

Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: An *in silico* analysis

John A. Jenkins, PhD,^a Sam Griffiths-Jones, PhD,^b Peter R. Shewry, PhD,^c
Heimo Breiteneder, PhD,^d and E.N. Clare Mills, PhD^a Norwich, Cambridge,
and Harpenden, United Kingdom, and Vienna, Austria

Background: The body of sequence and structural information on allergens and the sequence analysis of whole plant genomes are facilitating the application of bioinformatic approaches to identifying and defining plant allergens.

Objective: An *in silico* approach was used to quantify the distribution of plant food allergen sequences across protein families and to develop and apply a novel means of assessing conserved surface features important for IgE cross-reactivity. **Methods:** Plant food allergen sequences were classified into Pfam families on the basis of sequence homology. Contact surface areas of selected proteins were calculated with MOLMOL by using a 1.4-Å probe, corrected by removing contributions from IgE inaccessible main chains and side chains forming the ligand binding sites.

Results: A set of 129 food allergen sequences were classified into only 20 of 3849 possible Pfam families, with 4 families accounting for more than 65% of food allergens. Structural bioinformatic analysis of conserved exterior main chains and amino acid side chains in cross-reactive homologues of Bet v 1 and nonspecific lipid transfer proteins showed higher levels of similarity than shown by simple sequence comparisons. Thus, 75% of the Mal d 1 surface is likely to bind anti-Bet v 1 antibodies, compared with a sequence identity of ~56%. **Conclusion:** Most plant food allergens belong to only 4 structural families, indicating that conserved structures and biological activities may play a role in determining or promoting allergenic properties. Structural bioinformatic analysis shows that conservation of 3-dimensional structure should be included in any assessment of potential IgE cross-reactivity in, for example, novel proteins. (J Allergy Clin Immunol 2005;115:163-70.)

Key words: Allergen bioinformatics, distribution of plant food allergens across protein families, plant food allergen structures, Bet v 1 family cross-reactivity, nsLTP family cross-reactivity

One of the major challenges of molecular allergology is to pinpoint and evaluate characteristics of a protein that confer its allergenic properties when it comes into contact with an atopic immune system. It is clear that in some cases, the biochemical properties of the protein itself are important, such as the stability and proteolytic activity of the inhalant allergen Der p 1 from house dust mite.¹ However, our current knowledge of food allergens is much less complete. For example, we do not know why peanuts pose so much more of an allergenic risk than peas, despite their close botanical relationship. These gaps in understanding prevent us from giving rational advice on diverse issues, including the allergenic potential of novel foods and processes such as genetically modified organisms.

Both the ability of a substance to sensitize a naive individual and to elicit an allergic response in individuals who are already sensitized to a given protein are encompassed in its allergenicity. Although the molecular basis for these effects is still not completely understood, it is becoming evident that the level of exposure and the properties of the allergen itself play important roles in determining the allergenic potential.²

The last 10 years have seen an explosion in the identification and sequencing of food allergens. The sequences of more than 100 different food allergens have now been published, corresponding to 70% to 80% of the identified food allergens. When isoallergens are included, the number of reported sequences rises to more than 200 (www.ifr.ac.uk/protall, www.allergen.org, www.allergome.org, www.allergenonline.com). Bioinformatic tools are available to classify proteins into families on the basis of their shared amino acid sequences and conserved 3-dimensional structures. These form the basis of several protein family databases, including Pfam.³ Previous analyses of plant food allergens have indicated that most belong to only a small number of protein superfamilies,⁴ contributing to the development of a more molecular approach to the classification of plant food allergens.⁵ These studies have facilitated the application of *in silico* approaches to predict the allergenic potential of proteins directly from their amino acid sequence. However, Hileman et al⁶ reported that searches using a 6-amino acid window wrongly predicted that 41 of 50 corn proteins

From ^athe Institute of Food Research, Norwich; the ^bWellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge; ^cRothamsted Research, Harpenden; and ^dDepartment of Pathophysiology, Medical University of Vienna.

Partly supported by EU through FAIR CT-98-4356 (PROTALL), QLK1-CT-2000-01394 (SAFE), and BBSRC through the competitive strategic grant to the Institute of Food Research and Rothamsted Research. Dr Griffiths-Jones is supported by the Wellcome Trust. Dr Breiteneder is supported by Austrian Science Fund grant SFB F01802.

Received for publication August 31, 2004; revised October 6, 2004; accepted for publication October 7, 2004.

Reprint requests: Heimo Breiteneder, PhD, Department of Pathophysiology, Medical University of Vienna, AKH-EBO-3Q, Währinger Gürtel 18-20, 1090 Vienna, Austria. E-mail: Heimo.Breiteneder@meduniwien.ac.at.

0091-6749/\$30.00

© 2005 American Academy of Allergy, Asthma and Immunology

doi:10.1016/j.jaci.2004.10.026

Abbreviations used

nsLTP: Nonspecific lipid transfer protein
PLAT: Polycystin-1-lipoxygenase- α -toxin

were potential allergens. In contrast, the FASTA algorithm identified only 9 corn proteins which shared >35% identity over 80 amino acids with known allergens, with only 6 them similar over their entire length. Soeria-Atmadja et al⁷ and Stadler and Stadler⁸ have used local alignment and MEME, a motif discovery tool, in automated learning approaches to achieve good predictions and relatively low rates of false-positive identifications of allergens.

An alternative approach to allergen classification may be developed from the observation that most plant food allergens belong to only a small number of protein superfamilies.^{4,5} We have therefore exploited profile-based sequence homology methods³ to classify plant food allergens into families and then used a structural bioinformatic approach to analyzing conserved surface features that are thought to be important determinants of IgE cross-reactivity at the molecular level.

METHODS**Plant food allergen databases**

Lists of known plant food allergens were obtained from the PROTALL (www.ifr.ac.uk/protall) and FARRP (www.allergenonline.com) databases. The version of FARRP used defined proteins from plant foods as food allergens even when they had been found to act only as aeroallergens, such as Gly m 1. The PROTALL set was limited to proteins shown to bind serum IgE from at least 3 patients with clinical allergy to the food from which the allergen originated. Removing duplications between the 2 databases left a protein set composed of 136 entries mapped to SWISS-PROT and TrEMBL entries.⁹ A separate set of known pollen allergens was obtained from FARRP and mapped to 152 nonredundant SWISS-PROT and TrEMBL entries.

Protein family assignment in plant genomes

Arabidopsis thaliana. The *Arabidopsis* protein set was downloaded from the Proteome Web site (www.ebi.ac.uk/protome).¹⁰ The 2 allergen protein sets and the *Arabidopsis* protein set were searched against the Pfam 7.3 collection of hidden Markov model profiles characteristic of a family³ (www.sanger.ac.uk/Software/Pfam/) by using the HMMER software package, version 2.2g (hmmerr.wustl.edu/). The set of Pfam thresholds chosen using the lowest score for sequences known to belong to the family and the highest score for sequences not belonging to the family were used to give protein family classifications. The Pfam database classifies plant protein sequences into families on the basis of sequence homology, which is related to conserved 3-dimensional structures and possible function. Pfam 7.3 contains 3849 families, which include more than 70% of all plant protein sequences reported in the SWISS-PROT and TrEMBL databases. Where multiple domains are found within a single protein, these are counted only once, but nonidentical domains contribute to the scores of both families.

Oryza sativa. Currently there is no public domain protein set available for rice analogous to that available for *Arabidopsis*. Consequently, the assembled DNA sequence of *O sativa*¹¹ was obtained from Genbank (accession AAAA000000000) and searched against Pfam 7.3 by using the WISE2 software package version 2.1.22c (www.sanger.ac.uk/Software/Wise2/). The GENEWISEDB program was used to search for high-scoring (≥ 30 bits) full-length matches (global search mode) by using the alternate splice model (GT/AG), as suitable for plant genomes. Families were ranked only for *A thaliana* because the rice genome could be searched only with existing families that we had, rather than with the whole protein family universe. Each set is available together with the results of our analysis as supplementary information in the Journal's Online Repository (and also available online at www.sanger.ac.uk/Users/sgj/JACI_suppl).

Analysis of allergen surfaces

The 3-dimensional structures and contact surface areas of proteins were calculated with MOLMOL¹² by using a 1.4-Å radius probe. Surface areas were assigned to the nearest atom, and main chains and side chains were manually checked both for sequence conservation and to exclude main chains and side chains that contributed only to the surface of the IgE inaccessible lipid binding cavities (residues with no exposure to a probe of 3.5-Å radius included all IgE inaccessible residues in nonspecific lipid transfer proteins (nsLTPs) and Bet v 1a but also several in shallow depressions on the exterior surface). Sequences were aligned with T-Coffee¹³ for Bet v 1 homologues, and the insertion in corn nsLTP was positioned by comparing the wheat and barley structures.

RESULTS**Bioinformatic analysis of allergen sequences**

The Pfam database was used to assign 136 plant food allergen sequences compiled by using the FARRP and PROTALL databases. Most of these sequences are of clinically proven food allergens, but because inclusion criteria differ slightly between the databases, some sequences of proteins that are not proven food allergens were included in the data set. All but 7 sequences could be unambiguously assigned to Pfam families, with the 7 unmatched sequences mainly corresponding to short sequenced fragments of proteins. It is striking that the 129 matched sequences fall into only 20 of a possible 3849 Pfam families, with only 4 of these families accounting for more than 65% of the sequences (Fig 1, A, black bars). These 4 families are the cereal prolamin superfamily (composed of cereal storage proteins [eg, the ω -5 gliadin allergen of wheat], nonspecific lipid transfer proteins [such as Pru p3 from peach and Zea m 14 from corn], 2S storage albumins [eg, Ber e 1 from Brazil nut] and inhibitors of trypsin and α -amylase); the cupins (including the 7S and 11S globulin storage proteins of seeds, to which many allergens belong, notably Ara h 1, 3, and 4 from peanut); homologues of the major birch pollen allergen, Bet v 1 (such as Mal d 1 in apple and Api g 1 in celery); and profilins (such as Api g 4 from celery).

Allergens from the same species with greater than 67% sequence identity have been defined by the Allergen Nomenclature Committee as isoallergens.¹⁴ To ensure that the distribution of sequences was not biased by the inclusion of isoallergens (as is the case, for example, for

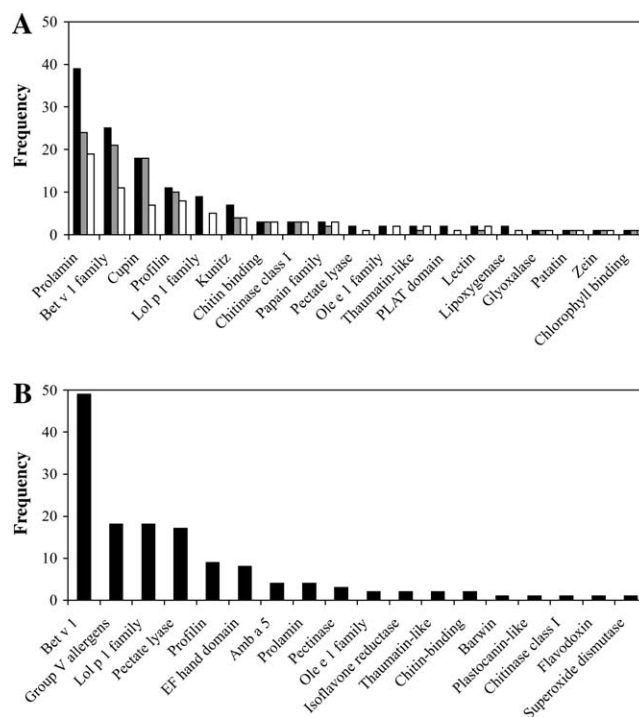


FIG 1. Distribution within protein families, as classified by Pfam using allergen sequences from the PROTALL and FARRP databases. **A**, PROTALL-FARRP (black); PROTALL food allergen (gray); PROTALL-FARRP (single sequence per allergen; white); **B**, FARRP respiratory allergen set.

various Bet v 1 homologues), sequence sets were edited to include only 1 sequence for each isoallergen. This analysis confirmed that the plant food allergens fell into the same protein families, although the ranking of the cupins and profilins was reversed (Fig 1, A, white bars). This is because, in contrast with the profilins, many of the cupin allergen sequences were derived from a single plant species, soybean. Restricting the analysis to the PROTALL sequence set, which includes only food allergen sequences, did not change the distribution of the main allergen families (Fig 1, A, gray bars). However, it did result in the loss of some allergen families, such as the Lol p 1 homologues, pectate lyases, and polycystin-1-lipoxygenase- α -toxin (PLAT) domain homologues. These proteins are not proven food allergens, but, by virtue of their being inhalant allergens from plants used as food, are included in the FARRP allergen sequence database and in the combined FARRP-PROTALL sequence set.

Many families of respiratory allergens have allergenic homologues in plant foods, notably Bet v 1, profilins, the olive pollen allergen Ole e 1, and chitin-binding and thaumatin-like proteins.⁵ Therefore, for the sake of comparison, a similar survey of the 152 sequences of plant respiratory allergens included in the FARRP database allowed 147 to be assigned to 18 Pfam families. As with the food allergens, a high proportion (more than 70%) fell into only 4 families (Fig 1, B), including the Bet v 1 and grass pollen allergen (Lol p 1) families. When considered together, these data show that, when combined,

plant food and respiratory allergens fall into only 29 Pfam families.

Complete plant genomes

Not all protein families are equally represented in plant genomes, and hence, the distribution of plant food allergens could simply be a reflection of the relative sizes of the prolamin, cupin, Bet v 1, and profilin protein families found in plants. Consequently, Pfam searches on the proteins predicted from the entire genome sequences of *A thaliana* (a model dicotyledonous plant species related to cultivated brassicas like cabbages) and the monocotyledonous cereal rice (*O sativa*) were performed. These genomes have been predicted to encode approximately 25,500 (*Arabidopsis* Genome Initiative 2000)¹⁵ and 50,000¹¹ proteins, respectively. In both species, the most prevalent Pfam family is the protein kinase family, with 985 predicted members in *Arabidopsis* and 1423 in rice. This family contains no known allergens in any plant species.

This analysis allowed the distribution of Pfam families containing characterized food or inhalant allergens to be assessed within these genomes. They could be ranked according to their abundance (Table I) for *Arabidopsis*, for which the complete universe of protein families has been defined. The allergen family with the greatest number of members in the entire plant genomes comprised proteins with the EF hand Ca^{2+} binding domain (132 members in *Arabidopsis*, 176 in rice), although it contained only 8,

TABLE I. Distribution of known plant food and pollen allergens among protein families as defined by Pfam compared with family counts in *A thaliana* and *O sativa* genomes*

Family description	Plant food allergens		Pollen allergens		<i>A thaliana</i>		<i>O sativa</i>
	Members, N	Rank	Members, N	Rank	Members, N	Genome rank	Members, N
Prolamin	39	1	4	6=	87	36	140
Bet v 1 family	25	2	49	1	4	614=	2
Cupin	18	3	0	—	11	305=	88
Profilin	11	4	9	4	0	—	3
Lol p 1 family	9	5	17	2	1	1129=	69
Pectate lyase	2	6	16	3	24	144=	14
Kunitz	7	7	0	—	7	431=	2
Chitin binding	3	8=	2	8=	13	258=	19
Chitinase I	3	8=	1	9=	18	195=	19
Papain family	3	8=	0	—	31	116=	45
PLAT domain	2	9=	0	—	8	385=	11
Lectin	2	9=	0	—	48	80=	34
Lipoxygenase	2	9=	0	—	7	431=	13
Ole e 1 family	2	9=	2	8=	13	258=	9
Thaumatococin-like	1	9=	2	8=	24	144=	43
Glyoxalase	1	10=	0	—	10	336=	24
Group V allergen	0	—	17	2=	0	—	0
Patatin	1	10=	0	—	9	355=	34
Zein	1	10=	0	—	0	—	0
Chlorophyll binding	1	10=	0	—	21	163=	19
EF hand	0	—	8	5	132	17	176
Amb a 5	0	—	4	6=	0	—	0
Pectinase	0	—	3	7	62	56=	35
Isoflavone reductase	0	—	2	8=	7	431=	12
Barwin	0	—	1	9=	0	—	5
Plastocyanin-like	0	—	1	9=	33	106=	58
Flavodoxin	0	—	1	9=	8	383=	13
Superoxide dismutase	0	—	1	9=	4	614=	6

*“=” Signifies an equal ranking.

exclusively respiratory, allergens. This was followed by the prolamin superfamily, which, although identified as the family containing the greatest number of food allergens, is only the 36th most abundant Pfam family in *Arabidopsis*. This was followed by the lectin (80th), plastocyanin-like protein (106th), papain-like cysteine proteinase (116th), pectate lyase (144th), and cupin (305th) families. The Bet v 1 family was ranked 614th. Although no ranking was possible within the rice genome set, the only major differences in membership with *Arabidopsis* were the Lol p 1 family (69 instead of only 1 in *A thaliana*), cupin (88 compared with 11), and prolamin (140 compared with 87). These data demonstrate that there is no relationship between the distribution of protein families in genomes and that in plant food and respiratory allergens.

In silico analysis of cross-reactive allergen sequences and structures

Members of the Bet v 1 family and 1 of the prolamin superfamily subfamilies, the nsLTPs, are known to be involved in IgE cross-reactive allergy syndromes. It is thought that members of the Bet v 1 family found in foods are allergenic by virtue of previous sensitization to the birch pollen allergen Bet v 1 itself.¹⁶ Similarly cross-

reactive allergies to nsLTPs may arise from initial sensitization to the peach allergen, Pru p 3, although it is possible that a pollen homologue may also be involved.¹⁷ For both protein families, it is likely that the sensitizing agent is presented to the immune system in a native, folded form, because Bet v 1 sensitization occurs from pollen in the lungs, whereas the nsLTP is highly resistant to digestion¹⁸ and so is likely to be presented in its native form even in the gut. Consequently, it is probable that the resulting IgE will preferentially recognize the native folded proteins. We therefore developed a method to quantify differences in surface residues accessible for antibody binding change in native folded homologues, combining information from comparison of both the sequences and 3-dimensional structures of selected allergens.

Bet v 1 homologues

A comparison was initially made between the known 3-dimensional structures of Bet v 1 and its homologues by using Bet v 1a (Protein Data Bank codes 1BV1 and 1B6F), Bet v 1l (1FM4), Pru av 1 from cherry (1E09), and Llp10.1a and b from lupin (1ICX and 1IFV). This was done to assess the extent to which the surface topography changed across the Bet v 1 family, because this would

TABLE II. Sequence identity to Bet v 1a and estimated conserved surface for some of the Bet v 1 family*

Types of surface area (Å ²)	Bet v 1a† (sequence, P15494; structure, 1BV1)	Mal d 1b (Q9S7M5)	Gly m 4 (P26987)	Api g 1.0101 (P49372)
Total conserved surface area (determined by using 1.4-Å radius probe)	9370.8	6781.8	5690.6	4520.3
Total conserved surface area from side chains	7035.2	4368.2	3359.6	2211.8
Total conserved surface area from main chains	2336.5	2242.9	2205.7	2183.2
Surface area of tunnel (calculated from the steroid in Bet v 1 structure 1FM4)	959.0			
Conserved exterior surface area (based on the total surface area and exterior residues identified by the 3.5-Å radius probe)	8779.3 (100%)	6285.9 (71%)	5238.5 (60%)	4090.0 (46.5%)
Conserved exterior surface area from side chains	6509.3 (100%)	4109.5 (63%)	3099.3 (48%)	1973.3 (30%)
Conserved exterior surface area from main chains	2270.0 (100%)	2176.4 (96%)	2139.2 (96%)	2116.7 (93%)

*Residues absolutely conserved between the homologues and Bet v 1a (159 residues) were Mal d 1b, 90/159 (56%); Gly m 4, 76/158 (47%); and Api g 1.0101, 63/154 (39%).
†The following side chains from Bet v 1a were omitted from exterior surface area calculations: 22, 23, 38, 41, 52, 56, 58, 67, 69, 71, 74, 81, 83, 85, 98, 100, 102, 104, 114, 116, 118, 120, 122, 136, 143, 147, 151.

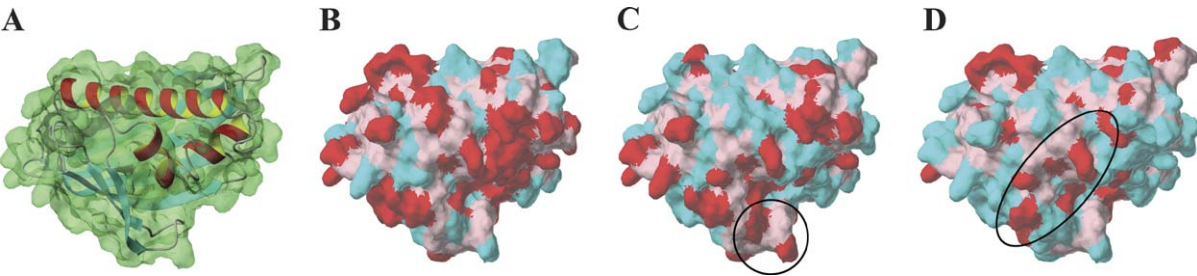


FIG 2. Conserved surface residues in the Bet v 1 family. **A**, Ribbon diagram of Bet v 1a. **B**, Conserved residues between Bet v 1a and Mal d 1 showing side chains (red), main chains (pink), and nonconserved side chains/main chains (cyan). **C**, Conserved residues, as in **B**, for Bet v 1a–Gly m 4 conservation, with P-loop marked. **D**, Conserved residues, as in **B**, for Bet v 1a–Api g 1.0101 conservation, with most similar surface region marked. Conception and production of video by John A. Jenkins, PhD (Institute of Food Research, Norwich, NR4 7UA, United Kingdom), Heimo Breiteneder, PhD (Department of Pathophysiology, Medical University of Vienna, Vienna, Austria), and E. N. Clare Mills, PhD (Institute of Food Research, Norwich, NR4 7UA, United Kingdom). This work was partly supported by EU through FAIR CT-98-4356 (PROTALL), QLK1-CT-2000-01394 (SAFE), and BBSRC through the competitive strategic grant to IFR and JACR. Dr Breiteneder acknowledges support by the Austrian Science Fund grant SFB F01802.

affect the validity of using this approach to compare the conservation of surface residues. This showed that the Bet v 1a structure, 1BV1, could be aligned with even the most distantly related structures, for Lpr10.1a (1ICX) and Lpr10.1b (1IFV) with root mean square deviations of 1.514 Å and 1.728 Å over the Cα atoms of 135 and 149 residues, respectively. These data indicated that the area of the surface contributed by conserved versus substituted residues in various Bet v 1 homologues could be determined.

The surface area of Bet v 1a was then measured by using a small (1.4-Å radius) probe, using an approach akin to rolling a ball of water across the surface of the protein. From this, the total surface area resulting from conserved residues and main chain (polypeptide backbone) was calculated (Table II). This included parts of the surface that might be hidden to an antibody but accessible to the 1.4-Å probe, such as the inside of the sterol-binding tunnel. Therefore, a second, larger (3.5-Å radius) probe, selected as being just too big to enter the tunnel, was rolled

across the surface to identify parts of the surface accessible to an antibody. In this way, the exterior surface area could be calculated, which excluded contributions from residues or main chains found in the sterol-binding tunnel. The 3.5-Å probe itself could not be used to measure the surface area accurately because it was too large.

Because IgE binding, like other antibody-binding reactions, will involve contacts between the antibody binding site and both the polypeptide backbone (comprising main chain atoms) and amino acid side chain residues of the allergen (cf, lysozyme),¹⁹ the conserved surface is expressed as a function of both (Table II). From this analysis, approximately 1000 Å² of total surface area was found to be contributed by the tunnel. The calculated exterior surface area for Bet v 1a was then approximately 8880 Å², of which approximately 75% arose from side chains. It is comparable with the surface area (approximately 959 Å²) of the 2 deoxycholate molecules found in the tunnel of Bet v 1.²⁰ Where the tunnel walls were thin, residues contributed to both the exterior surface and the

TABLE III. Sequence identity to peach nsLTP (Pru p 1, P81402, 91 residues) and estimated conserved surface for nsLTPs from maize and wheat*

Types of surface area (Å ²)	Corn nsLTP, Zea m 14 (sequence, P19656; structure, 1MZL)	Wheat nsLTP (sequence, P24296; structure, 1GH1 model 1)
Total surface area (determined by using 1.4-Å radius probe)	5206.1	5510.0
Total exterior surface area (based on residues contributing to exterior surface)	4857.2	5373.6
Conserved exterior surface area from side chains	1806.9 (55.5%)	1180.5 (39.0%)
Conserved exterior surface area from main chains	1375.0 (85.7%)	1500.8 (98.7%)

*Residues absolutely conserved between the homologues and Pru p 3 (91); Zea m 14, 57/93(63%); wheat 41/90 (46%).

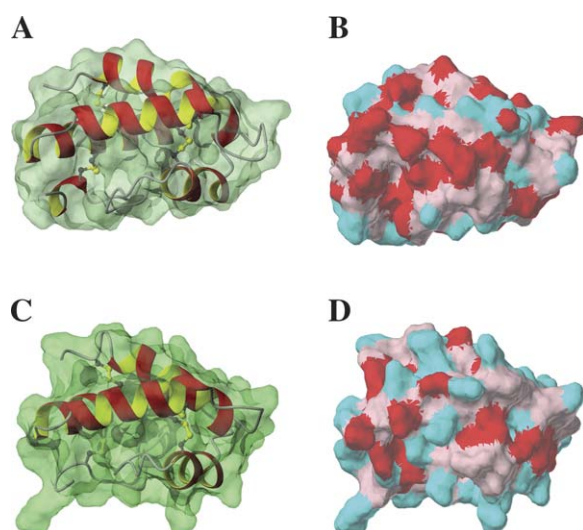


FIG 3. Conserved surface residues in the nsLTP family. **A**, Ribbon diagram of corn nsLTP, Zea m 14. **B**, Surface of Zea m 14, colored to show Pru p 3–Zea m 14 conservation, as in Fig 2. **C**, Ribbon diagram of wheat nsLTP1. **D**, Surface of wheat nsLTP, colored to show Pru p 3–wheat conservation, as in Fig 2. Conception and production of video by John A. Jenkins, PhD (Institute of Food Research, Norwich, NR4 7UA, United Kingdom), Heimo Breiteneder, PhD (Department of Pathophysiology, Medical University of Vienna, Vienna, Austria), and E. N. Clare Mills, PhD (Institute of Food Research, Norwich, NR4 7UA, United Kingdom). This work was partly supported by EU through FAIR CT-98-4356 (Protall), QLK1-C T-2000-01394 (SAFE), and BBSRC through the competitive strategic grant to IFR and JACR. Dr Breiteneder acknowledges support by the Austrian Science Fund grant SFB F01802.

tunnel surface. These were included in exterior surface area calculations and probably contributed to its being slightly overestimated. The sequence of Bet v 1a was then compared with homologous protein sequences from apple (Mal d 1b), soybean (Gly m 4), and celery (Api g 1.0101) by first aligning sequences (data not shown). This did not require more than a single internal deletion between pairs of sequences. In subsequent calculations, contributions to the conserved surface area from the main chains were omitted for the 3 residues around these deletions and the C-terminal extensions.

Identification of the exterior residues in the Bet v 1a structure allowed the conserved exterior surface residues to be identified in the different homologues. The compar-

ison among Bet v 1a, Mal d 1b, Gly m 4, and Api g 1.0101 (Table II; Fig 2) illustrates the extent to which the conserved surface areas form patches that might bind IgE. The most similar region of the surface between the most distant pair, Bet v 1a and Api g 1, is formed by the α -helix, loop, and β -sheet (Fig 2, D), with residues 25 to 45 containing a T-cell epitope identified by Bohle et al.²¹ By contrast, Bet v 1a and Gly m 4 have a similar surface on the P-loop (Fig 2, C). These data show that the level of conservation of surface in the Bet v 1 homologues is considerably higher than the level of sequence identity. Thus, in Mal d 1 and Gly m 4, approximately 71% and 60%, respectively, of the surface is conserved, compared with an overall sequence identity with Bet v 1 of 56% and 47%, respectively.

The overall levels of conserved external surface area relative to Bet v 1a are high, even in Api g 1, because the contribution from main chains is constant across all of the homologues (Table II). In contrast, the conserved external surface area contribution from side chains decreased with decreasing sequence identity relative to Bet v 1a. Thus, the sequence identity of Bet v 1a with Mal d 1 (56%) is lower than the conserved surface area (63%), whereas these values are similar for Gly m 4 (47% vs 48%). However, for Api g 1.0101, the conserved surface area from the side chains (30%) is lower than the sequence identity (39%). The deletion in Api g 1.0101 (also seen in Api g 1.0201) relative to Bet v 1a will contribute to this (Fig 2, D), and because it is located in the so-called P-loop around residue 45, where a major IgE epitope in Bet v 1 and Pru av 1 (originally known as Pru a 1) is located,^{22,23} it is likely to modify IgE binding to a considerable extent.

nsLTP homologues

A similar analysis of the nsLTP family of the prolamin superfamily was performed, comparing sequences from Pru p 3 from peach, Zea m 14 from corn, and wheat nsLTP1. The area of surface formed by conserved atoms is shown in Table III and Fig 3. These proteins were chosen because currently, it is thought that Pru p 3 is the sensitizing agent, with demonstrable cross-reactivity with Zea m 14²⁴ and none reported for wheat nsLTP1. As for the Bet v 1 homologues, several interior surface residues can be found in the lipid-binding tunnel and would not be expected to be accessible to IgE binding while the protein retained its native conformation. Residues in the loop

regions between the α -helices have both side chains, which contribute to the lipid binding cavity, and main chain atoms facing the exterior surface. Inclusion of these residues therefore results in overestimation of the calculated surface area. Once again, the peach and corn LTPs have a large proportion of their surface conserved, which could account for the observed IgE cross-reactivity. Although some of the surface of wheat nsLTP is conserved compared with peach, it is evident from the structures in Fig 3 that this protein has a very different surface topography. Because one of the surface loop regions is believed to be involved in IgE binding to peach nsLTP,²⁵ it seems unlikely that anti-Prü p 3 will show IgE cross-reactivity.

DISCUSSION

The data presented here indicate that, despite the number and diversity of proteins encoded in plant genomes, the universe of plant allergens is actually rather small, with approximately 65% of plant food allergens coming from only 4 protein families. This skewed distribution of different structural types of allergens has been inferred before, but this study represents the first quantitative comparison. It is also clear that the distribution of allergens into Pfam families does not merely reflect the relative abundance of the gene families in plants. Such information indicates that membership of one of these families is likely to be associated with a greater potential for a protein to become an allergen, providing additional factors such as presence in edible plant tissues are taken into account. Furthermore, comparisons of 3-dimensional structures suggest that the protein folds in the 3 dominating allergen families may predispose them to becoming allergens. Thus, for those allergen families such as the cupin and prolamin superfamilies in which sensitization via the gastrointestinal tract is thought to occur, the stability properties conferred by a fold may potentiate protein immunogenicity, a factor in determining allergenicity.^{4,5}

The remarkable conservation of both surface residues and main chain conformations in the Bet v 1 family plays an important role in conservation of IgE-binding epitopes²⁶ and underlies the fruit-vegetable-pollen cross-reactive allergy syndromes. The structural bioinformatic analysis undertaken here suggests that cross-reactive IgE binding can be retained even at approximately 35% to 40% of sequence identity, providing polypeptide backbone conformations are preserved. This extends to both members of the Bet v 1 family and the nsLTP subfamily of the prolamin superfamily. IgE cross-reactivity may also occur as a result of a single conserved region, such as an enzyme active site, despite a low overall level of sequence identity, as shown for the pectate lyase respiratory allergen of cedar pollen.²⁷

However, this combined conservation of sequence and main chain conformations does not occur across all subfamilies of the prolamin superfamily. This probably underlies the general lack of IgE cross-reactivity observed

both within and between the other subfamilies. Thus, although the 2S albumins have the same core of disulfide cross-linked α -helices as the nsLTPs, they contain hyper-variable loop regions that adopt a variety of conformations^{28,29} and are often the sites of IgE-binding.³⁰⁻³² The low level of sequence identity in these regions, combined with a divergent 3-dimensional structure, is probably responsible for the absence of IgE cross-reactivity between 2S albumins than might otherwise be expected from their overall level of sequence identity.

Despite having very low levels of sequence identity, members of the cupin family have highly conserved structures and functions. For example, there are no more than 5 identical contiguous amino acids between the 7S and 11S globulin sequences of soy. In contrast with the Bet v 1 family, there is little evidence of IgE cross-reactivity between cupin allergens with an overall sequence identity of $\leq 40\%$. This is probably because, in addition to changes in surface residues, the cupins do not maintain precisely the same main chain conformations across both plant species and between the 7 and 11S globulins. Thus, the surface seen by an IgE molecule changes much more for the cupins than for members of the Bet v 1 family, resulting in only very limited cross-reactivity between cupins from very closely related plant species such as legumes such as peanut and pea.³³ Furthermore, several cupin allergens are found only in foods consumed in a cooked form, making *in silico* analysis difficult because of protein denaturation induced by heating.

To summarize, combining structural information with analysis of conservation of primary structure indicates that the conservation of surface features (particularly main chain conformations that cannot be assessed by using simple amino acid homologies) would offer a sounder basis for assessing potential IgE cross-reactivity in novel proteins. A combined structural bioinformatic approach, such as the one we present, would contribute to the allergenic risk assessment process (FAO-WHO ALINORM 03/34, 2002; www.codexalimentarius.net/web/reports.jsp), particularly for proteins for which sensitization occurs with the native folded proteins. Implicit in the integrative approach adopted by FAO-WHO is that any such *in silico* analysis has to be considered along with other environmental factors, such as pollen exposure, dietary habits, the effect of food processing, and matrix, and individual factors, including genetic background, which may predispose individuals to becoming allergic.

We thank C. Radauer for critically reading the manuscript.

REFERENCES

- Sharma S, Lackie PM, Holgate ST. Uneasy breather: the implications of dust mite allergens. *Clin Exp Allergy* 2003;33:163-5.
- Taylor SL, Hefle SL. Genetically engineered foods: implications for food allergy. *Curr Opin Allergy Clin Immunol* 2002;2:249-52.
- Bateman A, Birney E, Cerruti L, Durbin R, Eddy SR, et al. The Pfam protein families database. *Nucleic Acids Res* 2002;30:276-80.
- Mills ENC, Jenkins J, Alcocer MJC, Shewry PR. Structural, biological and evolutionary relationships of plant food allergens sensitising via the gastrointestinal tract. *Crit Rev Food Sci Nutr* 2004;44:379-407.

5. Breiteneder H, Radauer C. A classification of plant food allergens. *J Allergy Clin Immunol* 2004;113:821-30.
6. Hileman RE, Silvanovich A, Goodman RE, Rice EA, Holleschak G, Astwood JD, et al. Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int Arch Allergy Immunol* 2002;128:280-91.
7. Soeria-Atmadja D, Zorzet A, Gustafsson MG, Hammerling U. Statistical evaluation of local alignment features predicting allergenicity using supervised classification algorithms. *Int Arch Allergy Immunol* 2004;133:101-12.
8. Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J* 2003;17:1141-3.
9. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003;31:365-70.
10. Pruess M, Fleischmann W, Kanapin A, Karavidopoulou Y, Kersey P, Kriventseva E, et al. The Proteome Analysis database: a tool for the in silico analysis of whole proteomes. *Nucleic Acids Res* 2003;31:414-7.
11. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 2002;296:79-92.
12. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics* 1996;14:51-5.
13. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205-17.
14. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen nomenclature. *Allergy* 1995;50:765-74.
15. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;408:796-815.
16. Vieths S, Scheurer S, Ballmer-Weber B. Current understanding of cross-reactivity of food allergens and pollen. *Ann N Y Acad Sci* 2002;964:47-68.
17. San Miguel-Moncin M, Krail M, Scheurer S, Enrique E, Alonso R, Conti A, et al. Lettuce anaphylaxis: identification of a lipid transfer protein as the major allergen. *Allergy* 2003;58:511-7.
18. Asero R, Mistrello G, Roncarolo D, de Vries SC, Gautier MF, Ciurana CL, et al. Lipid transfer protein: a pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. *Int Arch Allergy Immunol* 2000;122:20-32.
19. Li Y, Urrutia M, Smith-Gill SJ, Mariuzza RA. Dissection of binding interactions in the complex between the anti-lysozyme antibody HyHEL-63 and its antigen. *Biochemistry* 2003;42:11-22.
20. Markovic-Housley Z, Degano M, Lamba D, von Roepenack-Lahaye E, Clemens S, Susani M, et al. Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *J Mol Biol* 2003;325:123-33.
21. Bohle B, Radakovics A, Jahn-Schmid B, Hoffmann-Sommergruber K, Fischer GF, Ebner C. Bet v 1, the major birch pollen allergen, initiates sensitization to Api g 1, the major allergen in celery: evidence at the T cell level. *Eur J Immunol* 2003;33:3303-10.
22. Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, Joost van Neerven RJ, et al. X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nat Struct Biol* 1996;3:1040-5.
23. Scheurer S, Son DY, Boehm M, Karamloo F, Franke S, Hoffmann A, et al. Cross-reactivity and epitope analysis of Pru a 1, the major cherry allergen. *Mol Immunol* 1999;36:155-67.
24. Pastorello EA, Farioli L, Pravettoni V, Spano M, Scibola E, Trambaioli C, et al. The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. *J Allergy Clin Immunol* 2000;106:744-51.
25. Garcia-Casado G, Pacios LF, Diaz-Perales A, Sanchez-Monge R, Lombardero M, Garcia-Selles FJ, et al. Identification of IgE-binding epitopes of the major peach allergen Pru p 3. *J Allergy Clin Immunol* 2003;112:599-605.
26. Neudecker P, Lehmann K, Nerkamp J, Haase T, Wangorsch A, Fotisch K, et al. Mutational epitope analysis of Pru av 1 and Api g 1, the major allergens of cherry (*Prunus avium*) and celery (*Apium graveolens*): correlating IgE reactivity with three-dimensional structure. *Biochem J* 2003;376:97-107.
27. Midoro-Horiuti T, Mathura V, Schein CH, Braun W, Yu S, Watanabe M, et al. Major linear IgE epitopes of mountain cedar pollen allergen Jun a 1 map to the pectate lyase catalytic site. *Mol Immunol* 2003;40:555-62.
28. Pantoja-Uceda D, Bruix M, Gimenez-Gallego G, Rico M, Santoro J. Solution structure of RicC3, a 2S albumin storage protein from *Ricinus communis*. *Biochemistry* 2003;42:13839-47.
29. Pantoja-Uceda D, Shewry PR, Bruix M, Tatham AS, Santoro J, Rico M. Solution structure of a methionine-rich 2S albumin from sunflower seeds: relationship to its allergenic and emulsifying properties. *Biochemistry* 2004;43:6976-86.
30. Menendez-Arias L, Dominguez J, Moneo I, Rodriguez R. Epitope mapping of the major allergen from yellow mustard seeds, Sin a I. *Mol Immunol* 1990;27:143-50.
31. Monsalve RI, Gonzalez de la Pena MA, Menendez-Arias L, Lopez-Otin C, Villalba M, Rodriguez R. Characterization of a new oriental-mustard (*Brassica juncea*) allergen, Bra j IE: detection of an allergenic epitope. *Biochem J* 1993;293:625-32.
32. Robotham JM, Teuber SS, Sathe SK, Roux KH. Linear IgE epitope mapping of the English walnut (*Juglans regia*) major food allergen, Jug r 1. *J Allergy Clin Immunol* 2002;109:143-9.
33. Wensing M, Knulst AC, Piersma S, O'Kane F, Knol EF, Koppelman SJ. Patients with anaphylaxis to pea can have peanut allergy caused by cross-reactive IgE to vicilin (Ara h 1). *J Allergy Clin Immunol* 2003;111:420-4.

This user page is hosted by the [Sanger Institute](#) - [Disclaimer](#)

Plant food allergen families

The data here is supplementary information for the submitted paper:

**Structural relatedness of plant food allergens with specific reference to cross-reactive allergens:
An in silico analysis.**

John A. Jenkins, Sam Griffiths-Jones, Peter R. Shewry, Heimo Breiteneder, and E.N. Clare Mills

Lists of Swissprot identifiers used in the analysis

- [plant food set](#)
- [plant aero set](#)

[Complete list of Pfam 7.3 hits for the above proteins](#)

File format is:

protein_id	seq_start	seq_end	pfam_acc	model_start	model_end
bit_score	evaluate	pfam_id			

Please contact sgj@sanger.ac.uk for queries regarding this data.



► [Information, Teams, Faculty & People](#)

[News](#), [Introduction](#), [YourGenome](#),
[Press](#), [Seminars](#), [Travel](#), [Library](#)

► [Ensembl Genome Data Resources](#)

[Human](#), [Mouse](#), [Rat](#), [Fugu](#), [Zebrafish](#),
[Mosquito](#), [Fly](#), [Trace Repository](#)

► [Software, Blast & Databases](#)

[Acedb](#), [Artemis](#), [GeneDB](#), [Pfam](#), [Rfam](#), [SRS](#),
[Human BLAST](#), [C.elegans BLAST](#)

► [Genome Sequencing & Annotation](#)

[Vega](#), [Human](#), [Mouse](#), [Zebrafish](#), [X.tropicalis](#),
[Worm](#), [Pathogens](#), [S.pombe](#)

► [Genomics & Genetics](#)

[Cancer Genome Project](#), [Human Genetics](#),
[Gene Trap](#), [Mouse](#), [S.pombe](#)

► [Genome Infrastructure](#)

[Gene Resources](#), [Microarrays](#), [Genotyping](#),
[Bioinformatics](#), [Clone Requests](#)

► [Careers At The Sanger Institute](#)

[Graduate](#), [Postdoctoral](#) & [Faculty](#),
[Advertised Jobs](#), [Student Placements](#)

Finished Sequence: 2,521,447,355 bases

(05-Jan-2005, 06:52 PM)

[\[Detailed Statistics\]](#)



Search

☐ [www](#) ☐ [www.sanger.ac.uk](#)
[Contact us](#) [Sitemap](#)

20th Dec 2004

[More](#)

[2005 Sanger Institute Prize Competition](#)

Call for Expressions of Interest - The winner of the Sanger Institute Prize will be offered an all expenses paid ...



20th Dec 2004

[More](#)

[MEROPS Protease Database 6.90 released](#)

The MEROPS group at the Sanger Institute are pleased to announce the release of the MEROPS Protease Database version 6.90. ...



17th Dec 2004

[More](#)

[COSMIC Update](#)

The COSMIC team is proud to release somatic mutation data for CSF1R, RB1, RET and SMARCB1. This information has been ...



Database Versions

WormBase 136	4th Jan 2005
MEROPS 6.90	20th Dec 2004
COSMIC	17th Dec 2004
Vega	16th Dec 2004
Ensembl 27	15th Dec 2004
Glovar 2	17th Nov 2004
Pfam 16.0	5th Nov 2004
Rfam 6.1	1st Aug 2004



Sanger Institute User Page Content

The contents of the Sanger Institute user's pages are left largely to the discretion of the author, provided it falls within the guidelines set out in our Acceptable Use Policy below. Home pages provide the means to distribute directly information that does not form part of the official Sanger web site.

We make every effort to ensure that only acceptable content is available on Sanger Institute hosted web pages. If you feel that any content is offensive or else breaches our guidelines then please send email to webmaster@sanger.ac.uk.

Sanger Institute User Web Page Acceptable Use Policy (AUP)

1. Sanger Institute staff are responsible for the content of their own user pages, including obtaining the legal permission for any works they include and ensuring that the contents of these pages do not violate English law. They will be held responsible for any defamatory, confidential, secret or other proprietary material available via their user pages.
2. Sanger Institute user pages must not contain any offensive material.
3. Sanger Institute user pages must not be used to distribute or advertise any of the following:
 - software for sending SPAM (bulk emails, excessive news postings, etc.).
 - lists of email addresses except where all the addressees have given their explicit permission.
 - any collection of personal data other than in accordance with the [Data Protection Act](#) and the [Data Protection Principles](#).

Last content update: July 1998

2SS8_HELAN
2SS_BEREX
2SS_RICCO
ACTN_ACTCH
AL11_ARAHY
AL12_ARAHY
ALL1_APIGR
ALL1_BRAJU
ALL1_SINAL
ALL2_APIGR
C13_MAIZE
CB23_APIGR
DAU1_DAUCA
GLC1_SOYBN
GLC2_SOYBN
GLC3_SOYBN
GLC4_SOYBN
GLC5_SOYBN
GLCA_SOYBN
GLCB_SOYBN
GLCX_SOYBN
GLU2_MAIZE
IA03_WHEAT
IA16_WHEAT
IAA1_HORVU
IAAB_HORVU
IAP2_SOLTU
ICTD_SOLTU
ITRA_SOYBN
LA52_LYCES
LECG_ARAHY
LEC_SOYBN
LOX1_SOYBN
LOX2_SOYBN
MAL1_MALDO
MPO1_ORYSA
MPZ1_MAIZE
NLT1_PRUAR
NLT1_PRUDO
NLT1_PRUPE
NLTP_MAIZE
NLTP_MALDO
NLTP_PRUAV
O22120
O22121

O22318
O22521
O49178
O64458
O82546
O82580
P81943
P93198
P93680
PAPA_CARPA
PAT1_SOLTU
PC83_SOLTU
PR1_PHAVU
PR2_PHAVU
PRO1_LYCES
PRO1_MALDO
PRO1_SOYBN
PRO2_MALDO
PRO2_SOYBN
PRO3_MALDO
PROF_APIGR
PROF_ARAHY
PROF_HELAN
PROF_PRUAV
PROF_PYRCO
PRU1_PRUAR
PRU1_PRUAV
PYR1_PYRCO
Q39802
Q39853
Q39869
Q39898
Q39899
Q40280
Q40652
Q40653
Q40654
Q40655
Q40718
Q40719
Q40720
Q41043
Q41196
Q41277
Q41278
Q41279

Q41280
Q41576
Q42428
Q43549
Q43550
Q43551
Q43552
Q43861
Q43862
Q9AUD1
Q9AVK8
Q9FPK2
Q9FPK3
Q9FPK4
Q9FZP9
Q9LD01
Q9LD05
Q9LD06
Q9S7M5
Q9S8F2
Q9S8F3
Q9S8H1
Q9S8H2
Q9S8Q4
Q9SEW4
Q9SQG5
Q9SQH1
Q9SQH7
Q9SWR4
Q9SYV3
Q9SYV4
Q9SYW3
Q9ZR84
Q9ZWJ2
RA05_ORYSA
RA14_ORYSA
RA17_ORYSA
RAG1_ORYSA
RAG2_ORYSA
RAP_TAROF
TLP_PRUAV
YIR4_YEAST
YN62_ARCFU
ZEA3_MAIZE

ALL1_OLEEU
ALL3_BETVE
ALL6_OLEEU
BV1B_BETVE
BV1C_BETVE
BV1E_BETVE
BV1F_BETVE
BV1G_BETVE
BV1J_BETVE
BV1K_BETVE
BV1L_BETVE
CAA07321
CAA76557
CHAL_BRARA
EXB2_ARATH
LIV1_LIGVU
MP11_AMBAR
MP12_AMBAR
MP13_AMBAR
MP14_AMBAR
MP51_PHAAQ
MP52_PHAAQ
MP53_PHAAQ
MP54_PHAAQ
MP5A_AMBPS
MP5A_LOLPR
MP5A_PHLPR
MP5B_AMBPS
MP5B_LOLPR
MP5B_PHLPR
MP91_POAPR
MP92_POAPR
MP93_POAPR
MPA1_CARBE
MPA1_CHAOB
MPA1_CUPAR
MPA1_JUNAS
MPA1_PHAAQ
MPA2_AMBAR
MPA2_CARBE
MPA2_CRYJA
MPA3_AMBEL
MPA5_AMBEL
MPAA_CORAV
MPAG_ALNGL

MPC1_CYNDA
MPG2_DACGL
MPG3_DACGL
MPH1_HOLLA
MPL1_LOLPR
MPL2_LOLPR
MPL3_LOLPR
MPP1_PHLPR
MPP2_PHLPR
MPT5_AMBTR
NL11_PARJU
NL13_PARJU
NL21_PARJU
NLT6_AMBAR
O04828
O23207
O23746
O23747
O23748
O23749
O23750
O23751
O23752
O23753
O23754
O23971
O24642
O65013
O65318
O65319
O65320
O65321
O65868
O81341
O81342
O81343
O81344
O81355
P93105
P93466
P93467
POC1_BRARA
POC2_BRARA
POC2_JUNOX
POC3_OLEEU
POC4_ALNGL

POC4_BETVE
POC7_CYNDA
PRO1_ARATH
PRO1_OLEEU
PRO1_PHLPR
PRO2_OLEEU
PRO3_OLEEU
PROF_BETVE
PROF_CYNDA
PROF_MERAN
PROF_PYRCO
PRR3_JUNAS
PYR1_PYRCO
Q39425
Q39426
Q39427
Q39428
Q39429
Q39430
Q39431
Q39453
Q39454
Q39625
Q39626
Q40239
Q40967
Q42428
Q947S5
Q96365
Q96366
Q96367
Q96368
Q96370
Q96371
Q96377
Q96378
Q96379
Q96380
Q96381
Q96382
Q96501
Q96503
Q9FUW6
Q9FY19
Q9LD79
Q9LLT1

Q9LLT2
Q9M4S2
Q9M4S3
Q9M4S4
Q9M4S5
Q9M4S6
Q9S8S8
Q9SBE0
Q9SC98
Q9XG86
Q9ZP03
Q9ZP13
Q9ZP13
SBP_CRYJA
SODC_OLEEU

2SS8_HELAN	49	139	PF00234	1	129	45.9	5.7e-11	tryp_alpha_amyl
2SS_BEREX	53	137	PF00234	1	129	61.1	1.6e-15	tryp_alpha_amyl
2SS_RICCO	61	155	PF00234	1	129	56.9	2.9e-14	tryp_alpha_amyl
2SS_RICCO	162	256	PF00234	1	129	61.4	1.3e-15	tryp_alpha_amyl
ACTN_ACTCH	127	342	PF00112	1	337	387.6	8.1e-114	Peptidase_C1
AL11_ARAHY	176	303	PF00190	1	182	84.0	2e-22	Cupin
AL11_ARAHY	390	566	PF00190	1	182	141.9	7.6e-40	Cupin
AL12_ARAHY	182	309	PF00190	1	182	83.3	3.3e-22	Cupin
AL12_ARAHY	398	571	PF00190	1	182	140.2	2.4e-39	Cupin
ALL1_APIGR	1	154	PF00407	1	161	240.9	1.2e-69	Bet_v_I
ALL1_BRAJU	22	119	PF00234	1	129	48.8	8.1e-12	tryp_alpha_amyl
ALL1_OLEEU	13	133	PF01190	1	147	250.5	1.5e-72	Pollen_Ole_e_I
ALL1_SINAL	23	117	PF00234	1	129	54.4	1.6e-13	tryp_alpha_amyl
ALL2_APIGR	1	158	PF00407	1	161	271.6	6.6e-79	Bet_v_I
ALL3_BETVE	40	68	PF00036	1	29	33.2	3.8e-07	efhand
ALL3_BETVE	134	162	PF00036	1	29	22.8	0.00054	efhand
ALL3_BETVE	172	200	PF00036	1	29	29.6	4.6e-06	efhand
BV1B_BETVE	1	158	PF00407	1	161	375.4	3.7e-110	Bet_v_I
BV1C_BETVE	1	158	PF00407	1	161	370.2	1.4e-108	Bet_v_I
BV1E_BETVE	1	158	PF00407	1	161	371.8	4.5e-109	Bet_v_I
BV1F_BETVE	1	158	PF00407	1	161	373.0	2e-109	Bet_v_I
BV1G_BETVE	1	158	PF00407	1	161	371.5	5.6e-109	Bet_v_I
BV1J_BETVE	1	158	PF00407	1	161	374.7	6e-110	Bet_v_I
BV1K_BETVE	1	158	PF00407	1	161	374.0	9.7e-110	Bet_v_I
BV1L_BETVE	1	158	PF00407	1	161	369.0	3.1e-108	Bet_v_I
C13_MAIZE	37	157	PF01190	1	147	254.8	7.6e-74	Pollen_Ole_e_I
CAA07321	1	159	PF00407	1	161	395.1	4.5e-116	Bet_v_I
CAA76557	7	138	PF01620	1	195	259.5	3e-75	Pollen_allerg_2
CB23_APIGR	65	232	PF00504	1	192	300.6	1.2e-87	chloroa_b-bind
CHAL_BRARA	1	36	PF00187	1	40	40.9	1.8e-09	Chitin_bind_1
CHAL_BRARA	28	91	PF00967	1	128	35.9	4.6e-08	Barwin
DAU1_DAUCA	1	154	PF00407	1	161	234.2	1.2e-67	Bet_v_I
EXB2_ARATH	175	258	PF01357	1	84	142.8	4e-40	Pollen_allergen
GLC1_SOYBN	36	190	PF00190	1	182	209.7	2.9e-60	Cupin
GLC1_SOYBN	323	472	PF00190	1	182	193.8	1.8e-55	Cupin
GLC2_SOYBN	33	187	PF00190	1	182	214.0	1.4e-61	Cupin
GLC2_SOYBN	313	462	PF00190	1	182	176.6	2.6e-50	Cupin
GLC3_SOYBN	36	187	PF00190	1	182	212.1	5.6e-61	Cupin
GLC3_SOYBN	309	458	PF00190	1	182	191.6	7.9e-55	Cupin
GLC4_SOYBN	37	195	PF00190	1	182	188.0	9.6e-54	Cupin
GLC4_SOYBN	390	536	PF00190	1	182	179.5	3.6e-51	Cupin
GLC5_SOYBN	38	195	PF00190	1	182	181.7	7.8e-52	Cupin
GLC5_SOYBN	357	502	PF00190	1	182	135.6	5.8e-38	Cupin
GLCA_SOYBN	203	330	PF00190	1	182	90.5	2.1e-24	Cupin
GLCA_SOYBN	406	567	PF00190	1	182	141.4	1e-39	Cupin

GLCB_SOYBN	42	169	PF00190	1	182	87.3	2e-23	Cupin
GLCB_SOYBN	240	401	PF00190	1	182	135.8	5.1e-38	Cupin
GLCX_SOYBN	219	365	PF00190	1	182	74.9	1.1e-19	Cupin
GLCX_SOYBN	441	601	PF00190	1	182	140.7	1.7e-39	Cupin
GLU2_MAIZE	128	213	PF00234	1	129	53.0	4.4e-13	tryp_alpha_amyl
IA03_WHEAT	43	159	PF00234	1	129	63.2	3.7e-16	tryp_alpha_amyl
IA16_WHEAT	30	134	PF00234	1	129	80.0	3.1e-21	tryp_alpha_amyl
IAA1_HORVU	20	131	PF00234	1	129	78.7	8e-21	tryp_alpha_amyl
IAAB_HORVU	30	134	PF00234	1	129	76.7	3.2e-20	tryp_alpha_amyl
IAP2_SOLTU	35	215	PF00197	1	200	218.4	7e-63	Kunitz_legume
ICTD_SOLTU	7	187	PF00197	1	200	266.9	1.8e-77	Kunitz_legume
ITRA_SOYBN	26	199	PF00197	1	200	316.4	2.2e-92	Kunitz_legume
LA52_LYCES	29	149	PF01190	1	147	247.8	9.9e-72	Pollen_Ole_e_I
LECG_ARAHY	22	205	PF00139	1	207	352.0	4.2e-103	lectin_legB
LECG_ARAHY	210	257	PF00138	1	50	76.5	3.6e-20	lectin_legA
LEC_SOYBN	31	215	PF00139	1	207	430.9	7.5e-127	lectin_legB
LEC_SOYBN	221	269	PF00138	1	50	99.3	5e-27	lectin_legA
LIV1_LIGVU	13	133	PF01190	1	147	226.3	2.9e-65	Pollen_Ole_e_I
LOX1_SOYBN	40	143	PF01477	1	129	138.3	8.9e-39	PLAT
LOX1_SOYBN	155	823	PF00305	1	678	1663.7	0	lipoxygenase
LOX2_SOYBN	70	173	PF01477	1	129	141.8	8e-40	PLAT
LOX2_SOYBN	185	850	PF00305	1	678	1688.6	0	lipoxygenase
MAL1_MALDO	1	157	PF00407	1	161	284.3	1e-82	Bet_v_I
MP11_AMBAR	38	389	PF00544	1	403	276.9	1.6e-80	pec_lyase
MP12_AMBAR	40	391	PF00544	1	403	325.7	3.5e-95	pec_lyase
MP13_AMBAR	40	390	PF00544	1	403	306.4	2.3e-89	pec_lyase
MP14_AMBAR	40	385	PF00544	1	403	228.6	6e-66	pec_lyase
MP51_PHAAQ	7	147	PF01620	1	195	277.1	1.5e-80	Pollen_allerg_2
MP52_PHAAQ	7	157	PF01620	1	195	289.8	2.3e-84	Pollen_allerg_2
MP53_PHAAQ	7	142	PF01620	1	195	272.6	3.4e-79	Pollen_allerg_2
MP54_PHAAQ	1	30	PF01620	166	195	70.7	2e-19	Pollen_allerg_2
MP5A_AMBPS	23	66	PF03913	1	44	114.4	1.4e-31	Amb_V_allergen
MP5A_LOLPR	7	159	PF01620	1	195	23.1	2.8e-07	Pollen_allerg_2
MP5A_PHLPR	2	150	PF01620	1	195	251.2	9.1e-73	Pollen_allerg_2
MP5B_AMBPS	22	66	PF03913	1	44	76.6	3.4e-20	Amb_V_allergen
MP5B_LOLPR	7	177	PF01620	1	195	359.4	2.5e-105	Pollen_allerg_2
MP5B_PHLPR	2	142	PF01620	1	195	195.2	6.5e-56	Pollen_allerg_2
MP91_POAPR	42	237	PF01620	1	195	401.4	5.7e-118	Pollen_allerg_2
MP92_POAPR	7	178	PF01620	1	195	318.5	4.9e-93	Pollen_allerg_2
MP93_POAPR	7	171	PF01620	1	195	290.0	1.9e-84	Pollen_allerg_2
MPA1_CARBE	1	158	PF00407	1	161	365.0	5e-107	Bet_v_I
MPA1_CHAOB	39	367	PF00544	1	403	352.3	3.5e-103	pec_lyase
MPA1_CUPAR	18	346	PF00544	1	403	352.1	3.8e-103	pec_lyase
MPA1_JUNAS	39	367	PF00544	1	403	343.9	1.1e-100	pec_lyase
MPA1_PHAAQ	177	257	PF01357	1	84	136.7	2.8e-38	Pollen_allergen
MPA2_AMBAR	39	390	PF00544	1	403	314.7	7.2e-92	pec_lyase

MPA2_CARBE	1	158	PF00407	1	161	359.0	3.3e-105	Bet_v_I
MPA2_CRYJA	86	428	PF00295	1	362	384.4	7.2e-113	Glyco_hydro_28
MPA3_AMBEL	14	92	PF02298	1	89	104.9	1e-28	Cu_bind_like
MPA5_AMBEL	1	44	PF03913	1	44	120.0	2.9e-33	Amb_V_allergen
MPAA_CORAV	1	158	PF00407	1	161	353.4	1.5e-103	Bet_v_I
MPAG_ALNGL	1	158	PF00407	1	161	341.9	4.6e-100	Bet_v_I
MPC1_CYNDA	148	229	PF01357	1	84	142.9	3.8e-40	Pollen_allergen
MPG2_DACGL	4	36	PF01357	1	33	25.4	9.4e-07	Pollen_allergen
MPG3_DACGL	3	83	PF01357	1	84	84.6	1.3e-22	Pollen_allergen
MPH1_HOLLA	173	253	PF01357	1	84	130.5	1.9e-36	Pollen_allergen
MPL1_LOLPR	171	251	PF01357	1	84	137.1	2.1e-38	Pollen_allergen
MPL2_LOLPR	4	85	PF01357	1	84	105.0	9.4e-29	Pollen_allergen
MPL3_LOLPR	3	83	PF01357	1	84	72.3	6.6e-19	Pollen_allergen
MPO1_ORYSA	167	248	PF01357	1	84	151.5	9.3e-43	Pollen_allergen
MPP1_PHLPR	170	251	PF01357	1	84	145.8	5.1e-41	Pollen_allergen
MPP2_PHLPR	30	109	PF01357	1	84	106.5	3.3e-29	Pollen_allergen
MPT5_AMBTR	34	73	PF03913	1	44	98.5	8.5e-27	Amb_V_allergen
MPZ1_MAIZE	94	175	PF01357	1	84	140.9	1.5e-39	Pollen_allergen
NL11_PARJU	4	91	PF00234	1	129	79.3	5.3e-21	tryp_alpha_amyl
NL13_PARJU	40	127	PF00234	1	129	81.2	1.4e-21	tryp_alpha_amyl
NL21_PARJU	35	121	PF00234	1	129	76.5	3.5e-20	tryp_alpha_amyl
NLT1_PRUAR	3	87	PF00234	1	129	119.4	4.3e-33	tryp_alpha_amyl
NLT1_PRUDO	3	87	PF00234	1	129	117.2	2.1e-32	tryp_alpha_amyl
NLT1_PRUPE	3	87	PF00234	1	129	116.2	4.2e-32	tryp_alpha_amyl
NLT6_AMBAR	30	113	PF00234	1	129	70.5	2.4e-18	tryp_alpha_amyl
NLTP_MAIZE	31	116	PF00234	1	129	118.4	8.9e-33	tryp_alpha_amyl
NLTP_MALDO	27	111	PF00234	1	129	121.4	1.1e-33	tryp_alpha_amyl
NLTP_PRUAV	29	113	PF00234	1	129	116.8	2.7e-32	tryp_alpha_amyl
O04828	8	182	PF01620	1	195	374.2	8.5e-110	Pollen_allerg_2
O22120	141	268	PF00190	1	182	90.5	2.1e-24	Cupin
O22120	344	505	PF00190	1	182	141.4	1e-39	Cupin
O22121	19	146	PF00190	1	182	87.5	1.7e-23	Cupin
O22121	217	378	PF00190	1	182	135.8	5.1e-38	Cupin
O22318	13	52	PF00187	1	40	75.0	1.1e-19	Chitin_bind_1
O22318	71	122	PF00182	1	52	108.2	7.2e-31	Glyco_hydro_19
O22521	1	159	PF00407	1	161	295.3	5e-86	Bet_v_I
O23207	74	217	PF00258	1	173	97.2	2.1e-26	flavodoxin
O23746	1	158	PF00407	1	161	373.6	1.3e-109	Bet_v_I
O23747	1	158	PF00407	1	161	383.6	1.3e-112	Bet_v_I
O23748	1	158	PF00407	1	161	379.5	2.2e-111	Bet_v_I
O23749	1	158	PF00407	1	161	383.9	1.1e-112	Bet_v_I
O23750	1	158	PF00407	1	161	357.0	1.3e-104	Bet_v_I
O23751	1	159	PF00407	1	161	386.4	1.9e-113	Bet_v_I
O23752	1	159	PF00407	1	161	392.8	2.1e-115	Bet_v_I
O23753	1	159	PF00407	1	161	390.6	9.7e-115	Bet_v_I
O23754	1	159	PF00407	1	161	389.1	2.8e-114	Bet_v_I

O23971	1	106	PF01620	1	195	91.4	1.2e-24	Pollen_allerg_2
O24642	1	159	PF00407	1	161	394.6	6.1e-116	Bet_v_I
O49178	42	149	PF00234	1	129	38.0	1.4e-08	tryp_alpha_amyl
O64458	135	356	PF00112	1	337	350.3	1.3e-102	Peptidase_C1
O65013	164	257	PF01357	1	84	33.1	4.2e-07	Pollen_allergen
O65318	2	140	PF01620	1	195	235.1	6.3e-68	Pollen_allerg_2
O65319	2	140	PF01620	1	195	241.9	5.7e-70	Pollen_allerg_2
O65320	2	140	PF01620	1	195	241.9	5.7e-70	Pollen_allerg_2
O65321	2	140	PF01620	1	195	241.9	5.7e-70	Pollen_allerg_2
O65868	7	138	PF01620	1	195	260.2	1.8e-75	Pollen_allerg_2
O81341	7	176	PF01620	1	195	329.4	2.7e-96	Pollen_allerg_2
O81342	13	153	PF01620	1	195	193.9	1.6e-55	Pollen_allerg_2
O81343	7	148	PF01620	1	195	241.2	9.4e-70	Pollen_allerg_2
O81344	7	148	PF01620	1	195	238.3	7.2e-69	Pollen_allerg_2
O81355	2	308	PF02716	1	322	824.7	2.1e-245	Isoflavone_redu
O82546	30	245	PF00314	1	207	284.1	1.2e-82	thaumatin
O82580	14	181	PF00190	1	182	195.7	4.8e-56	Cupin
O82580	335	484	PF00190	1	182	173.4	2.4e-49	Cupin
P93105	1	140	PF00407	1	161	156.0	4.3e-44	Bet_v_I
P93198	52	132	PF00234	1	129	59.4	5.2e-15	tryp_alpha_amyl
P93466	1	139	PF01620	1	195	188.1	8.9e-54	Pollen_allerg_2
P93467	2	140	PF01620	1	195	233.2	2.5e-67	Pollen_allerg_2
P93680	25	64	PF00187	1	40	79.5	4.6e-21	Chitin_bind_1
P93680	78	315	PF00182	1	239	640.6	5.7e-190	Glyco_hydro_19
PAPA_CARPA	134	343	PF00112	1	337	337.8	8.2e-99	Peptidase_C1
PAT1_SOLTU	23	219	PF01734	1	268	218.8	5.1e-63	Patatin
PC83_SOLTU	45	219	PF00197	1	200	232.9	2.9e-67	Kunitz_legume
POC1_BRARA	5	33	PF00036	1	29	39.4	5.3e-09	efhand
POC1_BRARA	40	68	PF00036	1	29	31.4	1.3e-06	efhand
POC2_BRARA	9	37	PF00036	1	29	37.4	2.1e-08	efhand
POC2_BRARA	44	72	PF00036	1	29	38.0	1.4e-08	efhand
POC2_JUNOX	26	54	PF00036	1	29	39.4	5.2e-09	efhand
POC2_JUNOX	62	90	PF00036	1	29	33.0	4.6e-07	efhand
POC2_JUNOX	95	123	PF00036	1	29	30.2	3.2e-06	efhand
POC2_JUNOX	131	159	PF00036	1	29	33.1	4.2e-07	efhand
POC3_OLEEU	10	38	PF00036	1	29	32.6	6e-07	efhand
POC3_OLEEU	45	73	PF00036	1	29	39.2	6.1e-09	efhand
POC4_ALNGL	11	39	PF00036	1	29	34.8	1.3e-07	efhand
POC4_ALNGL	46	74	PF00036	1	29	38.3	1.2e-08	efhand
POC4_BETVE	11	39	PF00036	1	29	35.0	1.1e-07	efhand
POC4_BETVE	46	74	PF00036	1	29	35.9	6.1e-08	efhand
POC7_CYNDA	6	34	PF00036	1	29	36.7	3.4e-08	efhand
POC7_CYNDA	41	69	PF00036	1	29	39.9	3.8e-09	efhand
PR1_PHAVU	1	155	PF00407	1	161	340.0	1.7e-99	Bet_v_I
PR2_PHAVU	1	155	PF00407	1	161	335.0	5.5e-98	Bet_v_I
PRO1_ARATH	1	125	PF00235	1	139	210.2	2e-60	profilin

PRO1_LYCES	2	128	PF00235	1	139	234.9	7.5e-68	profilin
PRO1_MALDO	2	126	PF00235	1	139	242.4	4e-70	profilin
PRO1_OLEEU	2	129	PF00235	1	139	249.6	2.8e-72	profilin
PRO1_PHLPR	2	126	PF00235	1	139	242.1	5.1e-70	profilin
PRO1_SOYBN	2	126	PF00235	1	139	218.8	5.4e-63	profilin
PRO2_MALDO	2	126	PF00235	1	139	219.2	4e-63	profilin
PRO2_OLEEU	2	129	PF00235	1	139	249.7	2.5e-72	profilin
PRO2_SOYBN	2	126	PF00235	1	139	214.2	1.3e-61	profilin
PRO3_MALDO	2	126	PF00235	1	139	227.4	1.3e-65	profilin
PRO3_OLEEU	2	129	PF00235	1	139	247.3	1.4e-71	profilin
PROF_APIGR	2	129	PF00235	1	139	227.3	1.5e-65	profilin
PROF_ARAHY	2	126	PF00235	1	139	214.9	7.6e-62	profilin
PROF_BETVE	2	128	PF00235	1	139	250.0	2.1e-72	profilin
PROF_CYNDA	2	126	PF00235	1	139	237.1	1.7e-68	profilin
PROF_HELAN	2	128	PF00235	1	139	228.0	8.7e-66	profilin
PROF_MERAN	2	128	PF00235	1	139	244.1	1.3e-70	profilin
PROF_PRUAV	2	126	PF00235	1	139	224.2	1.3e-64	profilin
PROF_PYRCO	2	126	PF00235	1	139	244.4	1e-70	profilin
PRR3_JUNAS	33	225	PF00314	1	207	384.5	7.1e-113	thaumatin
PRU1_PRUAR	1	159	PF00407	1	161	330.1	1.6e-96	Bet_v_I
PRU1_PRUAV	1	159	PF00407	1	161	317.4	1.1e-92	Bet_v_I
PYR1_PYRCO	1	158	PF00407	1	161	289.4	3e-84	Bet_v_I
Q39425	1	159	PF00407	1	161	348.1	6.4e-102	Bet_v_I
Q39426	1	159	PF00407	1	161	389.5	2.1e-114	Bet_v_I
Q39427	1	159	PF00407	1	161	391.8	4.4e-115	Bet_v_I
Q39428	1	159	PF00407	1	161	388.3	4.8e-114	Bet_v_I
Q39429	1	159	PF00407	1	161	392.2	3.4e-115	Bet_v_I
Q39430	1	159	PF00407	1	161	355.6	3.4e-104	Bet_v_I
Q39431	1	159	PF00407	1	161	378.9	3.4e-111	Bet_v_I
Q39453	1	159	PF00407	1	161	368.1	5.8e-108	Bet_v_I
Q39454	1	159	PF00407	1	161	348.7	4e-102	Bet_v_I
Q39625	159	236	PF01357	1	84	176.2	3.5e-50	Pollen_allergen
Q39626	164	242	PF01357	1	84	174.4	1.2e-49	Pollen_allergen
Q39802	177	260	PF01357	1	84	142.3	5.7e-40	Pollen_allergen
Q39853	19	180	PF00190	1	182	137.9	1.2e-38	Cupin
Q39869	27	199	PF00197	1	200	152.0	6.9e-43	Kunitz_legume
Q39898	27	200	PF00197	1	200	316.4	2.2e-92	Kunitz_legume
Q39899	27	200	PF00197	1	200	299.0	3.7e-87	Kunitz_legume
Q40239	1	81	PF01357	1	84	94.2	1.7e-25	Pollen_allergen
Q40280	1	157	PF00407	1	161	285.5	4.4e-83	Bet_v_I
Q40652	41	153	PF00234	1	129	93.4	2.9e-25	tryp_alpha_amyl
Q40653	34	146	PF00234	1	129	95.0	9.7e-26	tryp_alpha_amyl
Q40654	41	149	PF00234	1	129	91.2	1.3e-24	tryp_alpha_amyl
Q40655	42	152	PF00234	1	129	96.2	4.2e-26	tryp_alpha_amyl
Q40718	8	98	PF00234	1	129	65.9	5.6e-17	tryp_alpha_amyl
Q40719	2	96	PF00234	1	129	67.0	2.6e-17	tryp_alpha_amyl

Q40720	10	100	PF00234	1	129	65.2	9.3e-17	tryp_alpha_amyl
Q40967	170	251	PF01357	1	84	142.6	4.5e-40	Pollen_allergen
Q41043	165	242	PF01357	1	84	160.9	1.4e-45	Pollen_allergen
Q41196	23	135	PF00234	1	129	67.1	2.4e-17	tryp_alpha_amyl
Q41277	23	135	PF00234	1	129	66.4	3.8e-17	tryp_alpha_amyl
Q41278	23	135	PF00234	1	129	68.5	9.1e-18	tryp_alpha_amyl
Q41279	23	135	PF00234	1	129	65.4	7.9e-17	tryp_alpha_amyl
Q41280	23	135	PF00234	1	129	67.8	1.5e-17	tryp_alpha_amyl
Q41576	25	105	PF01357	1	84	69.5	4.7e-18	Pollen_allergen
Q42428	18	57	PF00187	1	40	77.3	2e-20	Chitin_bind_1
Q42428	80	315	PF00182	1	239	605.6	2e-179	Glyco_hydro_19
Q43549	1	158	PF00407	1	161	308.2	6.4e-90	Bet_v_I
Q43550	1	159	PF00407	1	161	339.9	1.8e-99	Bet_v_I
Q43551	1	159	PF00407	1	161	328.1	6.6e-96	Bet_v_I
Q43552	1	159	PF00407	1	161	332.4	3.3e-97	Bet_v_I
Q43861	2	62	PF00544	296	365	29.7	2.8e-08	pec_lyase
Q43862	107	431	PF00544	1	403	289.0	3.8e-84	pec_lyase
Q96365	1	159	PF00407	1	161	390.3	1.2e-114	Bet_v_I
Q96366	1	159	PF00407	1	161	392.8	2.2e-115	Bet_v_I
Q96367	1	159	PF00407	1	161	391.4	5.9e-115	Bet_v_I
Q96368	1	159	PF00407	1	161	385.1	4.7e-113	Bet_v_I
Q96370	1	159	PF00407	1	161	392.4	2.9e-115	Bet_v_I
Q96371	1	159	PF00407	1	161	393.3	1.6e-115	Bet_v_I
Q96377	1	159	PF00407	1	161	379.9	1.7e-111	Bet_v_I
Q96378	1	159	PF00407	1	161	378.7	3.7e-111	Bet_v_I
Q96379	1	159	PF00407	1	161	378.7	3.9e-111	Bet_v_I
Q96380	1	159	PF00407	1	161	362.6	2.8e-106	Bet_v_I
Q96381	1	159	PF00407	1	161	351.3	6.8e-103	Bet_v_I
Q96382	1	159	PF00407	1	161	358.9	3.6e-105	Bet_v_I
Q96501	1	159	PF00407	1	161	366.5	1.8e-107	Bet_v_I
Q96503	1	159	PF00407	1	161	378.5	4.5e-111	Bet_v_I
Q9AUD1	56	146	PF00234	1	129	52.9	4.6e-13	tryp_alpha_amyl
Q9AVK8	37	167	PF00190	1	182	81.0	1.6e-21	Cupin
Q9AVK8	276	426	PF00190	1	182	121.2	1.3e-33	Cupin
Q9FPK2	1	160	PF00407	1	161	342.3	3.4e-100	Bet_v_I
Q9FPK3	1	160	PF00407	1	161	340.6	1.1e-99	Bet_v_I
Q9FPK4	1	160	PF00407	1	161	343.6	1.4e-100	Bet_v_I
Q9FUW6	2	300	PF02716	1	322	773.9	4e-230	Isoflavone_redu
Q9FY19	87	429	PF00295	1	362	364.2	9.1e-107	Glyco_hydro_28
Q9FZP9	157	284	PF00190	1	182	90.1	2.8e-24	Cupin
Q9FZP9	360	521	PF00190	1	182	150.6	1.8e-42	Cupin
Q9LD01	170	251	PF01357	1	84	155.6	5.7e-44	Pollen_allergen
Q9LD05	170	251	PF01357	1	84	169.3	4.1e-48	Pollen_allergen
Q9LD06	172	253	PF01357	1	84	150.7	1.6e-42	Pollen_allergen
Q9LD79	13	96	PF00314	1	207	34.7	1.1e-11	thaumatin
Q9LLT1	39	367	PF00544	1	403	342.3	3.5e-100	pec_lyase

Q9LLT2	39	367	PF00544	1	403	334.2	9.6e-98	pec_lyase
Q9M4S2	39	367	PF00544	1	403	346.1	2.5e-101	pec_lyase
Q9M4S3	39	367	PF00544	1	403	347.0	1.3e-101	pec_lyase
Q9M4S4	39	367	PF00544	1	403	338.1	6.5e-99	pec_lyase
Q9M4S5	39	367	PF00544	1	403	344.6	6.9e-101	pec_lyase
Q9M4S6	39	367	PF00544	1	403	345.6	3.4e-101	pec_lyase
Q9S7M5	1	158	PF00407	1	161	295.2	5.3e-86	Bet_v_I
Q9S8Q4	1	24	PF00407	1	25	26.7	1.8e-07	Bet_v_I
Q9S8S8	1	36	PF00407	1	37	64.8	3.5e-18	Bet_v_I
Q9SBE0	1	123	PF01620	1	195	165.8	4.7e-47	Pollen_allerg_2
Q9SC98	171	251	PF01357	1	84	133.1	3.2e-37	Pollen_allergen
Q9SEW4	187	316	PF00190	1	182	97.4	1.8e-26	Cupin
Q9SEW4	386	556	PF00190	1	182	191.8	7.2e-55	Cupin
Q9SQG5	31	120	PF00234	1	129	46.6	3.5e-11	tryp_alpha_amyl
Q9SQH1	45	124	PF00234	1	82	6.8	0.12	tryp_alpha_amyl
Q9SQH7	37	204	PF00190	1	182	199.0	4.7e-57	Cupin
Q9SQH7	358	507	PF00190	1	182	147.8	1.3e-41	Cupin
Q9SWR4	1	160	PF00407	1	161	344.2	9.1e-101	Bet_v_I
Q9SYV3	1	158	PF00407	1	161	288.6	5e-84	Bet_v_I
Q9SYV4	1	158	PF00407	1	161	291.5	6.9e-85	Bet_v_I
Q9SYW3	1	158	PF00407	1	161	296.7	1.9e-86	Bet_v_I
Q9XG86	51	392	PF00295	1	362	517.5	6.5e-153	Glyco_hydro_28
Q9ZR84	170	255	PF01357	1	84	146.9	2.4e-41	Pollen_allergen
Q9ZWJ2	31	155	PF00903	1	121	134.5	1.2e-37	Glyoxalase
Q9ZWJ2	162	289	PF00903	1	121	113.6	2.4e-31	Glyoxalase
RA05_ORYSA	41	149	PF00234	1	129	90.7	2e-24	tryp_alpha_amyl
RA14_ORYSA	41	152	PF00234	1	129	92.0	8e-25	tryp_alpha_amyl
RA17_ORYSA	39	149	PF00234	1	129	91.7	9.7e-25	tryp_alpha_amyl
RAG1_ORYSA	39	94	PF00234	1	75	29.7	5.1e-08	tryp_alpha_amyl
RAG2_ORYSA	41	153	PF00234	1	129	95.0	9.7e-26	tryp_alpha_amyl
RAP_TAROF	1	157	PF00407	1	161	198.0	9.9e-57	Bet_v_I
SBP_CRYJA	4	367	PF00544	1	403	413.5	1.3e-121	pec_lyase
SODC_OLEEU	1	29	PF00080	1	32	25.1	5.5e-08	sodcu
TLP_PRUAV	30	245	PF00314	1	207	290.5	1.3e-84	thaumatin
ZEA3_MAIZE	1	230	PF01559	1	271	113.6	2.5e-31	Zein