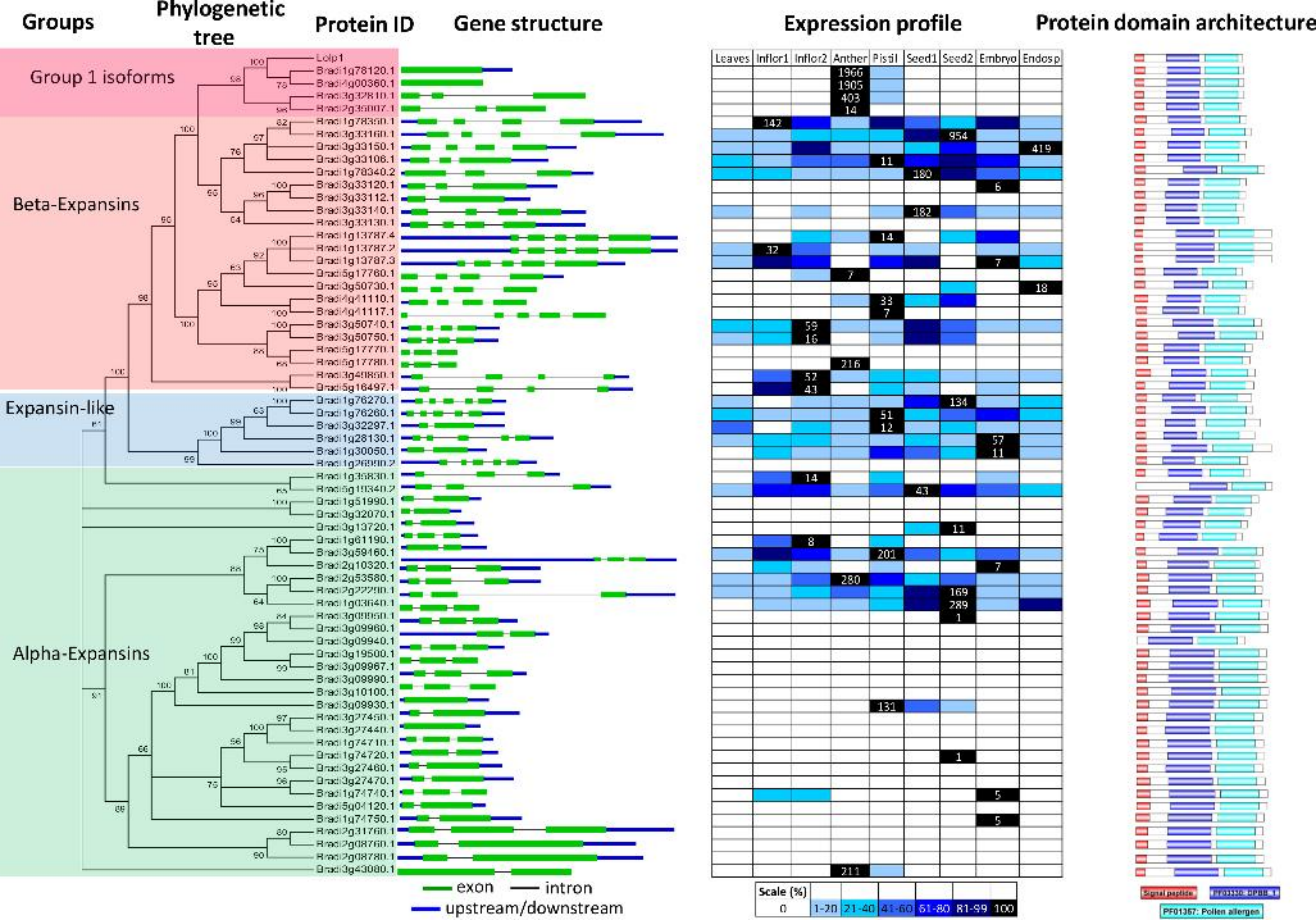
55. Rajashankar K, Bufe A, Weber W, Eschenburg S, Lindner B, Betzel C. Structure of the functional domain of the major grass-pollen allergen Phlp 5b. Acta Crystallogr D Biol Crystallogr. 2002; 58: 1175–1181. PMID: 12077438

56. Khurana R, Kapoor S, Tyagi AK. Anthology of anther/pollen-specific promoters and transcription factors. Crit Rev Plant Sci. 2012; 31: 359–390.

57. Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, et al. Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. Plant Mol Biol. 2001; 45: 577–585. PMID: 11414616

58. Filichkin SA, Leonard JM, Monteros A, Liu PP, Nonogaki H. A novel endo-beta-mannanase gene in tomato LeMAN5 is associated with anther and pollen development.Plant Physiol. 2004; 134: 1080–1087. doi: 10.1104/pp.103.035998 PMID: 14976239

59. Hamilton DA, Schwarz YH, Mascarenhas JP. A monocot pollen-specific promoter contains separable pollen-specific and quantitative elements. Plant Mol Biol. 1998; 38: 663–669. PMID: 9747811



Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Group 1 isoforms
Beta-Expansins
Expansin-like
Alpha-Expansins
Lolp1
Leaves
Inflor1
Inflor2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
exon
intron
upstream/downstream
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100
Signal peptide
PF01357: Pollen allergen

```
Lolp1(P14946)        IAKVPPGPNITAEYGDKWLDAKSTWYG-KPTGAGPKDNGGACGYKNVDKAPFNGMTGCG
Phaa1(Q41260)        IAKVPPGPNITAEYGDKWLDAKSTWYG-KPTGAGPKDNGGACGYKDVDKAPFNGMTGCG
Dacg1(CAD20406)      IPKVPPGPNITATYGDKWLDAKSTWYG-KPTGAGPKDNGGACGYKDVDKAPFNGMTGCG
Holl1(P43216)        IAKVPPGPNITATYGDEWLDAKSTWYG-KPTGAGPKDNGGACGYKDVDKPPFSGMTGCG
Phlp1(P43213)        IPKVPPGPNITATYGDKWLDAKSTWYG-KPTGAGPKDNGGACGYKDVDKPPFSGMTGCG
Poap1(CAA10520)      IAKVPPGPNITATYGDKWLDAKSTWYG-KPTGAGPKDNGGACGYKDVDKAPFSGMTGCG
Bradi1g78120.1       IPKVPPGPNITATYGDKWLDAKSTWYG-KPTGAGPKDNGGACGYKDVDKEPFSGMTSCG
Bradi4g00360.1       IPKVPPGPNITATYNGKWLDAKSTWYG-RPEGAGPKDNGGACGYKDVDKPPFNGMTSCG
Orys1(AAA86533)     GHPKVPPGPNITTSYGDKWLEARPPGMV-RPRVLAPKDNGGACGYKDVDKAPFLGMNSCG
Cynd1(O04701)        AIGDKPGPNITATYGSKWLEARATFYGSNPRGAAPDDHGGACGYKDVDKPPFDGMTACG
Zeam1(P58738)       GPPKVPPGPNITTNYNGKWLTARATWYG-QPNGAGAPDNGGACGIKNVNLPPYSGMTACG
Bradi3g32810.1       IPKVPPGPNVTANYNGQWLNAMATWYG-KPTGAGPKDNGGACGIKDVNLPPYSAMTACG
Bradi2g35007.1       FVVVPPGPNITTSYDGHWLHAKATWYG-SPTGAGPPDNGGSCGIKETNLPPYNSLTSCG

Lolp1(P14946)       NTPIFKDGRGCGSCFEIKCTKP-ESCSGEAVTVTITDDNEEPIAPYHFDLSGHAFGSMAK
Phaa1(Q41260)       NTPIFKDGRGCGSCFELKCSKP-ESCSGEPITVHITDDNEEPIAPYHFDLSGHAFGSMAK
Dacg1(CAD20406)     NTPIFKDGRGCGSCFEIKCTKP-ESCSGEAVTVHITDDNEEPIAPYHFDLSGHAFGSMAK
Holl1(P43216)       NTPIFKDGRGCGSCFEIKCTKP-ESCSGEPVTVHITDDNEEPIAPYHFDLSGHAFGSMAK
Phlp1(P43213)       NTPIFKSGRGCGSCFEIKCTKP-EACSGEPVVVHITDDNEEPIAPYHFDLSGHAFGAMAK
Poap1(CAA10520)     NTPIFKSGRGCGSCFEIKCTKP-ESCSGEPVLVHITDDNEEPIAAYHFDLSGKAFGAMAK
Bradi1g78120.1      NTPIFRDGRGCGSCFEIKCTKP-DACSGDPVLVHITDDNEEPIAAYHFDLSGHAFGSMAK
Bradi4g00360.1      NTPIFRDGRGCGSCFEIKCDKPAEFCSGQPVLVHITDDNEEPIAAYHFDLSGKAFGSMAK
Orys1(AAA86533)     NDPIFKDGKGCGSCFEIKCSKP-EACSDKPALIHVTDMNDEPIAAYHFDLSG---LAMAK
Cynd1(O04701)       NEPIFKDGLGCRACYEIKCKEP-VECSGEPVLVKITDKNYEHIAAYHFDLSGKAFGAMAK
Zeam1(P58738)       NVPIFKDGKGCGSCYEVRCKE-KPECSGNPVTVYITDMNYEPIAPYHFDLSGKAFGSLAK
Bradi3g32810.1      NVPIFKDGRGCGSCYEVKCNVPAKLCSNKPITVFITDMNYEPIAPYHLDLSGTAFGLMAQ
Bradi2g35007.1      NLPIFKDGKGCGSCFKIKCGN--KVCSKKPVVVFITDMNYYQFAPYHFDLSGTAFGALAM

Lolp1(P14946)       KGEEQNVRSAGELELQFRRVKCKYPDDTKPTFHVEKASNPNYLAILVKYVDGDGDVVAVD
Phaa1(Q41260)       KGEEENVRGAGELELQFRRVKCKYPDGTKPTFHVEKGSNPNYLALLVKYVDGDGDVVAVD
Dacg1(CAD2040)      KGEEQKLRSAGELELQFRRVKCKYPEGTKVTFHVEKGSNPNYLALLVKYVDGDGDVVAVD
Holl1(P43216)       KGEEQKLRSAGELELKFRRVKCKYPDGTKPTFHVEKGSNPNYLALLVKYIDGDGDVVAVD
Phlp1(P43213)       KGDEQKLRSAGELELQFRRVKCKYPEGTKVTFHVEKGSNPNYLALLVKYVNGDGDVVAVD
Poap1(CAA1052)      KGEEQKLRSAGELELKFRRVKCEYPEGTKVTFHVEKGSNPNYLALLVKYVTGDGDVVAVD
Bradi1g78120.1      KGKEQDLRSAGEVEIQFRRVKCKYPEGTKVTFHVEKGSSPNYLAILVKYVGGDGDVVAVD
Bradi4g00360.1      KGQEQKLRGCGEVEIQFRRVKCYYPLGTKVTYHVEKGSNPNYLALLVKFVGGDGDVVAVE
Orys1(AAA86533)     DGKDEELRKAGIIDTQFRRVKCKYPADTKITFHIEKASNPNYLALLVKYVAGDGDVVEVE
Cynd1(O04701)       KGQEDKLRKAGELTLQFRRVKCKYPSGTKITFHIEKGSNDHYLALLVKYAAGDGNIVAVD
Zeam1(P58738)       PGLNDKIRHCGIMDVEFRRVRCKYPAGQKIVFHIEKGCNPNYLAVLVKYVADDGDIVLME
Bradi3g32810.1      PGKEQLLRNCGELQLQFRRVRCKLPPGTKITFHIEKGSNPNYLAVLVKFASDDGDIVQMD
Bradi2g35007.1      PGRESELRRYGIIDLRFRRVRCKLAPKMKIAFHVEEGSNPEYLAVLVKFVAGDGDIVQMD

Lolp1(P14946)       IKEKGKDKWIELKESWGAVWRIDT--PDKLTGPFTVRYTTEGGTKSEFEDVIPEGWKADT
Phaa1(Q41260)       IKEKGKDKWIELKESWGAIWRIDT--PDKLTGPFTVRYTTEGGTKAEFEDVIPEGWKADT
Dacg1(CAD2040)      IKEKGKDKWIALKESWGAIWRVDT--PDKLTGPFTVRYTTEGGTKSEVEDVIPEGWKADT
Holl1(P43216)       IKEKGKDKWIELKESWGAVWRVDT--PDKLTGPFTVRYTTEGGTKGEAEDVIPEGWKADT
Phlp1(P43213)       IKEKGKDKWIELKESWGAIWRIDT--PDKLTGPFTVRYTTEGGTKTEAEDVIPEGWKADT
Poap1(CAA1052)      IKEKGKDKWIELKESWGSIWRVDT--PDKLTGPFTVRYTTEGGTKGEAEDVIPEGWKADT
Bradi1g78120.1      VKEKGKDEWVALKESWGAVWRLDT--AKPLKGPLTVRYTTDGGTKGESEDVIPEDWKPDT
Bradi4g00360.1      VQEKGKYNWIPLKESWGAVWRIDT--AKPLKGPLSVRYTTDGGTKAVSPDVIPEKWKPDT
Orys1(AAA86533)     IKEKGSEEWKALKESWGAIWRIDT--PKPLKGPFSVRVTTEGARRSSAEDAIPDPGRRQR
Cynd1(O04701)       IKPRDSDEFIPMKSSWGAIWRIDP--KKPLKGPFSIRLTSEGGAHLVQDDVIPANWKPDT
Zeam1(P58738)       IQDKLSAEWKPMKLSWGAIWRMDT--AKALKGPFSIRLTSESGKKVIAKDVIPANWRPDA
Bradi3g32810.1      LQDKISPEWKPMKESWGAVWRMDS--IKPLKGPYSIRLTSESGKKLIAKDIIPLNWKPDT
Bradi2g35007.1      LKQEGWPEWKPMRESWGAIWRMDVD--HPLQGPFSIRLTSESGKVLVAPHVIPADWKPKT

Lolp1(P14946)       SYSAK-----
Phaa1(Q41260)       HDASK-----
Dacg1(CAD2040)      SYEAK-----
Holl1(P43216)       AYEAK-----
Phlp1(P43213)       SYESK-----
Poap1(CAA1052)      AYASK-----
Bradi1g78120.1      MYESK-----
Bradi4g00360.1      MYVAKY----
Orys1(AAA86533)     VQVNVQAK--
Cynd1(O04701)       VYTSKLQFGA
Zeam1(P58738)       VYTSNVQFY-
Bradi3g32810.1      FLKSNIQF—
Bradi2g35007.1       --------
```

Yellow: hydroxyproline residues at positions 5 and 8

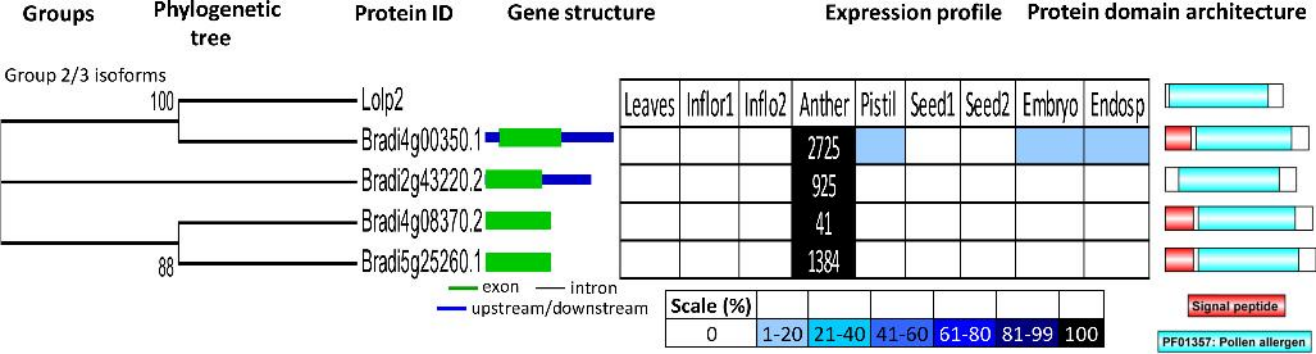
Pink: N-glycosylation site

Green: 7 conserved cysteine residues

Light Gray: Conserved residue

Dark Gray: Conserved substitutions

Red: IgE-binding residues


Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Group 2/3 isoforms
100
88
Lolp2
Bradi4g00350.1
Bradi2g43220.2
Bradi4g08370.2
Bradi5g25260.1
exon
intron
upstream/downstream
Leaves
Inflor1
Inflo2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
2725
925
41
1384
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100
Signal peptide
PF01357: Pollen allergen

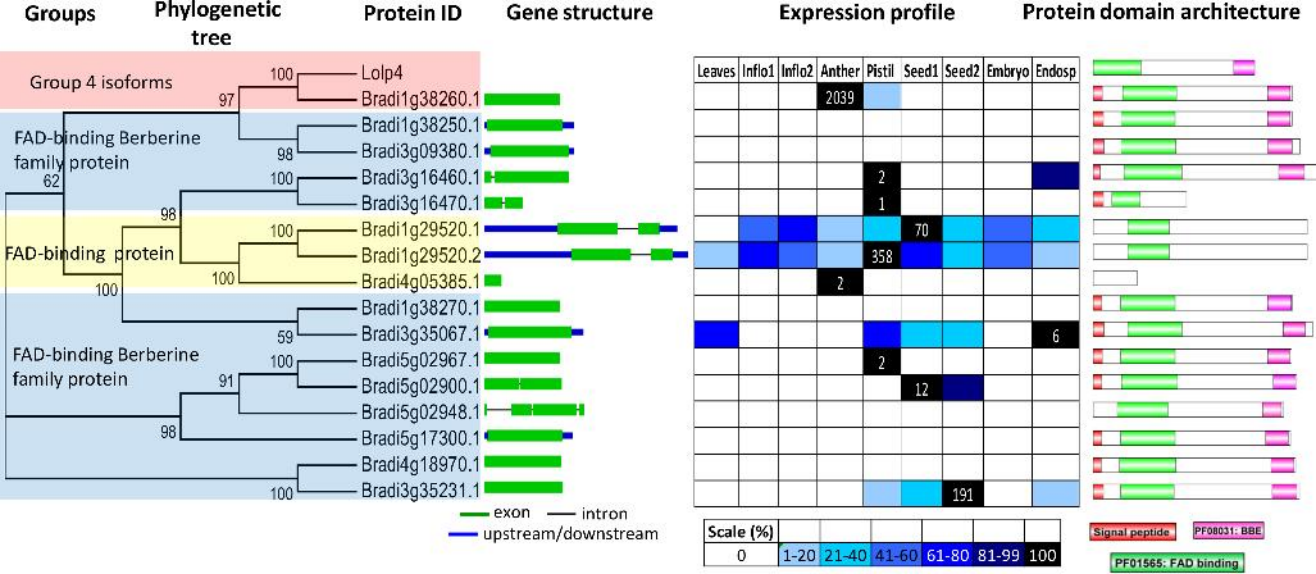
```
Poap2(CAA10348)      --------VPKVTFTVEKGSNEK--HLAVLVKYEGDTMAEVELREHGSDEWVAMTKGEGG
Holl2(CAA10347)      --------VPKVTFTVEKGSNEK--HLAVLVKYEGDTMAEVELREHGSDEWVAMTKGEGG
Phlp2(P43214)        --------VPKVTFTVEKGSNEK--HLAVLVKYEGDTMAEVELREHGSDEWVAMTKGEGG
Dacg2(CAA10345)      --------VPKVTFTVEKGSNEK--HLAVLVKYEGDTMAEVELREHGSDEWVAMTKGEGG
Tria2(CAA10349)      --------VPKVTFTVEKGSNEK--HLAVLVKYEGDTMAEVELREHGSDEWVAMTKGEGG
Cynd2(CAA10346)      --------VPKVTFTVEKGSNEK--HLAVLVKYEGDTMAEVELREHGSDEWVAMTKGEGG
Bradi4g00350.1       ---------VKPTFTVQPGSTTK--KLGVKVNKPGHSVAEVELRQHGSETWLTMKKTGPD
Bradi4g08370.2       --------APKVTFTVEKGSDPK---KIVKYDKEGDGMAEVELKQKGSNEWLAMSKCKDT
Bradi5g25260.1       --------APKVTFTVEKGSDPKKIVLQVKYDKEGDSMAEVELKQKGSNEWLAMSKCKDT
Bradi2g43220.2       MLFAGVWCAPKVTFTVEKGSDPKKIVLDVKYNKEGDSMAEVELKQKGSNEWLAMSKCKDT
Dacg3(P93124)        ---------VKVTFKVEKGSDPKKLVLDIKYTRPGDTLAEVELRQHGSEEWEPLTKKGN-
Lolp3(P14948)        ---------TKVDLTVEKGSDAKTLVLNIKYTRPGDTLAEVELRQHGSEEWEPMTKKGN-
Lolp2 (P14947)       --------AAPVEFTVEKGSDEKNLALSIKYNKEGDSMAEVELKEHGSNEWLALKKNGDG

Poap2(CAA10348)      -VWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE-
Holl2(CAA10347)      -VWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE-
Phlp2(P43214)        -VWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE-
Dacg2(CAA10345)      -VWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE-
Tria2(CAA10349)      -VWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE-
Cynd2(CAA10346)      -VWTFDSEEPLKGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE-
Bradi4g00350.1       -TFTVQSPTPLKGPYNFRIVTEKGLRGVFDDVVPETFKCGTTYVPDEY
Bradi4g08370.2       GAWKYESPEPPKCPLNIRFQSEKGMRNVFDDVIPENYKIGSSYAPQEY
Bradi5g25260.1       GAWKYESPEPLKCPLNIRFQSEKGMRNVFDDVIPENYKIGSTYSPQEY
Bradi2g43220.2       GVWKYESPEALKCPLNIRFQSEKGMRNVFDDVIPENYKVGSTYAPQEY
Dacg3(P93124)        -LWEVKSSKPLTGPFNFRFMSKGGMRNVFDEVIPTAFKIGTTYTPEE-
Lolp3(P14948)        -LWEVKSAKPLTGPMNFRFLSKGGMKNVFDEVIPTAFTVGKTYTPEYN
Lolp2 (P14947)       -VWEIKSDKPLKGPFNFRFVSEKGMRNVFDDVVPADFKVGTTYKPE--
```

Red: fully conserved epitope residues

Turquoise: conservatively substituted epitope residues

Light gray: conserved residues

Dark gray: conservative substitutions

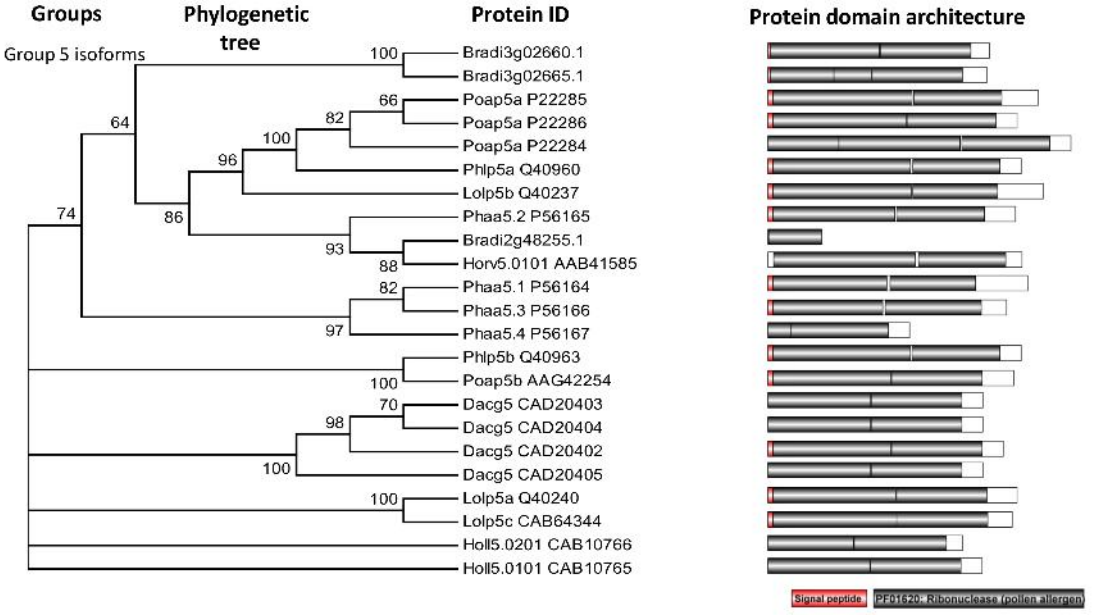

Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Group 4 isoforms
FAD-binding Berberine family protein
FAD-binding protein
FAD-binding Berberine family protein
Lolp4
Bradi1g38260.1
Bradi1g38250.1
Bradi3g09380.1
Bradi3g16460.1
Bradi3g16470.1
Bradi1g29520.1
Bradi1g29520.2
Bradi4g05385.1
Bradi1g38270.1
Bradi3g35067.1
Bradi5g02967.1
Bradi5g02900.1
Bradi5g02948.1
Bradi5g17300.1
Bradi4g18970.1
Bradi3g35231.1
Leaves
Inflo1
Inflo2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
exon
intron
upstream/downstream
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100
Signal peptide
PF08031: BBE
PF01565: FAD binding

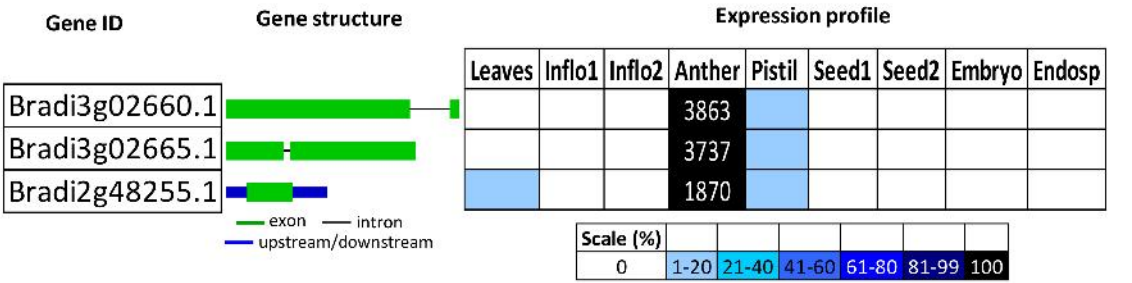
```
Lolp4(Q5T1W3)              ------------------------------------------------------------
Phlp4.0201(CAD54671)       SSCQVAFSYFPPPAAKEDFLGCLVKEIPPRLLYAKSSPAYPSVLGQTIRNSRWSSPDNVK
Phlp4.0101(CAD54670)       SSCEVALSYYPTPLAKEDFLRCLVKEIPPRLLYAKSSPAYPSVLGQTIRNSRWSSPDNVK
Tria4.0101(CAH92632)       --------TYAPVPAKEDFLGCLMKEIPARLLYAKSSPDFPTVLAQTIRNSRWLSPQNVK
Tria4.0201(CAH92633)       --------TYAPVPAKEDFLGCLMKEIPARLLYAKSSPDYPTVLAQTIRNSRWSTQQNVK
Secc4.0101(CAH92627)       -------AAYAPVPAKADFLGCLMKEIPARLLYAKSSPDYPTVLAQTIRNSRWSSPQNVK
Secc4.0201(CAH92630)       ----AVSYAAAPVPAKEDFFGCLVKEIPARLLYAKSSPAFPTVLAQTIRNSRWSSPQSVK
Bradi1g38260.1             --------YPPVLAPKEDFLACLVKEIPPRLLYAKSSPAYPTVLSATIRNSRWSSPQNVK

Lolp4(Q5T1W3)              ---------------AVVCGRRYDVRIRVRSGGHDYEGLSYRSL-QPENFAVVDLNQMRA
Phlp4.0201(CAD54671)       PLYIITPTNVSHIQSAVVCGRRHSVRIRVRSGGHDYEGLSYRSL-QPETFAVVDLNKMRA
Phlp4.0101(CAD54670)       PIYIVTPTNASHIQSAVVCGRRHGVRIRVRSGGHDYEGLSYRSL-QPEEFAVVDLSKMRA
Tria4.0101(CAH92632)       PLYIITPTNASHIQSAVVCGRRHSVRLRVRSGGHDYEGLSYRSE-KPETFAVVDLNKMRA
Tria4.0201(CAH92633)       PLYIITPTNASHIQSAVVCGRRHGVRLRVRSGGHDYEGLSYRSE-KPETFAVVDLNKMRA
Secc4.0101(CAH92627)       PIYIITPTNASHIQSAVVCGRRHGIRLRVRSGGHDYEGLSYRSE-KPETFAVVDLNKMRA
Secc4.0201(CAH92630)       PLYIITPTNASHIQSAVVCGRRHGVRIRVRSGGHDYEGLSYRSE-RPEAFAVVDLNKMRA
Bradi1g38260.1             PLYIVTPTNVSHIQSAVVCGRRHGVRIRVRSGGHDYEGLSYRSE-RAESFAVVDLNMMRA

Lolp4 (Q5T1W3)             VLVDGKARTAWVDSGAQLGELYYAISKYS-RTLAFPAGVCPTIGVGGNLAGGGFGMLLRK
Phlp4.0201(CAD54671)       VWVDGKARTAWVDSGAQLGELYYAIYKAS-PTLAFPAGVCPTIGVGGNFAGGGFGMLLRK
Phlp4.0101(CAD54670)       VWVDGKARTAWVDSGAQLGELYYAIHKAS-PVLAFPAGVCPTIGVGGNFAGGGFGMLLRK
Tria4.0101(CAH92632)       VLIDGYARTAWVESGAQLGELYYAIAKNS-PVLAFPAGVCPTIGVGGNFAGGGFGMLLRK
Tria4.0201(CAH92633)       VVVDGYARTAWVESGAQLGELYYAIAKNS-PVLAFPAGVCPSIGVGGNFAGGGFGMLLRK
Secc4.0101(CAH92627)       VSVDGYARTAWVESGAQLGELYYAIAKNS-PVLAFPAGVCPSIGVGGNFAGGGFGMLLRK
Secc4.0201(CAH92630)       VVVDGKARTAWVDSGAQLGELYYAIAKNS-PVLAFPAGVCPTIGVGGNFAGGGFGMLLRK
Bradi1g38260.1             VSVDANARTAWVESGAQIGELYYAISKAS-PSLAFPAGVCPSIGVGGHFSGGGFGMLLRK

Lolp4(Q5T1W3)              YGIAAENVIDVKLVDANGKLHDKKSMGDDHFWAVRGGGG-ESFGIVVSWQVKLLPVPPTV
Phlp4.0201(CAD54671)       YGIAAENVIDVKLVDANGKLHDKKSMGDDHFWAVRGGGG-ESFGIVVAWQVKLLPVPPTV
Phlp4.0101(CAD54670)       YGIAAENVIDVKLVDANGTLHDKKSMGDDHFWAVRGGGG-ESFGIVVAWKVRLLPVPPTV
Tria4.0101(CAH92632)       YGIAAENVIDVKVVDPNGKLLDKSSMSPDHFWAVRGGGG-ESFGIVVSWQVKLLPVPPTV
Tria4.0201(CAH92633)       YGIAAENVIDVKVVDPDGKLLDKSSMSADHFWAVRGGGG-ESFGIVVSWQVKLMPVPPTV
Secc4.0101(CAH92627)       YGIAAENVIDVKVVDPNGKLLDKSSMSADHFWAVRGGGG-ESFGIVVSWQVKLLPVPPTV
Secc4.0201(CAH92630)       YGIAAENVIDVKVVDANGTLLDKSSMSADHFWAVRGGGG-ESFGIVVSWQVKLLPVPPTV
Bradi1g38260.1             FGIAAENVLDAKLVDANGKLHDRKSMGEDHFWAIRGGGG-ESFGIVVGWEVKLLPVPPVV

Lolp4(Q5T1W3)              TIFKIPKSVSEGAVDIINKWQLVAPQLPADLMIRIIAMGPKATFEAMYLGTCKTLTPMMQ
Phlp4.0201(CAD54671)       TIFKISKTVSEGAVDIINKWQVVAPQLPADLMIRIIAQGPKATFEAMYLGTCKTLTPLMS
Phlp4.0101(CAD54670)       TVFKIPKKASEGAVDIINRWQVVAPQLPDDLMIRVIAQGPTATFEAMYLGTCQTLTPMMS
Tria4.0101(CAH92632)       TVFKIPKTVQEGAVDLVNKWQLVGPALPGDLMIRVIAAGNTATFEGMYLGTCQTLTPLMS
Tria4.0201(CAH92633)       TVFKIPKTVQEGAVDLVNKWQLVGPALPGDLMIRVIAAGNTATFEALYLGTCKTLTPLMS
Secc4.0101(CAH92627)       TVLKIPKTVQEGAIDLVNKWQLVGPALPGDLMIRIILAGNSATFEAMYLGTCSTLTPLMS
Secc4.0201(CAH92630)       TVFKIPKTVQEGAVELINKWQLVAPALPDDLMIRIIAFGGTAKFEAMYLGTCKALTPLMS
Bradi1g38260.1             TVFKVSKTLKDGAIDIVNKWQTVAPALPGDLMIRILAMAQQATFEGMYLGTCNNLLPLIT

Lolp4(Q5T1W3)              SKFPELGMNASHCNEMSWIESIPFVHLGHRDSLEGDLLNRNNTFKPFAEYKSDYVYEPFP
Phlp4.0201(CAD54671)       SKFPELGMNPSHCNEMSWIQSIPFVHLGHRDALEDDLLNRNNSFKPFAEYKSDYVYQPFP
Phlp4.0101(CAD54670)       SKFPELGMNASHCNEMSWIQSIPFVHLGHRDNIEDDLLNRNNTFKPFAEYKSDYVYEPFP
Tria4.0101(CAH92632)       SQFPELGMNPYHCNEMPWIKSIPFIHLG-KEASLVDLLNRNNTFKPFAEYKSDYVYQPFP
Tria4.0201(CAH92633)       SQFPELGMNPYHCNEMPWIKSVPFIHLG-KQAGLDDLLNRNNTFKPFAEYKSDYVYQPFP
Secc4.0101(CAH92627)       SKFPELGMNPSHCNEMSWIKSIPFIHLG-KQN-LDDLLNRNNTFKPFAEYKSDYVYQPFP
Secc4.0201(CAH92630)       SRFPELGMNASHCNEMPWIKSVPFIHLG-KQATLSDLLNRNNTFKPFAEYKSDYVYQPVP
Bradi1g38260.1             SKFPELGFNRGQCNEMPWAQTIPFIHLG--NRDLGDLTNRNNNFKPFAEYKSDYVYQPIP

Lolp4(Q5T1W3)              KSVWEQIFGTWLVKPGAGIMIFDPYGATISATPEAATPFPHRKGVLFNIQYVNYWFA--P
Phlp4.0201(CAD54671)       KTVWEQILNTWLVKPGAGIMIFDPYGATISATPESATPFPHRKGVLFNIQYVNYWFA--P
Phlp4.0101(CAD54670)       KEVWEQIFSTWLLKPGAGIMIFDPYGATISATPEWATPFPHRKGVLFNIQYVNYWFA--P
Tria4.0101(CAH92632)       KPVWEQIFG-WLTKPGGGMMIMDPYGATISATPEAATPFPHRQGVLFNIQYVNYWFA--E
Tria4.0201(CAH92633)       KPVWEQIFG-WLAKPGAGIMIMDPYGATISATPEAATPFPHRQGVLFNIQYVNYWFA--E
Secc4.0101(CAH92627)       KPVWEQIFG-WLVKPGAGIMIMDPYGATISATPEAATPFPHRQGVLFNIQYVNYWFA--E
Secc4.0201(CAH92630)       KPVWAQIFV-WLVKPGAGIMVMDPYGAAISATPEAATPFPHRKDVLFNIQYVNYWFD--E
Bradi1g38260.1             KNVWEQIFG-WLTKPGAGIMIMDPYGATISATPETATPFPHRKGVLFNIQYVNYWFA—-E

Lolp4(Q5T1W3)              GAGAAPLSWSKEIYNYMEPYVSKNPRQAYANYRDIDLGRNEVVNG----VSTYSSGKVWG
Phlp4.0201(CAD54671)       GAAAAPLSWSKDIYNYMEPYVSKNPRQAYANYRDIDLGRNEVVND----VSTYASGKVWG
Phlp4.0101(CAD54670)       GAGAAPLSWSKEIYNYMEPYVSKNPRQAYANYRDIDLGRNEVVND----VSTFSSGLVWG
Tria4.0101(CAH92632)       AAAAAPLQWSKDMYNFMEPYVSKNPRQAYANYRDIDLGRNEVVND----ISTYSSGKVWG
Tria4.0201(CAH92633)       PAGAAPLQWSKDIYNFMEPYVSKNPRQAYANYRDIDLGRNEVVND----ISTYSSGKVWG
Secc4.0101(CAH92627)       SAGAAPLQWSKDIYKFMEPYVSKNPRQAYANYRDIDLGRNEVVND----ISTYSSGKVWG
Secc4.0201(CAH92630)       AGGAAPLQWSKDMYRFMEPYVSKNPRQAYANYRDIDLGRNEVVND----ISTYASGKVWG
Bradi1g38260.1             GAGAAPLQWSKDMYKFMEPYVSKNPRQAYANYRDIDLGRNEVVND----VSTYSSGKVWG

Lolp4(Q5T1W3)              QKYFKGNFERLAITKGKVDPTDYFRNE--------------------------
Phlp4.0201(CAD54671)       QKYFKGNFERLAITKGKVDPTDYFRNEQSIPPLIKKY---------------
Phlp4.0101(CAD54670)       QKYFKGNFQRLAITKGKVDPTDYFRNEQSIPPLIQKY---------------
Tria4.0101(CAH92632)       EKYFKGNFQRLAITKGKVDPQDYFRNEQSIPPLLEKY---------------
Tria4.0201(CAH92633)       EKYFKSNFQRLAITKGKVDPQDYFRNEQSIPPLIEKY---------------
Secc4.0101(CAH92627)       EKYFKGNFQRLAITKGKVDPQDYFRNEQSIPPLVEKY---------------
Secc4.0201(CAH92630)       EKYFKGNFQRLAITKGKVDPQDYFRNEQSIPPLLGK----------------
Bradi1g38260.1             EKYFKGNFQRLAMIKGKVDPEDYFRNEQSIPPLLTKY---------------
```

Light gray: conserved residues

Dark gray: Conservative substitutions

Turquoise: N-glycosylation sites


Groups
Phylogenetic tree
Protein ID
Protein domain architecture
Group 5 isoforms
100
64
66
82
100
96
74
86
93
88
82
97
100
70
98
100
100
Bradi3g02660.1
Bradi3g02665.1
Poap5a P22285
Poap5a P22286
Poap5a P22284
Phlp5a Q40960
Lolp5b Q40237
Phaa5.2 P56165
Bradi2g48255.1
Horv5.0101 AAB41585
Phaa5.1 P56164
Phaa5.3 P56166
Phaa5.4 P56167
Phlp5b Q40963
Poap5b AAG42254
Dacg5 CAD20403
Dacg5 CAD20404
Dacg5 CAD20402
Dacg5 CAD20405
Lolp5a Q40240
Lolp5c CAB64344
Holl5.0201 CAB10766
Holl5.0101 CAB10765
Signal peptide
PF01620: Ribonuclease (pollen allergen)


Gene ID
Gene structure
Expression profile
Bradi3g02660.1
Bradi3g02665.1
Bradi2g48255.1
exon
intron
upstream/downstream
Leaves
Inflo1
Inflo2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
3863
3737
1870
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100

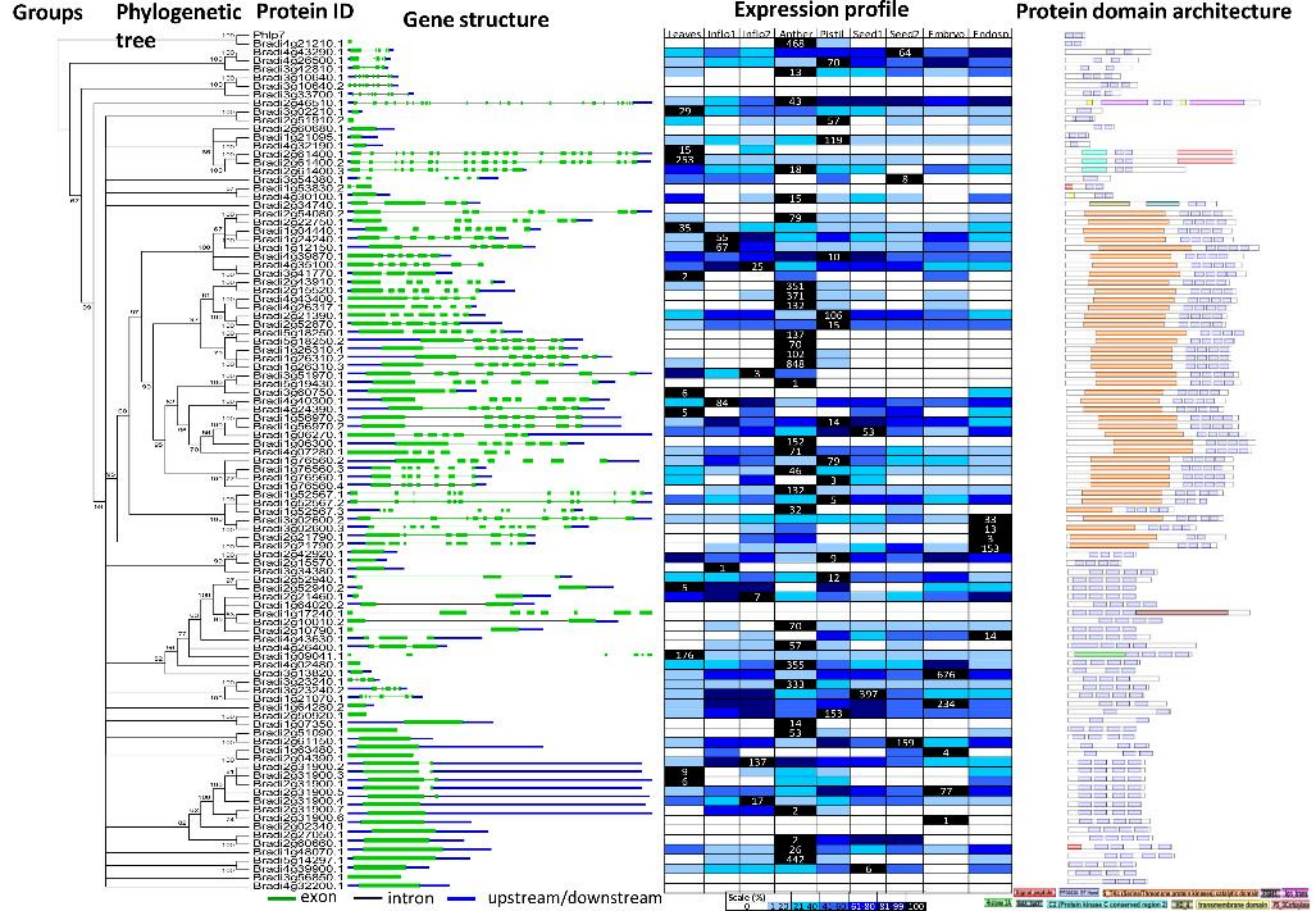
```
Phlp5 (Q40960)          ------------------------------------------------------------
Lolp5 (Q40237)          ------------------------------------------------------------
Lolp5 (Q40240)          ------------------------------------------------------------
Holl5.0201(CAB10766)    ------------------------------------------------------------
Holl5.0101(CAB10765)    ------------------------------------------------------------
Dacg5(CAD20402)         ------------------------------------------------------------
Dacg5(CAD20403)         ------------------------------------------------------------
Dacg5(CAD20404)         ------------------------------------------------------------
Dacg5(CAD20405)         ------------------------------------------------------------
Phlp5 (Q40963)          ------------------------------------------------------------
Poap5 (AAG42254)        ------------------------------------------------------------
Phaa5  (P56164)         ------------------------------------------------------------
Bradi3g02660.1          ------------------------------------------------------------
Bradi3g02665.1          ------------------------------------------------------------
Horv5.0101(AAB41585)    -------------------------------------------------------MANS
Bradi2g48255.1          ------------------------------------------------------------

Phlp5 (Q40960)          -------------------------------ADLGYGPATPAAPAAGYTPATPAAPAE--
Lolp5 (Q40237)          ----------------------------ADAGYAPATPATPAAPATAATPATPATPATPA
Lolp5 (Q40240)          ---------------------------------------ADAGYTPAAAATPATPAATP
Holl5.0201(CAB10766)    ------------------------------------------------------------
Holl5.0101(CAB10765)    ---------------------------------------ADAGYTPAAPAAAG------
Dacg5(CAD20402)         ---------------------------------------ADAGYTPAAAATPAT-----
Dacg5(CAD20403)         ---------------------------------------ADAGYTPAAAATPAT-----
Dacg5(CAD20404)         ---------------------------------------ADAGYTPAAAATPAT-----
Dacg5(CAD20405)         ---------------------------------------ADAGYTPAAAATPAA-----
Phlp5 (Q40963)          ---------------------------AAAAVPRRGPRGGPGRSYTADAGYAPATPAAAG
Poap5 (AAG42254)        ------------------------------------------------ADAGYAPATPAAAG
Phaa5  (P56164)         ------------------------------------------------TPPTPRTPPLLPP
Bradi3g02660.1          ------------------------------------------------------YAPVTPATP
Bradi3g02665.1          ------------------------------------------------------YAPVTPATP
Horv5.0101(AAB41585)    GREHSAVPRRRNLVALVPRHGCYAEFSLYVCVGNINAPFPVFNRTTFIANAGIEAELEPH
Bradi2g48255.1          ------------------------------------------------------------

Phlp5 (Q40960)          AAPAGKATTEEQKLIEKINAGFKAALAAAAGVQPADKYRTFVATFGAASNKAFAEGLSGE
Lolp5 (Q40237)          AVPSGKATTEEQKLIEKINAGFKAAVAAAAVVPPADKYKTFVETFGTATNKAFVEGLASG
Lolp5 (Q40240)          AAAGGKATTDEQKLLEDVNAGFKAAVAAAANAPPADKFKIFEAAFSESSKGLLATS---A
Holl5.0201(CAB10766)    -----------QKLLEDVNASFKAAVAAAAKVPPADKYKTFLRAFTVLDRGSTEQSKAEE
Holl5.0101(CAB10765)    --AGGKATTDEQKLLEDVNAGFKTAVAAAANVPPADKYKTFEAAFTASSKASIAAA---A
Dacg5(CAD20402)         --AGGKAMTEEQTLIEDVNAGFKAAVAAASSAPPADKFKTFEATFTAACKANIAAA---A
Dacg5(CAD20403)         --AGGKAMTEEQTLIEDVNAGFKAAVAAASSAPPADKFKTFEATFTAACKANIAAA---A
Dacg5(CAD20404)         --AGGKAMTEEQTLIEDVNAGFKAAVAAASSAPPADKFKTFEATFTAACKANIAAA---A
Dacg5(CAD20405)         --AGGKAMTEEQKLIEDVNAGFKAAVAAASSAPPADKFKTFEATFTAACKANIAAA---A
Phlp5 (Q40963)          AAAGKAT-TEEQKLIEDINVGFKAAVAAAASVPAADKFKTFEAAFTS-----SSKAA--A
Poap5 (AAG42254)        AAAGKITPTQEQKLMEDINVGFKAAVAAAAGAPPADKFKTFQAAFSASVEASAAKLN--A
Phaa5  (P56164)         PRARDKATLTSRSVEDINAASRRPWWASVP---PADKFKTFADHVLCVPNADVTSAAT--
Bradi3g02660.1          AAGYPKASTDEQKMIEDINAGFKTAVAAANAAPPADKYNTFEAAFKKASTGYKAGVSA--
Bradi3g02665.1          AAGYPKASTDEQKMIEDINAGFKTAVAAANAAPPADKYNTFEAAFKKASTGYKAG-----
Horv5.0101(AAB41585)    FLLLLFTFSSSSSFFTLLKTMIHFTDRSDNKNKAMMRGREFRKAFAEVLKGAATGQIAGQ
Bradi2g48255.1          ------------------------------------------------------------

Phlp5 (Q40960)          PKGAAESSSKAALTSKLDAAYKLAYKTAEGATPEAKYDAYVATLSEALRIIAGTLEVHAV
Lolp5 (Q40237)          ----YADQSKNQLTSKLDAALKLAYEAAQGATPEAKYDAYVATLTEALRVIAGTLEVHAV
Lolp5 (Q40240)          AKAPGLIP-------KLDTAYDVAYKAAE-ATPEAKYDAFVTALTEALRVIAGALEVHAV
Holl5.0201(CAB10766)    TKMPELSS-------KLVDAYMAAFKASTGGTQEAKYDAFVTTLTEALRVIAGALEVHAV
Holl5.0101(CAB10765)    TKAPGLIP-------QLNAATNTAYAAAQGATPEAKYDAFVTTLTEALRVIAGALEVHAV
Dacg5(CAD20402)         TKVPLFVA-------KLDAAYAVAYKTATGPTPEAKYDAFVAALTEALRVIAGALEVHAV
Dacg5(CAD20403)         TKVPLFVA-------KLDAAYAVAYKTATGPTPEAKYDAFVAALTEALRVIAGALEVHAV
Dacg5(CAD20404)         TKVPLFVA-------KLDAAYAVAYKTATGPTPEAKYDAFVAALTEALRVIAGALEVHAV
Dacg5(CAD20405)         TKVPLFVA-------KLDAAYAVAYKTAAGPTPEAKYDAFVAALTEALRVIAGALEVHAV
Phlp5 (Q40963)          AKAPGLVP-------KLDAAYSVAYKAAVGATPEAKFDSFVASLTEALRVIAGALEVHAV
Poap5 (AAG42254)        AQAPGFVS-------HVAATSDATYKAAVGATPEAKFDSFVAAFTEALRIIAGVLKVHAV
Phaa5  (P56164)         -KAPQLKA-------KLDAAYRVAYEAAEGSTPEAKYDAFIAALTEALRVIAGAFEVHAV
Bradi3g02660.1          ---------------KLDTAYQAAYKAADGATPEAKYDVFVATLTEALRVVAGAVEAHAV
Bradi3g02665.1          ----------------------AAYKAADGATPEAKYDVFVATLTEALRVVAGAVEAHAV
Horv5.0101(AAB41585)    SSSMAKLSS------SLELSYKLAYDKAQGATPEAKYDAYVATLTESLRVISGTLEVHSV
Bradi2g48255.1          ------------------------------------------------------------

Phlp5 (Q40960)          KPAAEEVKV-------IPAGELQVIEKVDAAFKVAATAAN---------AAPANDKFTVF
Lolp5 (Q40237)          KPAAEEVKVGA-----IPAAEVQLIDKVDAAYRTAATAAN---------AAPANDKFTVF
Lolp5 (Q40240)          KPATEEVLAAK-----IPTGELQIVDKIDAAFKIAATAAN---------AAPTNDKFTVF
Holl5.0201(CAB10766)    KPATEEVPAAK-----IPAGDLQVVDKIDASFKIAATAAN---------AAPANDKFTVF
Holl5.0101(CAB10765)    KPATEEVGAAK-----IPAGELQIVDKIDAAFRIAATAAN---------AAPVNDKFTVF
Dacg5(CAD20402)         KPAAEEVPAAK-----IPAGELQIVDKIDAAYKIAATAAN---------AAPANDKFTVF
Dacg5(CAD20403)         KPAAEEVPAAK-----IPAGELQIVDKIDAAYKIAATAAN---------AAPANDKFTVF
Dacg5(CAD20404)         KPAAEEVPAAK-----IPAGELQIVDKIDAAYKIAATAAN---------AAPANDKFTVF
Dacg5(CAD20405)         KPAAEEVPAAK-----IPAGELQIVDKIDAAYKIAATAAN---------AAPANDKFTVF
```

```
Phlp5 (Q40963)          KPVTEEPGMAK-----IPAGELQIIDKIDAAFKVAATAAA---------TAPADDKFTVF
Poap5 (AAG42254)        KPITEETGAAK-----IPAGEQQIIDKIDAAFKVAATAAN---------AAPANDKFTVF
Phaa5  (P56164)         KPATEEVVAD-------PVGELQIVDKIDAAFKIAATAAN---------SAPANDKFTVF
Bradi3g02660.1          KPVAEELAAGAGGK--MPADELKIVDKIDAAFKTASTAAN---------AAPANDKFNVF
Bradi3g02665.1          KPVAEELAAGAGGK--MPADELKIVDKIDSAFKTASTAAN---------AAPANDKFNVF
Horv5.0101(AAB41585)    KPAAEEVKG-------VPAGELKAIDQVDAAFRTAATAAD---------AAPANDKFTVF
Bradi2g48255.1          ------------------------------------------------------------

Phlp5 (Q40960)          EAAFNDAIKASTGGAYESYKFIPALEAA-VKQAYAATVATAPEVKYTVFETALKKAITAM
Lolp5 (Q40237)          ENTFNNAIKVSLGAAYDSYKFIPTLVAA-VKQAYAAKQATAPEVKYTVSETALKKAVTAM
Lolp5 (Q40240)          ESAFNKALNECTGGAYETYKFIPSLEAA-VKQAYAATVAAAPEVKYAVFEAALTKAITAM
Holl5.0201(CAB10766)    ETAFNKALKESTGGAYESYKFIPSLEAA-VKQAYASTVAAAPEVKYAVFEAALTKAITAM
Holl5.0101(CAB10765)    EGAFNKAIKESTGGAYEAYKFIPSLETA-VKQAYAATVATAPEVKYTVFETALKKAITAM
Dacg5(CAD20402)         EGAFNKAIKESTGGAYESYKFIPTLEAA-VKQAYAATVAAAPEVKYAVFEAALTKAITAM
Dacg5(CAD20403)         EGAFNKAIKESTGGAYESYKFIPTLEAA-VKQAYAATVAAAPEVKYAVFEAALTKAITAM
Dacg5(CAD20404)         EGAFNKAIKESTGGAYESYKFIPTLEAA-VKQAYAATVAAAPEVKYAVFEAALTKAITAM
Dacg5(CAD20405)         EGAFNKAIKESTGGAYESYKFIPTLEAA-VKQAYAATVAAAPEVKYAVFEAALTKAITAM
Phlp5 (Q40963)          EAAFNKAIKESTGGAYDTYKCIPSLEAA-VKQAYAATVAAAPQVKYAVFEAALTKAITAM
Poap5 (AAG42254)        EAAFNNAIKESTGGAYDTYKSIPSLEAA-VKQAYAATIAAAPEVKFAVFKAALTKAITAM
Phaa5  (P56164)         EGAFNKAIKESTAGAYETYKFIPSLEAA-VKQAYGATVARAPEVKYAVFEAGLTKAITAM
Bradi3g02660.1          EATFNKALKESTGGAYESYKFIPSLEAA-VKQAYAATIAAAPEVKYAVFEAALTKAITAM
Bradi3g02665.1          EATFNKALKESTGGAYESYKFIPSLEAA-VKQAYAATIAAAPEVKYAVFEAALTKAITAM
Horv5.0101(AAB41585)    ESLQQGPSRKPRGGAYESYKFIPALEAA-VKQAYAATVAAAPEVKFTVFQTALSKAINAM
Bradi2g48255.1          -----MAGMSIDGKAYTDYKFIPPLDAA-FKEAYASTVAAKPEVKYAVFEAALTKAIGAM

Phlp5 (Q40960)          SEAQKAAKPAAAATATATAAVGAATGAATAATGGYKV-----------------------
Lolp5 (Q40237)          SEAEKEATPAAAATATPTPAAATATATPAAAYATATPAAATATATPAAATATPAAAGGYK
Lolp5 (Q40240)          TQAQKAGKPAAAAATAAATVATAAATAAAVLPPPLLVVQSLI------------------
Holl5.0201(CAB10766)    SQAQKVAQPAAAATGAATVAAGAATTAAGGYKV---------------------------
Holl5.0101(CAB10765)    SEAQKEAKPVAAATGAATAAAGVAAGAATAAAGGYKV-----------------------
Dacg5(CAD20402)         SEAQKVATPAAVATGAATAAASAATGAATAAAGGYKV-----------------------
Dacg5(CAD20403)         SEAQKVATPAAVATGAATAAASAATGAATAAAGGYKV-----------------------
Dacg5(CAD20404)         SEAQKVATPAAVATGAATAAASAATGAATAAAGGYKV-----------------------
Dacg5(CAD20405)         SEAQKVATPAAAATGAATAAASAATGAATAAAGGYKV-----------------------
Phlp5 (Q40963)          SEVQKVSQPATGAATVAAGAATTAAGAAS-------GAATVA------------------
Poap5 (AAG42254)        AEVQKVSKPVAGAATVAAGAATAATGAATGAAGAATGAATVS------------------
Phaa5  (P56164)         SEAQKVAKPP-LSPQPPQVLPLAAGGAATVAAASDVRVCRSHGTLQDACLLRCRGGCQPV
Bradi3g02660.1          SEAQKAKGGVTVTATVSAGAAAKPAAPGYKFYIYSLILIYHFY-----------------
Bradi3g02665.1          SEAQKAKGAAGYAGAAAGAGATVAAGAGAAVSAGGYKV----------------------
Horv5.0101(AAB41585)    TQAGKVAKPAAAATATATVAAGAAATAGNYKV----------------------------
Bradi2g48255.1          AEAEKAASTKPAA-----------------------------------------------

Phlp5 (Q40960)          ---------------
Lolp5 (Q40237)          V--------------
Lolp5 (Q40240)          ---------SLLIYY
Holl5.0201(CAB10766)    ---------------
Holl5.0101(CAB10765)    ---------------
Dacg5(CAD20402)         ---------------
Dacg5(CAD20403)         ---------------
Dacg5(CAD20404)         ---------------
Dacg5(CAD20405)         ---------------
Phlp5 (Q40963)          ---------AGGYKV
Poap5 (AAG42254)        ---------AGGYKV
Phaa5  (P56164)         VWRGGSHRARGGYKV
Bradi3g02660.1          ---------------
Bradi3g02665.1          ---------------
Horv5.0101(AAB41585)    ---------------
Bradi2g48255.1          ---------------
```

Light gray: conserved residues

Dark gray: Conservative substitutions

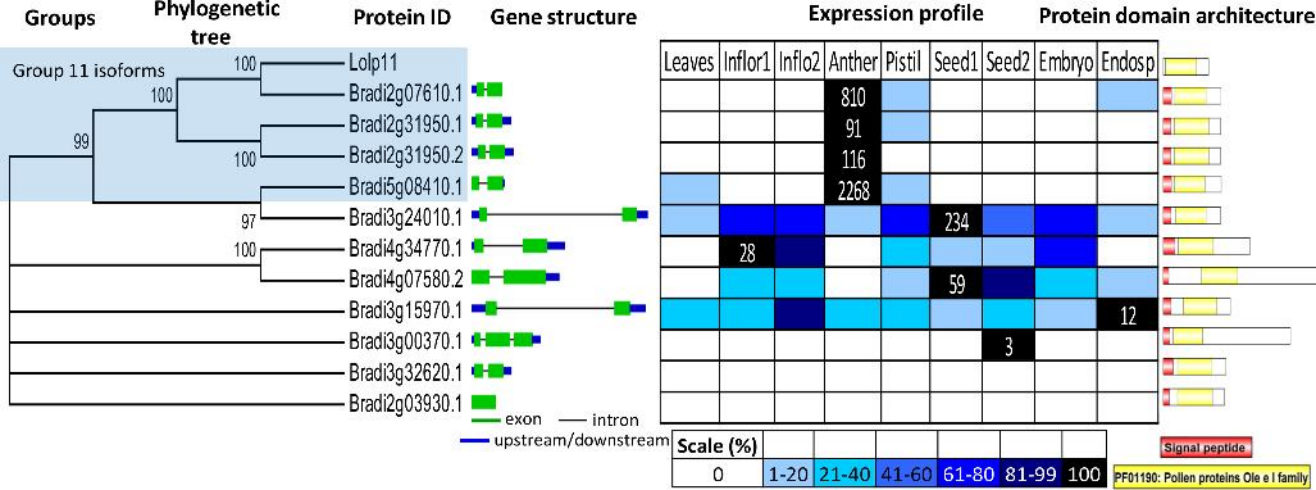
Red: IgE-binding residues


Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Leaves
Infio1
Infio2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
exon
intron
upstream/downstream

```
Phlp7 (O82040) --MADDMERIFKRFDTNGDGKISLSELTDALRTLGSTSADEVQRMMAEIDTDGDGFIDFN
Bradi4g21210.1 MAAAEDMERIFKRFDTNGDGKISLSELTDALRTLGSTSADEVQRMMAEIDTDGDGFIDFD
Cynd7 (P94092) MADTGDMEHIFKRFDTNGDGKISLAELTDALRTLGSTSADEVQRMMAEIDTDGDGFIDFD

Phlp7 (O82040) EFISFCNANPGLMKDVAKVF
Bradi4g21210.1 EFISFCNANPGLMKDVAKVF
Cynd7 (P94092) EFISFCNANPGLMKDVAKVF
```

Yellow: Two Calcium-binding domains (EF-hands)

Gray: Conserved residues


Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Group 11 isoforms
100
100
99
100
97
100
Lolp11
Bradi2g07610.1
Bradi2g31950.1
Bradi2g31950.2
Bradi5g08410.1
Bradi3g24010.1
Bradi4g34770.1
Bradi4g07580.2
Bradi3g15970.1
Bradi3g00370.1
Bradi3g32620.1
Bradi2g03930.1
exon
intron
upstream/downstream
Leaves
Inflor1
Inflo2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
810
91
116
2268
234
28
59
12
3
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100
Signal peptide
PF01190: Pollen proteins Ole e I family

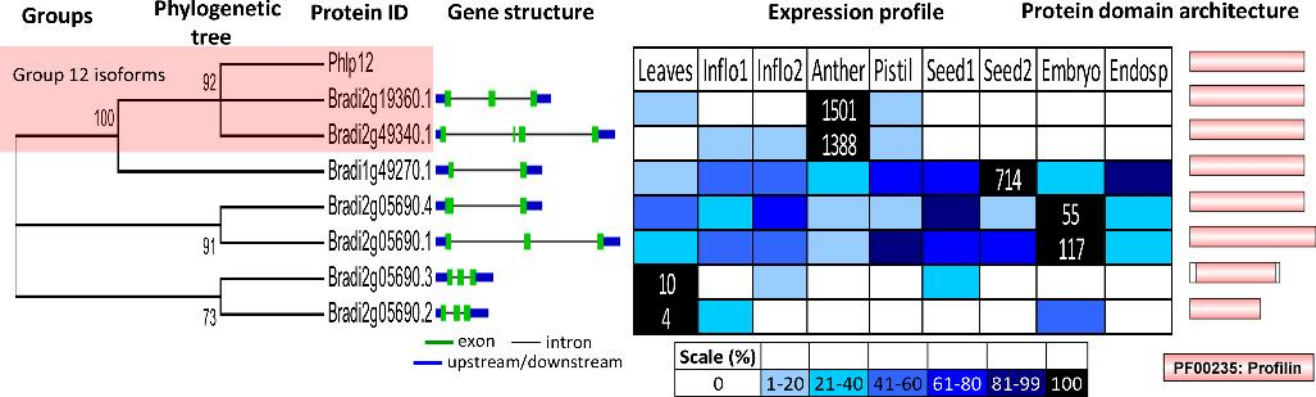
```
Bradi2g31950.1        QGYSKPGFLVTGRVYCDTCRAGFETNASHSIPGAVVQMECRHFETNELHHKAEATTDAQG
Bradi2g31950.2        QGYSKPGFLVTGRVYCDTCRAGFETNASHSIPGAVVQMECRHFETNELHHKAEATTDAQG
Lolp11(Q7M1X5)        --DKGPGFVVTGRVYCDPCRAGFETNVSHNVEGATVAVDCRPFDGGESKLKAEATTDKDG
Phlp11(AAN32987)      --DKGPGFVVTGRVYCDPCRAGFETNVSHNVQGATVAVDCRPFNGGESKLKAEATTDGLG
Bradi2g07610.1        YEKGPGGFVVTGRVYCDPCRAGFETNVSKNIGGATVAVDCRPFNGGDSKLKAEATTDQYG
Bradi5g08410.1        --VDLPDYIVQGRVYCDTCRAGFETNVTTYIKGAKVRLECKRFGTEKVERALDGVTDETG

Bradi2g31950.1        WYKMEVGEDHQEEICEVALLRSPEKDCAEIEKSRDRCRVPLTRNNGIKQSGVRYANPIAF
Bradi2g31950.2        WYKMEVGEDHQEEICEVALLRSPEKDCAEIEKSRDRCRVPLTRNNGIKQSGVRYANPIAF
Lolp11(Q7M1X5)        WYKIEIDQDHQEEICEVVLAKSPDKSCSEIEEFRDRARVPLTSNXGIKQQGIRYANPIAF
Phlp11(AAN32987)      WYKIEIDQDHQEEICEVVLAKSPDTTCSEIEEFRDRARVPLTSNNGIKQQGIRYANPIAF
Bradi2g07610.1        WYKIDIDQDHQEEICEVLLARSPDPACSEIEEFRDRARVPLTRNNGLKQQGTRYANPIAF
Bradi5g08410.1        TYKIELKDSHPEDICEVVLIQSPLADCNKIQALRDRARVELTRNIGISDN-LRLANSLGY

Bradi2g31950.1        FRKEPLANCGDVLRKYDLYDETSENS---
Bradi2g31950.2        FRKEPLANCGDVLRKYDLYDETSENS---
Lolp11(Q7M1X5)        FRKEPLKECGGILQAY-------------
Phlp11(AAN32987)      FRKEPLKECGGILQAYDLRDAPETP----
Bradi2g07610.1        FRKEPLKDCGGILQAYDLKDAPETP----
Bradi5g08410.1        LKDVPLPVCAQLLKQFKSADDDDDDQVKA
```

Green: Conserved cysteine residues

Turquoise: N-glycosylation sites

Light gray: conserved residues

Dark gray: Conservative substitutions

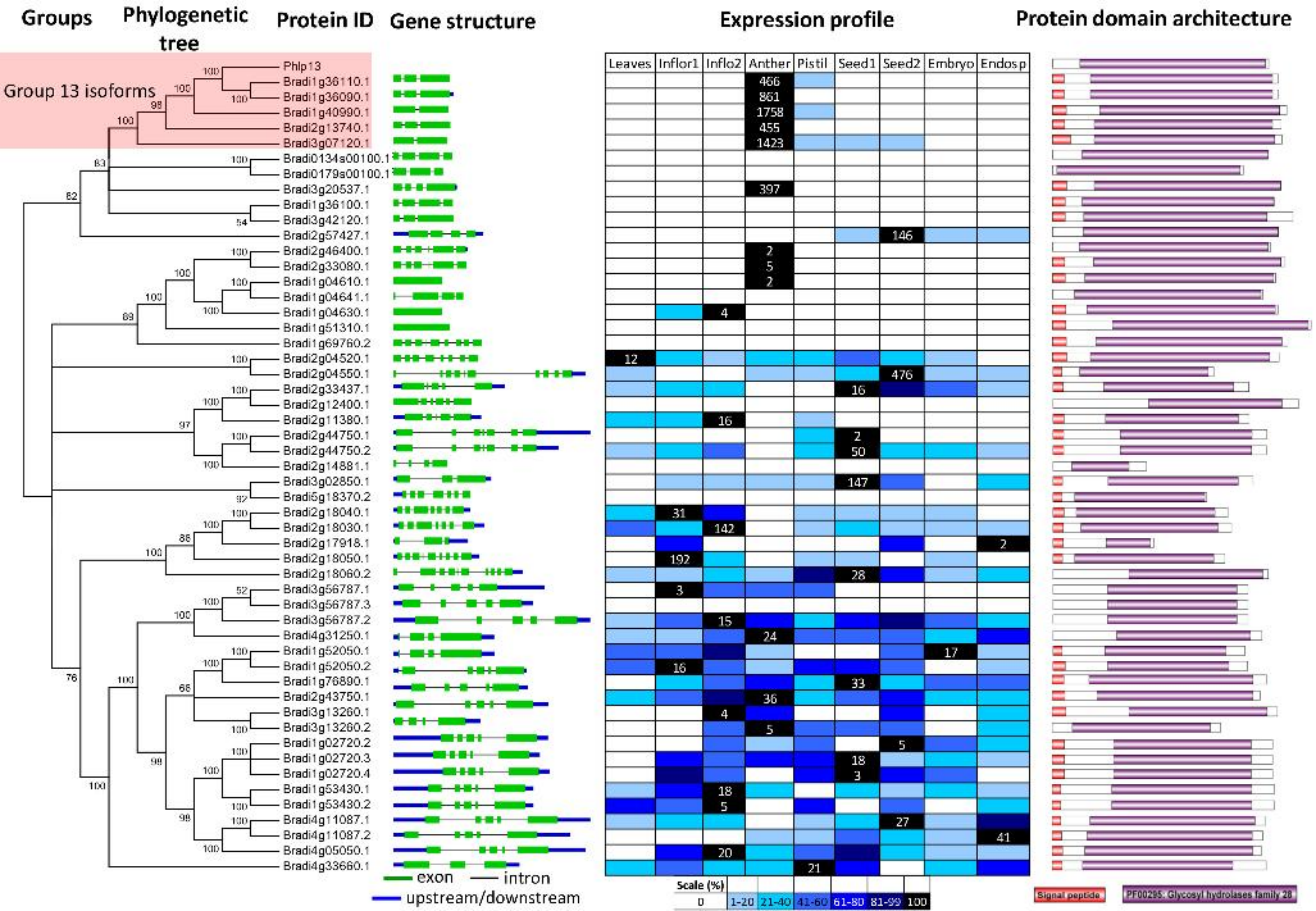

Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Group 12 isoforms
92
100
91
73
Phlp12
Bradi2g19360.1
Bradi2g49340.1
Bradi1g49270.1
Bradi2g05690.4
Bradi2g05690.1
Bradi2g05690.3
Bradi2g05690.2
exon
intron
upstream/downstream
Leaves
Inflo1
Inflo2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
1501
1388
714
55
117
10
4
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100
PF00235: Profilin

```
Zeam12 (P35081)    MSWQTYVDEHLMCEIEGHHLTSAAIVGHDGATWAQSTAFPEFKPEEMAAIMKDFDEPGHL
Zeam12 (P35082)    MSWQAYVDEHLMCEIEGHHLAAAAIVGHDGAAWAQSTAFPEFKTEDMANIMKDFDEPGHL
Cynd12 (O04725)    MSWQAYVDDHLMCEIEGHHLTSAAIIGHDGTVWAQSAAFPAFKPEEMANIMKDFDEPGFL
Orys12 (AAG32056)  MSWQTYVDEHLMCEIEGHHLTSAAIVGHDGTVWAQSAAFPQFKPEEMTNIMKDFDEPGFL
Zeam12 (P35083)    MSWQTYVDEHLMCEIEGHHLSSAAIVGHDGAVWAQSTAFPQFKPEEMTNIIKDFDEPGFL
Phlp12 (P35079)    MSWQTYVDEHLMCEIEGHHLASAAILGHDGTVWAQSADFPQFKPEEITGIMKDFDEPGHL
Bradi2g49340.1     MSWQTYVDEHLMCDIEGHHLVSAAILGHDGTVWAQSADFPSFKPEEMTNIMKDFDEPGTL
Bradi2g19360.1     MSWQTYVDEHLMCEIEGHHLGSAAILGHDGTVWAQSAAFPAFEPKEMTDIMKDFDEPGHL

Zeam12 (P35081)    APTGLILGGTKYMVIQGEPGAVIRGKKGSGGITVKKTGQSLIIGIYDEPMTPGQCNLVVE
Zeam12 (P35082)    APTGLFLGPTKYMVIQGEPGAVIRGKKGSGGITVKKTGQALVVGIYDEPMTPGQCNMVVE
Cynd12 (O04725)    APTGLFLGPTKYMVIQGEPGAVIRGKKGSGGVTVKKTGQALVIGIYDEPMTPGQCNMVIE
Orys12 (AAG32056)  APTGLFLGPTKYMVIQGEPGAVIRGKKGSGGITVKKTGQALVVGIYDEPMTPGQCNMVVE
Zeam12 (P35083)    APIGLFLGPTKYMVIQGEPGAVIRGKKGSGGITVKKTGQALVIGIYDEPMTPGQCNMVVE
Phlp12 (P35079)    APTGMFVAGAKYMVIQGEPGRVIRGKKGAGGITIKKTGQALVVGIYDEPMTPGQCNMVVE
Bradi2g49340.1     APTGLFLASAKYMVIQGEPGAVIRGKKGSGGITLKKTGQALVVGIYDEPMTPGQCNMVVE
Bradi2g19360.1     APTGMFLGGAKYMVIAGEPGAVIRGKKGSGGITIKKTGQALVIGIYDEPMTPGQCNMVVE

Zeam12 (P35081)    RLGDYLLEQGM
Zeam12 (P35082)    RLGDYLLEQGM
Cynd12 (O04725)    KLGDYLIEQGM
Orys12 (AAG32056)  RLGDYLVEQGL
Zeam12 (P35083)    RLGDYLVEQGL
Phlp12 (P35079)    RLGDYLVEQGM
Bradi2g49340.1     RLGDYLVEQGM
Bradi2g19360.1     RLGDYLVEQGM
```

Yellow: Conserved cysteine residues

Light Gray: conserved residues

Dark Gray: conserved substitutions


Groups
Phylogenetic tree
Protein ID
Gene structure
Expression profile
Protein domain architecture
Group 13 isoforms
Phlp13
Bradi1g36110.1
Bradi1g36090.1
Bradi1g40990.1
Bradi2g13740.1
Bradi3g07120.1
Bradi0134s00100.1
Bradi0179s00100.1
Bradi3g20537.1
Bradi1g36100.1
Bradi3g42120.1
Bradi2g57427.1
Bradi2g46400.1
Bradi2g33080.1
Bradi1g04610.1
Bradi1g04641.1
Bradi1g04630.1
Bradi1g51310.1
Bradi1g69760.2
Bradi2g04520.1
Bradi2g04550.1
Bradi2g33437.1
Bradi2g12400.1
Bradi2g11380.1
Bradi2g44750.1
Bradi2g44750.2
Bradi2g14881.1
Bradi3g02850.1
Bradi6g18370.2
Bradi2g18040.1
Bradi2g18030.1
Bradi2g17918.1
Bradi2g18050.1
Bradi2g18060.2
Bradi3g56787.1
Bradi3g56787.3
Bradi3g56787.2
Bradi4g31250.1
Bradi1g52050.1
Bradi1g52050.2
Bradi1g76890.1
Bradi2g43750.1
Bradi3g13260.1
Bradi3g13260.2
Bradi1g02720.2
Bradi1g02720.3
Bradi1g02720.4
Bradi1g53430.1
Bradi1g53430.2
Bradi4g11087.1
Bradi4g11087.2
Bradi4g05050.1
Bradi4g33660.1
exon
intron
upstream/downstream
Leaves
Inflor1
Inflo2
Anther
Pistil
Seed1
Seed2
Embryo
Endosp
Scale (%)
0
1-20
21-40
41-60
61-80
81-99
100
Signal peptide
PF00295: Glycosyl hydrolases family 28

```
Bradi1g36110.1        ------KAAEKEAAGSAAAASGPGATLDIVKLG---------ATGDGKTDSTKAVEEAWA
Bradi1g36090.1        ------KAAEKEAAGSAAAASGPGATLDIVKLG---------ATGDGKTDSTKAVEEAWA
Phlp13 (CAB42886)     ---GKKEEKKEEKKESGDAASGADGTYDITKLG---------AKPDGKTDCTKEVEEAWA
Zeam13 (89892725)     --------EKAKSKDNDAKASGPGGSFDITKLG---------ASGNGKTDSTKAVQEAWA
Bradi1g40990.1        --EKKEKKEEKGGKDKEADAEGPAAEGPSSGGGELDISKSGKCKGDGKADCTEALEEAWA
Bradi2g13740.1        KRPVAKESAGPAPAPAPEEASGDGGTFDITKLG---------AASDGKTDCSKELEEAWK
Bradi3g07120.1        -----------TYAAAGGPPSVPAGPLDIVLLG---------AKGDGKTDATEAVQKAWK

Bradi1g36110.1        SACGGTGKQTILIPKGDFLTGPLNFTGPCTG-DITIQVDGNLLGSNDLAKYKAN--WIEI
Bradi1g36090.1        SACGGTGKQTILIPKGDFLTGPLNFTGPCTG-DITIQVDGNLLGSNDLAKYKAN--WIEI
Phlp13 (CAB42886)     SACGGTGKNTIVIPKGDFLTGPLNFTGPCKGDSVTIKLDGNLLSSNDLAKYKAN--WIEI
Zeam13 (89892725)     SACGGTGKQTILIPKGDFLVGPLNFTGPCKG-DVTIQVNGNLLATTDLSQYKDHGNWIEI
Bradi1g40990.1        SACKGTGKQTIQIPKGDYLTGPLNFTGPCTG-DVTIQLDGNLLGSTDMALYKSN--WIEI
Bradi2g13740.1        SACGAAGQHTVVIPKGEFLCGPVNFSGPCEG-DVTIQIDGTLLATNDLPKYGGG-SWINI
Bradi3g07120.1        NACAATGVQKIIIPPGNFLIGALALEGPCTS-SIIIRLDGNLLGTGDLNAYKKN--WIEV

Bradi1g36110.1        MRVKNLVITGK-GKLDGQGPAVWGKNSCAKNYNCKILPNTLVLDFCNNTLIEGITLLNSK
Bradi1g36090.1        MRVKNLVITGK-GKLDGQGPAVWGKNSCAKNYNCKILPNTLVLDFCNNTLIEGITLLNSK
Phlp13 (CAB42886)     MRIKKLTITGK-GTLDGQGKAVWGKNSCAKNYNCKILPNTLVLDFCDDALIEGITLLNAK
Zeam13 (89892725)     LRVDNLVITGK-GKLDGQGPAVWSKNSCVKKYDCKILPNSLVMDFVNNGEVSGITLLNSK
Bradi1g40990.1        MRVENLVISGK-GTLDGQGPKVWSKNACAKKYDCKILPNSLVLDFITNGTISGITLLNAK
Bradi2g13740.1        LKVDNLVITGS-GTLDGQGSNVWTKDPQAKAG----YPNTLVLDYVKNGSVSGITLLNSK
Bradi3g07120.1        MHVDNFAINGH-GTIDGQGPLVWQKNQCNKNYNCKILPNSLVLDYVTNASIRGVTLKNAK

Bradi1g36110.1        FFHLNIYECRGITVKDVTITSPGDSPNTDGIHMGDASNITITDTKIGTGDDCISIGPGTS
Bradi1g36090.1        FFHLNIYECRGITVKDVTITSPGDSPNTDGIHMGDASNITITDTKIGTGDDCISIGPGTS
Phlp13 (CAB42886)     FFHMNIYECKGVTVKDVTITAPGDSPNTDGIHIGDSSKVTITDTTIGTGDDCISIGPGST
Zeam13 (89892725)     FFHMNMYKCKDMLIKDVNVTAPGDSPNTDGIHMGDSSGVTITNTVIGVGDDCISIGPGTS
Bradi1g40990.1        FFHMNVFQCKGVTIEDVTVTAPGDSPNTDGIHIGDSSGITITGTTIGVGDDCISIGPGST
Bradi2g13740.1        FFHMNIYMSSDVKIDNVTITAPGDSPNTDGIHIGDSSNIHVTGATIGTGDDCISIGGGSA
Bradi3g07120.1        FFHMNLFNCKNVVVDKVNITAPGDSPNTDGIHMGDSDNITITNTNIGVGDDCISIGPGTQ

Bradi1g36110.1        GVNISGVNCGPGHGISVGSLGRYKDEKDVTDITVKNCVLNKSTNGVRIKSYEDAKSPLVA
Bradi1g36090.1        GVNISGVNCGPGHGISVGSLGRYKDEKDVTDITVKNCVLNKSTNGVRIKSYEDAKSPLVA
Phlp13 (CAB42886)     GLNITGVTCGPGHGISVGSLGRYKDEKDVTDITVKNCVLKKSTNGLRIKSYEDAKSPLTA
Zeam13 (89892725)     KVNITGVTCGPGHGISIGSLGRYKDEKDVTDINVKDCTLKKTANGVRIKAYEDAASVLTA
Bradi1g40990.1        KINITGVTCGPGHGISIGSLGRYKDEKDVTDINVKDCTLKKATNGLRIKSYQSAESSLTA
Bradi2g13740.1        SITVTGVTCGPGQGISVGCLGRYKDEKDVSDVTVKDCVLRSSTNGVRIKTYVDAVKSITA
Bradi3g07120.1        RVRVHGSRCGPGHGISVGSLGRYKDEKNVEDIQVTNCTIKGATNGLRIKSYEDSKSVLKA

Bradi1g36110.1        SKLTYENIKMMDVGYPIIIDQKYCPNKICPPKPNSAKVTVKDVTFRNITGTSSTPEAVSL
Bradi1g36090.1        SKLTYENIKMMDVGYPIIIDQKYCPNKICPPKPNSAKVTVKDVTFRNITGTSSTPEAVSL
Phlp13 (CAB42886)     SKLTYENVKMEDVGYPIIIDQKYCPNKICTSKGDSARVTVKDVTFRNITGTSSTPEAVSL
Zeam13 (89892725)     SKIHYENIKMEDSGYPIIIDMKYCPNKLCTANG-ASKVTVKDVTFKNITGTSSTPEAVNL
Bradi1g40990.1        SEIHYENVNMEDVANPIIIDMNYCPNKICPAKG-SSHVTIKDVTFKNITGTSSTPEAVSL
Bradi2g13740.1        SNLTFENIKMEDVANPIIIDQNYCPEKICTAKS-KSAVTVKDIIFRNITGTSSTPEAVSL
Bradi3g07120.1        SRFVYDQIVMDNVSFPIVIDQKYCPNNICNKNG-QSTVTIQDIVFKNIVGTSATPEAVTL

Bradi1g36110.1        LCSDKQPCSGVQMFDVKVEYSGT-NNKTMAVCTNAKVTAKGCSEALACAA--
Bradi1g36090.1        LCSDKQPCSGVQMFDVKVEYSGT-NNKTMAVCTNAKVTAKGCSEALACAA--
Phlp13 (CAB42886)     LCSDKQPCNGVTMNDVKIEYSGT-NNKTMAVCTNAKVTAKGVSEANTCAA--
Zeam13 (89892725)     LCSAKIPCTGVTMDDVNIKYSGT-NNKTMAVCKNAKGSAKGCLKELACF---
Bradi1g40990.1        LCSDKLPCSGVELNDVKVEYSGK-NNKTMAVCKNAKGTAKGCLEALACL---
Bradi2g13740.1        LCSEKQPCSGVELIDVNVEYSGK-NNKTMGVCTNAKGIAKDTLQALACVSDV
Bradi3g07120.1        NCPNNLPCQGVQLVNVNLKYVGARNNKTMAVCHNAVGKSTNVAKELACL---
```

Yellow: Conserved polygalacturonase domains

Light Gray: conserved residues

Dark Gray: conserved substitutions

Green: cysteine residues (14 conserved) Red: N-glycosylation sites