

Transcriptome-Based Examination of Putative Pollen Allergens of Rice (*Oryza sativa* ssp. *japonica*)

Scott D. Russell^{a,1}, Prem L. Bhalla^b and Mohan B. Singh^b

^a Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA

^b Plant Molecular Biology and Biotechnology Laboratory, Australian Research Council Centre of Excellence for Integrative Legume Research, Faculty of Land and Food Resources, The University of Melbourne, Parkville, Vic. 3010, Australia

ABSTRACT Pollen allergens are among the most abundantly transcribed and translated products in the life history of plants, and particularly grasses. To identify different pollen allergens in rice, putative allergens were identified in the rice genome and their expression characterized using the Affymetrix 57K rice GeneChip microarray. Among the most abundant pollen-specific candidate transcripts were *Ory s 1* beta-expansin, *Ory s 2*, *Ory s 7* EF hand, *Ory s 11*, *Ory s 12* profilin A, *Ory s 23*, glycosyl hydrolase family 28 (polygalacturonase), and FAD binding proteins. Highly expressed pollen proteins are frequently present in multiple copy numbers, sometimes with mirror images located on nearby regions of the opposite DNA strand. Many of these are intronless and inserted as copies that retain nearly exact copies of their regulatory elements. *Ory s 23* reflects low variability and high copy number, suggesting recent gene amplification. Some copies contain pseudogenes, which may reflect their origin through activity of retrotransposition; some putative allergenic sequences bear fusion products with repeat sequences of transposable elements (LTRs). The abundance of nearby repetitive sequences, activation of transposable elements, and high production of mRNA transcripts appear to coincide in pollen and may contribute to a syndrome in which highly transcribed proteins may be copied and inserted with streamlined features for translation, including grouping and removal of introns.

INTRODUCTION

Pollen allergens represent a major human health challenge that is particularly pernicious among wind-pollinated seed plants and particularly the grasses (Mohapatra et al., 2005). Often, these proteins are among the most abundantly produced proteins in their life history. These proteins may play a biological role in facilitating and enabling the elongation of the pollen tube, which may occur through controlling the internal metabolism of the tube and modifying the immediate environment of the tube. In anemophilous plants such as rice, pollen is at least partially desiccated and is protected by a thick layer of proteins and a lipid coat which together provide a barrier effect. Both proteins and lipids in rice may be implicated in allergic response to pollen (Asero et al., 2007; Sen et al., 2003). Not surprisingly, genes that produce these proteins are frequently present in multiple copies, and are frequently closely situated to one another. Grasses have evolved a group of allergens that appear to be largely restricted to species of Poaceae. In maize, the recent proliferation of Group 1 pollen allergens has been well documented, suggesting a rapid diversification of cell wall modifying genes that also occupy a role in human allergy (Valdivia et al., 2007). A similar pattern of expression

was proposed using the rice genome to identify putative pollen allergens with these domains and other motifs (Jiang et al., 2005).

In the rice genome, pollen allergens represent conspicuous and highly transcribed groups of genes with conserved sequences. In order to further understand their relationship and diversification, an investigation was conducted on their genomic sequence and placement, as well as the nature of conservation of coding and regulatory sequences. Given the known doubling of chromosomes in rice and the age of this event (Wang et al., 2007), conservation of pollen allergens could be reflected in copy numbers, positioning and diversification of function of expressed proteins. In general, the pollen transcriptome has been found to be the most highly divergent

¹ To whom correspondence should be addressed. E-mail: srussell@ou.edu, fax +1-405-325-7619, tel. +1-405-325-7619.

© The Author 2008. Published by the Molecular Plant Shanghai Editorial Office in association with Oxford University Press on behalf of CSPP and IPPE, SIBS, CAS.

doi: 10.1093/mp/ssn036, Advance Access publication 21 July 2008

Received 28 April 2008; accepted 12 June 2008

of flowering plant life history according to microarrays of *Arabidopsis* (Becker and Feijo, 2007 and references therein) and MPSS (massively parallel signature sequencing of transcripts, Nobuta et al., 2007). Proteomic data indicate that pollen transcripts are translated at high expression levels at anthesis in *Arabidopsis* (Holmes-Davis et al., 2005; Noir et al., 2005; Sheoran et al., 2006), tomato (Sheoran et al., 2007) and in rice (Dai et al., 2006), and that dynamic changes occur during pollen germination and initial elongation of the tube (Dai et al., 2007). We undertook this examination of pollen expression in rice employing the Affymetrix 57K rice genome chip to assess whether suspected allergens indeed express a pollen enhanced profile. Genomic features will then be examined that may contribute to explaining their high level of expression.

RESULTS AND DISCUSSION

Pollen allergens of *Oryza sativa* recognized by the International Union of Immunological Societies (IUIS) official list of allergens (www.allergen.org/) include Ory s 1, which represents a beta-expansin gene; Ory s 7, which represents an EF-hand protein; and Ory s 12, which represents profilin A. Non-IUIS Ory s 2, Ory s 11, and Ory s 23 are largely designated from rice genome annotations of homologues of Phl p 2, Phl p 11, and Cyn d 23. In a survey of the rice genome for potential allergens, Jiang et al. (2005) used homologues of three general pollen allergen motifs: Ole e 1, expansins and ribonucleases, but their occurrence was not validated for pollen expression. There are additional candidate protein families that were not considered, as well. In the current study, Affymetrix rice genomic microarray tests revealed the most highly expressed and up-regulated pollen genes bearing similarities with grass allergens. These findings were confirmed by MPSS signatures of anthesis rice pollen (dataset NPO at www.mpss.udel.edu/rice/), expressed sequence tags (ESTs, at TIGR and NCBI), and proteome data from developing and mature rice anthers (Kerim et al., 2003a, 2003b; Imin et al., 2004; Dai et al., 2006, 2007).

Canonical pollen allergens are characteristically highly expressed in pollen, often largely restricted to gametophytic expression, and involved in unique attributes of pollen physiology. Radauer and Breiteneder (2006) examined pollen allergens known to elicit allergenic response and categorized them into prevalent protein groups, which we then screened to identify candidates based on high expression and up-regulation in pollen.

In the only other genomic survey of rice pollen allergens, Jiang et al. (2005) identified, as putative pollen putative allergens, 45 proteins with 'Ole e 1-like motifs', 65 proteins with Pfam motifs designated as 'Protein_allerg_1' (expansin motif), and seven proteins with Pfam motifs designated as 'Protein_allerg_2' (ribonuclease motif). Once sequences were assigned to TIGR and IRGSP loci, transcriptional expression was determined for each candidate using the Affymetrix chip, MPSS and proteome data. Of Jiang et al. (2005) putative allergens, 10 of the 117 proteins were highly expressed in pollen. Two

candidates of the so-called OsOle series showed pollen expression (OsOle24 and OsOle28), eight candidates of the OsAllerg1 group showed high activity (3 Ory s 1 proteins, 3 Ory s 2 proteins and two expansin candidates), and none of the seven OsAllerg2 group candidates showed high up-regulation in pollen (see supplementary spreadsheet). Specific candidates are further characterized in the supplementary data, and grouped below into major allergen groups.

Pollen Major Group 1 Allergens

Related to the β -expansins, these proteins account for most of the major group I allergens of grasses (Sampedro and Cosgrove, 2005), with which they share high sequence similarity (Jiang et al., 2005). In rice, pollen proteins containing this characteristic protein motif include Ory s 1, which has been validated as an allergen on the basis of its recognition by IgE antibodies from allergic individuals (Xu et al., 1995), and Ory s 2, which represents a smaller protein containing a similar motif. Extensive supplementary data complement this and other topics in the results.

Ory s 1

The Ory s 1 family of pollen allergens is closely related to the β -expansins, major group I allergens of grasses, such as Zea m 1; β -expansin shares high similarity in a number of motifs and its genomic representatives reflect a dynamic history (Sampedro et al., 2005). The sequence of Ory s 1 is consistent with a group I allergen, encoding a putatively secreted water-soluble protein with allergenic properties. Ory s 1-encoded protein sequences contain a signal peptide at position 22 (VSC-GP), a 'rare lipoprotein A-like double-psi beta-barrel' motif (PF03330) at positions 78–157 and a pollen_allerg_I motif (PF01357) from position 170 to 251.

Ory s 1 bears closest sequence relationship in the grasses to Phl p 1 of *Poa*, sharing 74% identity, 84% positives and $p < 1e^{-129}$. In *Oryza sativa* ssp. *japonica*, Ory s 1 is represented by four gene loci that are in tandem association near the origin of chromosome 3 (Figure 1). The loci range from LOC_Os03g01610 to LOC_Os03g01650 and consist of a coding region that is typically 804 bp, encoding a putative protein of 267 amino acids. Two loci are identical in the coding region (LOC_Os03g01610 and LOC_Os03g01650), but interestingly are encoded in opposite orientations on chromosome 3; LOC_Os03g01610 is on the (+) strand (as are the other two loci) and LOC_Os03g01650 on the (–) strand (Figure 1). They share a similar promoter motif (Xu et al., 1999).

Ory s 2

Ory s 2 is a Group II/III pollen allergen of 117 amino acid length that has a core 'pollen_allerg_I' motif (Pfam: PF01357) at positions 26–102. A signal peptide at position 23 (most often SCATE) is consistent with this gene locus, representing a secreted protein product. Similarly to Ory s 1, the role of Ory s 2 is expected to involve cell wall modification, sharing one major motif held in common with expansins. Unlike Ory s 1 and canonical expansins, however, Ory s 2 lacks a RlpA-like DPBB_1

motif (Pfam: PF03330). Available data suggest that Ory s 2 is highly transcribed, with eight probe sets of the 57K rice genomic microarray report up-regulation $>6 \log_2$ in pollen—and six of these higher than $11 \log_2$ (>2000 -fold). The high degree of up-regulation of the transcripts of Ory s 2 is even more strongly supported by MPSS, which reports significant Ory s 2 transcription for 10 signature sequences, of which seven are in the top 25 for pollen intensity. Abundance of translated products is supported by proteome data (Dai et al., 2006, 2007). According to Kerim et al. (2003b), Ory s 2 has multiple isoforms displaying some amino acid differences and therefore they named them as three different forms, largely on the basis of their N-terminus.

The genomic arrangement of the Ory s 2 family suggests that its loci natively lack introns. The intron-bearing loci appear to represent altered copies that may be functionless pseu-

dogenes (see LOC_Os04g26190 and LOC_Os06g45200, supplementary data). There appear to be four major groups of Ory s 2 genes—two groups are located on chromosome 4 and two on chromosome 6, with triplicate pairings being a common pattern (Figure 2). Chromosome 4 repeats the theme of sharing strong sequence similarity with inverted loci on the opposite strand. Additional data are available as supplementary data.

Additional Potential Group I, II/III Allergens

In addition to Ory s 1 and Ory s 2, there are three additional very highly up-regulated expansin-related proteins in pollen. Each has an expansin motif, as well as a rare lipoprotein A (RlpA)-like double-psi beta-barrel motif (PF03330), as in Ory s 1. One includes LOC_Os10g40090.1, which is a Beta-expansin 1a precursor with the third highest abundance in the

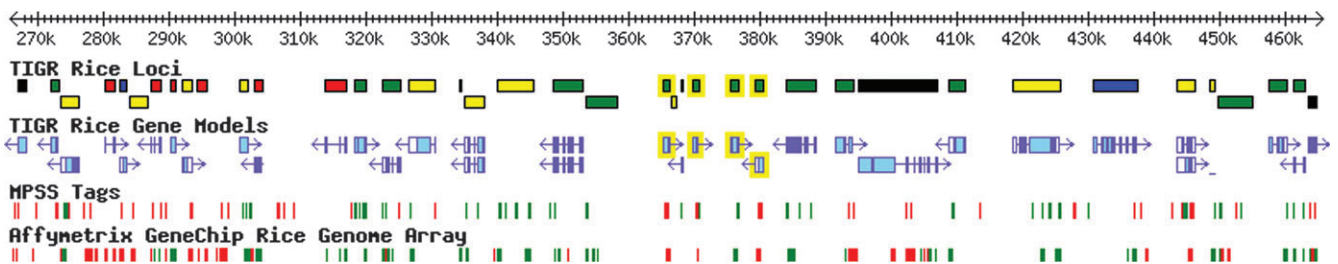


Figure 1. Genomic Organization of Ory s 1 Loci on Chromosome 3.

Ory s 1 loci (green, highlighted in yellow) are, from left: LOC_Os03g01610, LOC_Os03g01630, LOC_Os03g01640, and LOC_Os03g01650. TIGR Rice Gene Loci line shows gene placement, with orientation indicated in line labeled TIGR Rice Gene Models. The high degree of sequence overlap is reflected in overlapping MPSS signatures and Affymetrix GeneChip probe sets, indicated in red, contrasted with unique sequences in green. (Graphics modified from TIGR Rice Genome Browser.)

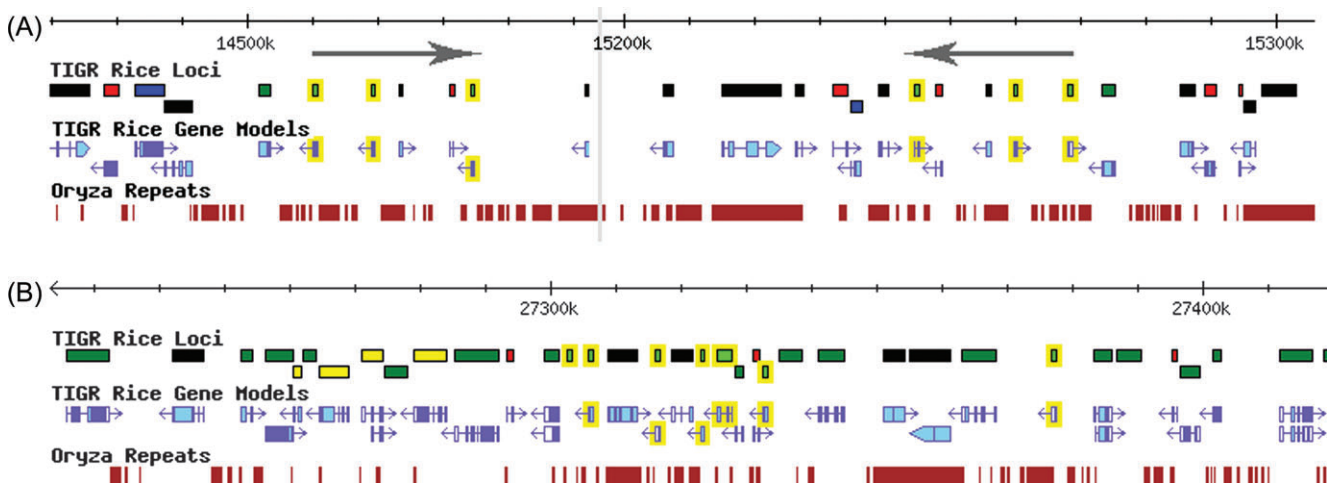


Figure 2. Genomic Organization of Ory s 2 Loci (Green, Highlighted in Yellow) on Chromosome 4.

(A) Reflects mirror images of loci (left: LOC_Os04g25150, LOC_Os04g25160, LOC_Os04g25190, and right: LOC_Os04g26190, LOC_Os04g26220, and LOC_Os04g26230).

(B) Chromosome 6 loci have a tight distribution of related sequences on (-) strand. Oryza repeats illustrate repeat elements, including LTRs, flanking the coding regions of Ory s 2 loci. (Graphics modified from TIGR Rice Genome Browser.)

Affymetrix microarray, 177 ESTs, 34th MPSS abundance and strong proteome support (see supplementary data). Two others (LOC_Os08g44790 and LOC_Os12g36040) are also highly up-regulated and apparently pollen-specific. These are obvious candidates but have not yet been tested for allergenic activity.

Profilin A (Ory s 12)

Ory s 12 is the pollen allergen corresponding to pollen-expressed profilin, which has long been known as a potent allergen (Valenta et al., 1991). Since the known functional role of profilin—principally as a protein involved with controlling actin polymerization—is exclusively intracellular, the release of profilin from pollen as an allergen is likely the result of protein released from damaged, non-intact or dead pollen, possibly forming small aerosol particles that may be inhaled (Swoboda et al., 2004). In rice, pollen-related profilin A transcripts appear to be highly expressed (intensity 8.120 log₂) and up-regulated by 11.74 log₂ times (>2000×) above seedlings, according to microarray results. MPSS indicates that transcript abundance is among the top 30 signatures, of which two loci were specific for pollen-expressed profilin; there was strong proteome support (supplementary data). In rice, the two loci active exclusively in pollen are LOC_Os10g17660 and LOC_Os10g17680. These are inverted on opposite strands and share high sequence similarity, differing by only six nucleotides. In the rice genome, there are only three loci devoted to profilin, of which two seem only to be expressed in pollen (see supplementary data). All other sporophytic expression seems to correspond to a remaining profilin-2 (LOC_Os06g05880). Profilins are typically involved in the sequestration of actin monomers and are highly expressed in pollen, presumably as a protein that modulates polymerization and the dynamic nature of the elaborate actin cytoskeleton of elongating pollen tubes. These appear to be strong candidates.

Ory s 23

Ory s 23 was the sixth most abundant transcript according to the 57K microarray, second according to MPSS (mpss.udel.edu/rice/, dataset NPO) with a fold up-regulation of between ×1000 (MPSS) and ×8000 (57K chip). These transcripts are translated and abundant, occurring as water-soluble secreted proteins according to proteome evidence from mature pollen and in-vitro germinated pollen (Dai et al., 2006, 2007). Surely this degree of transcription and translation would suggest an important role in facilitating rice pollen function, but its role is as yet undetermined. Ory s 23 has as its closest homologue Cyn d 23, which was first described in *Cynodon dactylon* as a 112 amino acid pollen allergen peptide (AAP80170.1). The similarity between this gene and the version in *Cynodon dactylon* is 37% amino acid identity with 55% positives and two significant gaps, $p < 5e^{-14}$ and therefore there may be significant differences in function of the protein, its potential substrate target or both. While allergenic response is suggested by its possessing an IUIS designation (www.allergen.org/), the med-

ical possibilities of this potential allergen protein are not currently characterized in rice. If this pollen product is similar to other described secreted pollen allergens, as a secreted product, it might be expected to have enzymatic, possibly cell wall-altering, activity, which may function in facilitating pollen tube passage within the gynoecium (Valdivia et al., 2007) or some other highly up-regulated pollen function. These appear to be strong candidates.

The Ory s 23 gene family in rice consists of 17 very closely related loci, sharing high sequence synonymy and separation by less than 200 kbp, with all copies of the gene located within the interval from TIGR loci LOC_Os09g23899 to LOC_Os09g24150. All are on the Crick (–) strand and 13 of these bear UTR sequences and strong EST support (<http://rice.tigr.org/>). The coding region of each locus consists of an intronless 405 nucleotide sequence encoding a 134 amino acid protein with signal peptide at position 16 (ASA-VL). These peptide sequences have two distinct forms, differing at positions 48, 79, 81, 82, 128, and 130. The two alternative peptide sequences are highlighted as yellow and light blue in Figure 3. Six single nucleotide polymorphisms are present that code for three single amino acid substitutions in LOC_Os09g24020, LOC_Os09g24100, and LOC_Os09g24130, each of which is annotated as an expressed product.

Interestingly, sequence synonymy also extends to upstream sequences, expanding sometimes to the full 1.4-kbp distance between coding regions of adjacent loci; these typically differ by only a few basepairs, despite the fact that this is a non-coding-region, with the exception of three upstream sequences that have apparently been displaced (see supplementary data). The low sequence variability expressed by Ory s 23 suggests either a very short evolutionary history for this gene at its current copy number, or a mechanism that has extraordinarily strict sequence conservation. Given that non-coding upstream and UTR sequences provide an almost complete match, with relatively little variability except along two motifs, the suggestion is that this has a short history and that a faithful copy mechanism that can capture non-coding regions seems a likely explanation. The grouping of different sequence motifs as well as the distance between coding sequences suggests that multiple loci may have been inserted at the same time, potentially in pairs.

Transposable element (TE) motifs are linked to three additional nearly complete Ory s 23 sequences. One open reading frame has a long leading retrotransposon sequence that constitutes a properly terminated Ory s 23 sequence at LOC_Os09g23980. Two shorter, trailing TE motifs that appear anomalously terminated are connected to reading frames at LOC_Os09g24150, with a 14-bp overlapping transposon motif, and LOC_Os09g23930, with a 17-bp overlapping retrotransposon motif at the end of an otherwise normal Ory s 23 sequence. The anomalous flanking sequences have high homology with TE sequences located on other chromosomes, suggesting multiple independent insertions of the TE sequences, which may have potentially arisen as an inadvertent consequence of TE activity itself. The presence of multiple repetitive elements in broad intervals between tandem Ory s 23 sequences and

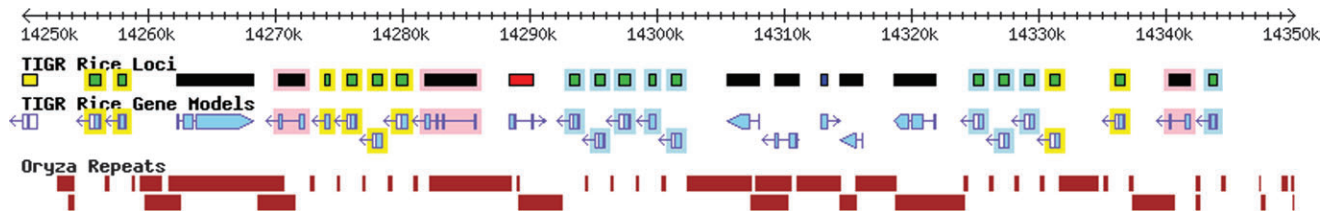


Figure 3. Genomic Organization of Ory s 23 Loci.

Rectangles with green interiors display two variants (highlighted in light blue and yellow) that differ by six amino acids on chromosome 9; leftmost Ory s 23 locus is LOC_Os09g23899; rightmost is LOC_Os09g24159. Ory s 23 loci are intronless, separated by narrow repetitive intervals of typically ≈ 1400 bp—a feature of the Ory s 23 gene family. Transposable element (TE) motifs (highlighted in pink) form three nearly complete Ory s 23 sequences with leading or trailing TE motifs. Oryza repeats indicate repeat elements, including LTRs, flanking the coding regions. (Graphic modified from TIGR Rice Genome Browser.)

short repetitive elements between Ory s 23 sequences in non-coding regions suggests that this region may be a target of TE insertion. The alternative of DNA repair genes selecting this region as a target remains a possibility but the presence of multiple TE loci between Ory s 23 tandem repeats implicates action by TEs in their origin. The presence of such sequences may represent pseudogenes that may have been inserted by retrotransposable elements. Wang et al. (2006), for example, unexpectedly observed that retroposition was involved in generating large numbers of intronless genes in rice and provided strong evidence for involvement of RTEs as a mechanism of insertion. RTEs, representing a ‘copy and paste’ transposition, can rapidly increase gene copy numbers and plant genome size, as well as providing, over time, a source of paralog formation (Bennetzen, 2007). The rapid and highly related sequences of Ory s 23 noted here, along with LTR-repeat linked variants, make RTE insertion an attractive mechanism for amplification of transcription of this sequence pair (see supplementary data).

Other Putative Allergens

Among the 2615 protein families of angiosperms, Radauer and Breiteneder (2006) found that allergenic pollen proteins occur in 29 of these, eight of which are common in Poaceae; half of these groups are covered above: expansin-related products, ‘Ole e 1-like’ proteins, and profilin. Among these, the remaining groups include EF hand, glycosyl hydrolase family 28, and FAD binding.

The EF-hand protein family is responsible for the rice allergen known as Ory s 7 and is highly up-regulated and expressed in rice pollen at two loci: LOC_Os08g44660, with an intensity of $7.894 \log_2$, up-regulated $12.526 \log_2$ -fold ($\sim 4000\times$), making it one of the most highly up-regulated sequences, and LOC_Os12g12730, with an intensity of $7.538 \log_2$, up-regulated $9.671 \log_2$ -fold ($\sim 700\times$). A third product (LOC_Os04g45180) is highly expressed, but only about four times more than seedling, suggesting wide sporophytic expression as well.

Pectate lyase is highly up-regulated ($>10 \log_2$ -fold) and expressed ($>7 \log_2$) in pollen at five loci on the rice genome in the 57K chip, with one locus also highly expressed in MPSS (LOC_Os06g05260), but no proteome support. The five pollen-

expressed loci share sequences that distinguish them from the predominantly sporophytically expressed loci; four are located on chromosome 6, one on chromosome 2. The most closely related are LOC_Os06g05209 and LOC_Os06g05272, which are on (–) and (+) strands, respectively. These appear to be weak candidates (supplementary data).

Glycosyl hydrolase family 28 proteins are most abundantly represented in rice pollen by the exopolysaccharuronases, which have well documented representatives in the grasses that elicit an allergic response (Swoboda et al., 2004). Of 68 loci that have been annotated as polygalacturonases in the rice genome (TIGR, v5), there were eight with ESTs that displayed $>6 \log_2$ higher expression than in seedlings, and expression is absent in other tissues (GEOS data). Three are most likely to be pollen allergen candidates: LOC_Os06g35320, LOC_Os06g35370, and LOC_Os02g10300, each of which is highly supported by microarray, MPSS and proteome (supplementary data). Each has a signal peptide and is expected to represent a secreted protein that may express wall-modifying properties that may alter PT walls or their penetration into the gynoecium. All polygalacturonases expressed highly in pollen seem to have introns, but there seems to be an inverse relationship between the numbers of introns and pollen expression (supplementary data).

FAD-binding motif is present in only one representative, annotated as a reticuline oxide precursor, an oxidoreductase gene typically involved in secondary metabolism (similar to tetrahydroprotoberberine synthase, At5g44410, $1.00e^{-137}$); the predicted protein has a signal peptide at position 29 and is likely secreted. With a $\log_2$ intensity of 8.508 and very high up-regulation over seedling controls ($12.414 \log_2$), the expression profile is similar to that of pollen allergens, and has MPSS and proteome support.

Evaluation of Proteins for Potential Allergenic Activity

Clearly, clinical testing is needed to determine the medical relevance of each of these protein candidates to determine their allergenic potential in humans. Skin testing, radioallergosorbent test (RAST), immunoblotting and allergen-induced lymphoproliferation have been used to assess asthmatic children and revealed three major allergen molecular weight classes of approximately 16, 26, and 32 kDa, as well as other

minor allergens not described in detail (Tsai et al., 1990). These molecular weights correspond to a number of candidates; for example, Ory s 1 is undoubtedly represented by the 32-kDa protein (Ory s 1 has a reported MW of 30 kDa, according to Kerim et al., 2003a, 2003b, and 34 kDa according to Xu et al., 1995). The 26-kDa line likely also represents a form of expansin, presumably one of the top two candidates under 'additional potential Group I, II/III allergens'. The 16-kDa line may represent a composite of Ory s 2, which has been measured at 12–13 kDa (Kerim et al., 2003b), profilin Ory s 12 measured at ~14 kDa and Ory s 23, predicted at 13.99 kDa. The others may represent minor allergens that could become problematic with chronic exposure, as with most pollens. Rice pollen allergy has remained relatively little studied compared to other grasses. Population centers are typically remote to rice paddies and have more conspicuous allergenic threats. Another contributing factor may be that rice pollen is extremely short-lived. *Oryza sativa* pollen viability is reduced below 50% within 6 min after leaving the anther and essentially all are non-viable in a mere 20 min (Song et al., 2001).

Expression and Potential Amplification of Putative Pollen Allergens

Pollen allergens are among the most highly expressed among any cell or tissue type in rice, independently of the method of normalization used (in this study, Affymetrix GCOS, GCRMA, dChip, RMA, and PLM yielded similarly high intensities compared to seedling controls and GEO datasets). That pollen expresses a highly divergent and abundant complement of pollen transcripts is supported by MPSS (Nobuta et al., 2007; www.mpss.udel.edu/rice/, dataset NPO) and has been noted in regard to the *Arabidopsis* transcriptome as well (Becker and Feijo, 2007). Proteome studies also support that rice pollen allergens are translated at high levels of abundance (Kerim et al., 2003a, 2003b; Imin et al., 2004; Dai et al., 2006, 2007).

Most highly expressed pollen allergens are present in multiple copies, often are intronless, and the most abundant are often grouped. Additionally, not only the directly expressed sequences but also some non-coding portions of these gene loci may be highly conserved, including UTRs and upstream regulatory sequences. These multiple genes could possibly have arisen through genomic duplication, as occurred in rice at approximately 57 MYA (Wang et al., 2007), but the similarity of these sequences, as well as their typical restriction to a single chromosome, suggests a more recent origin. Many of the most abundantly transcribed pollen proteins are encoded by loci lacking introns or containing a reduced number of introns. Since introns in some cases may be recruited in post-transcriptional control (such as through the generation of siRNAs, Piriyapongsa and Jordan, 2008), elimination of introns could potentially release translational regulation, leading to further enhanced expression.

Commonly, highly expressed sequences were encoded in pairs on opposite DNA strands grouped within several kbp. Sequence duplication, accompanied by production of nearby pseudogenes, has been suggested as a signature for the

activity of transposable elements, particularly retrotransposons, in creating chimeral genes (Wang et al., 2006). Retrotransposon-mediated gene duplication has been reported in other crop plants and appears to account for much of the relatively rapid origin and high degree of variability in tomato fruit morphology (Xiao et al., 2008). These gene insertions and duplications in tomato appeared to be primed by the occurrence of repeat sequences near critical existing genes. In *Oryza*, many of the most highly transcribed pollen allergen loci also appear to be linked with nearby LTR-associated sequences. Frequently, gene duplication in eukaryotes has been reported to occur through non-homologous DNA end-joining as a common motif (Gu and Lieber, 2008), as well as illegitimate recombination whereby a DNA repair mismatch may insert a duplicated strand of DNA that is incorporated into the genome (Guan et al., 2007). However, reported bursts of LTR-retrotransposon activation (Vitte et al., 2007) may leave LTR repeats flanking duplicated sequences.

The extreme degree of precision that seems to be evident in the insertion of Ory s 23 sequences (including nearly 100% duplication of non-coding, as well as coding sequences) seems highly unusual unless it is a relatively recent occurrence. Although genomic sequences have reportedly been precisely removed by retrotransposable elements (van de Lagemaat et al., 2005), similar precision in insertion has not. These segments of repeated nucleotide sequences may thus contribute to retrotransposon insertion but also favor alternative methods of illegitimate DNA mismatch as well. In general, retrotransposable elements tend to activate in the context of hypomethylated DNA. Hypomethylation may also trigger read-through transcription that inadvertently activates linked genes (Cheng et al., 2006). It may be noteworthy that pollen is reported to display especially low levels of DNA methylation compared to sporophytic cells (Oakeley et al., 1997), which, if true, could potentially favor such high transcription rates.

Amplification of pollen allergen transcription, particularly in grasses, is well known, and evidence suggests a role for retrotransposition. Wang et al. (2006) found numerous examples of chimeric retrogenes in rice representing captured mRNA messages that have been reverse transcribed and inserted, lacking introns, often in reverse direction; current bioinformatic data on putative pollen allergens concur with this possibility, as well as attachment to signature LTR sequences. As younger retrogenes accumulate indels rapidly, LTR flanking regions will decay, and the retrogene sequence will soon lose direct evidence of its LTR origin (Ma et al., 2004). Such a mechanism could provide positive feedback for the production of highly transcribed sequences and account for amplification over time.

METHODS

Microarray

Affymetrix 57K Rice GeneChip™ microarray provides essentially genomic coverage of genes. The microarray consists of 57 381 probe sets with 182 control and reporter probe sets.

Specific putative pollen allergy candidates were selected from expression in protein groups noted for human allergenic responses in grass pollen (Radauer and Breiteneder, 2006) and from genomic predictions (Jiang et al., 2005). Screened pollen allergens in this study were identified as candidates based on high up-regulation and expression in pollen. Transcription and translation were detected using the Affymetrix 57K rice genome microarray, MPSS and proteome data. Overall expression profiles of candidate genes are available on Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/, platform GPL2025).

Pollen Isolation

Mature anthers were collected from field-grown rice (*Oryza sativa* ssp. *japonica*, cultivar Katy) at the Dale Bumpers National Rice Research Center and University of Arkansas Extension Station near Stuttgart, Arkansas, to obtain pollen isolates. Collections were made on three different days using three different field sites, forming biological/technical replicates that were kept separated at all subsequent steps in processing. Field material displayed high vigor and no sign of disease, which was supported by absence calls for major reporter pathogens on the Affymetrix 57K rice microarray controls.

RNA Preparation

Mature viable rice pollen (Figure 4) was isolated by grinding collected anthers in a cold 45% sucrose solution. Pollen cytoplasm was collected by filtering out debris using a 100- μ m nylon mesh membrane and pelleting the resulting fraction at 300 g for 3 min. Pollen was washed once in cold 45% sucrose solution and stored frozen at or below -80°C until use. Seedlings were collected as a sporophytic control from rice seed of Katy germinated in soil and cultured to developmental stage V3 (collar forms on leaf 3 on main stem), according to Counce et al. (2000), and harvested. Seedling tissue was frozen in liquid nitrogen, ground into a fine powder and then stored until needed. All samples were stored at or below -80°C until RNA isolation. Total RNA was purified using the RNeasy plant mini kit according to manufacturer's instructions (Qiagen). RNA

concentration and quality of pollen and seedlings were determined by routine spectrophotometric measurement and agarose gels. 100 μ g total RNA of seedlings and mature pollen were used for probe preparation for each of the three biological/technical replicates performed.

Oligonucleotide Microarray Hybridization and Data Collection

Since the amount of starting total RNA was low, the GeneChip Two-Cycle cDNA Synthesis Kit from Affymetrix (Santa Clara, CA, USA) was used for target preparation with signal amplification. The Affymetrix GeneChip Rice Genome oligonucleotide array was used for the hybridization (45°C for 16 h) with a mixture containing 15 μ g of fragmented cRNA. Subsequent washing and staining steps were performed on a GeneChip Fluidics Station 450 and the chips were scanned on a GeneChip Scanner 3000. All experimental procedures strictly followed instructions specified in the Affymetrix GeneChip Expression Analysis Technical Manual. Instrument control and data collection were carried out with GeneChip Operating Software (GCOS, ver. 1.1.1). In order to minimize experimental variability, standardized microarray operation was performed by an experienced investigator throughout the study. The quality and quantity of the original RNA samples and of the cRNA probes generated were determined with the Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA), and by spectrophotometric measurements at 260 and 280 nm on a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). The microarray data generated from all chips met the quality-control criteria set by Affymetrix. Consistency of the data was supported by Pearson's coefficient of correlation (r), which, for pollen biological/technical replicates, $r = 0.9898, 0.9931, \text{ and } 0.9971$, and, for control seedling replicates, $r = 0.9918, 0.9925, \text{ and } 0.9929$. Calls for transcripts were calculated based on the Affymetrix PM/MM method using GCOS software (version 1.1.1) as being present ($p < 0.05$), absent ($p > 0.10$) or marginal ($0.05 < p < 0.10$). The values for probe set intensity and up-regulation were calculated using GC Robust Multi-array Average (GCRMA)

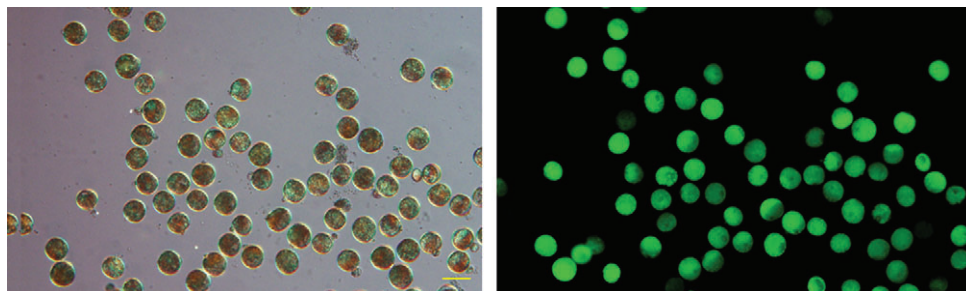


Figure 4. Photomicrograph of Isolated Pollen in Differential Interference Contrast (Left) and Fluorescein Diacetate Viability Screen in Epifluorescence Microscopy (Right).

Bar = 50 μ m.

background adjustment, quantile normalization, and median-polish summarization on Affymetrix microarray probe-level data using afflimGUI (Irizarry et al., 2003; Wu et al., 2004). Values correspond to the A and M parameters, with B parameter (p values) indicating consistency of the grouped data. To determine probe matches, probe sequences were compared with TIGR rice pseudochromosome sequences (v. 5) using blastn to determine number of probes with exact 25-mer match. These are listed in parentheses in the tables of the charts. dChip software (Li and Wong, 2003; www.biostat.harvard.edu/complab/dchip/) was used for additional independent analysis of data and for GEO comparisons (at a CEL file level) of different sporophytic tissues.

SUPPLEMENTARY DATA

Supplementary Data are available at www.mplant.oxfordjournals.org.

FUNDING

Financial support was provided by Australian Research Council (PLB, MBS), University of Melbourne and University of Oklahoma (SDR).

ACKNOWLEDGMENTS

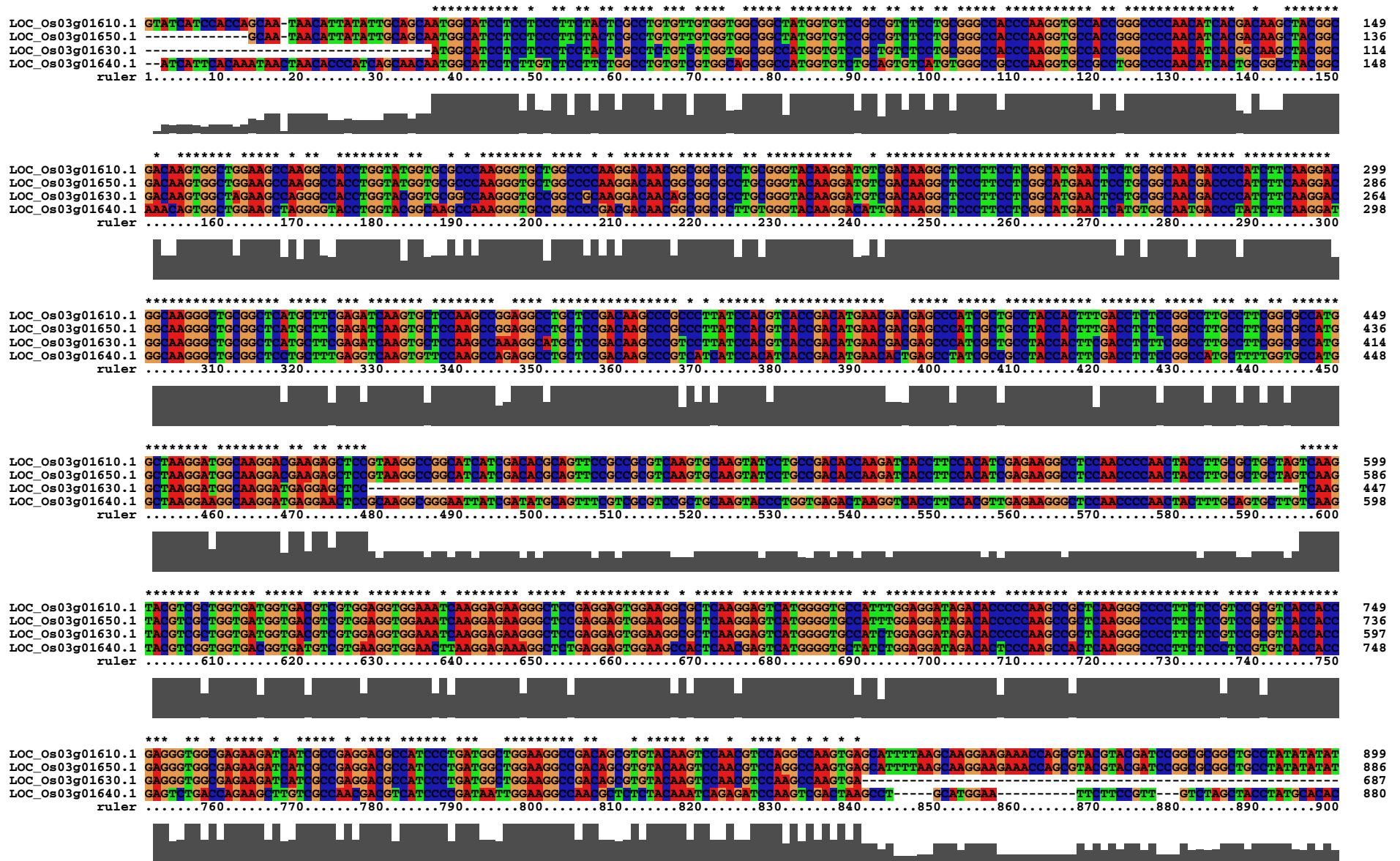
We thank Dr Xiaoping Gou, Dr Tong Yuan, Xiaoping Wei, and Cal Lemke, University of Oklahoma, for excellent field, technical, and growth assistance, and Dr Xinkun Wang, University of Kansas Genomics Facility for Affymetrix processing. Special thanks are due to Dr Yulin Jia, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, USA, for providing field material and advice during our collections and to Prof. Karen Moldenhauer, University of Arkansas Extension Station, Stuttgart, for the seed. We thank Professor Terry Speed and the staff of WEHI software development for help with statistical analysis. No conflict of interest declared.

REFERENCES

- Asero, R., Amato, S., Alfieri, B., Folloni, S., and Mistrello, G. (2007). Rice: another potential cause of food allergy in patients sensitized to lipid transfer protein. *Int. Arch. Allergy Immunol.* **143**, 69–74.
- Becker, J.D., and Feijo, J.A. (2007). How many genes are needed to make a pollen tube? Lessons from transcriptomics. *Ann. Bot.* **100**, 1117–1123.
- Bennetzen, J.L. (2007). Patterns in grass genome evolution. *Curr. Opin. Plant Biol.* **10**, 176–181.
- Cheng, C., Daigen, M., and Hirochika, H. (2006). Epigenetic regulation of the rice retrotransposon Tos17. *Mol. Genet. Genomics.* **276**, 378–390.
- Counce, P.A., Keisling, T.C., and Mitchell, A.J. (2000). A uniform, objective, and adaptive system for expressing rice development. *Crop Science.* **40**, 436–443.
- Dai, S.J., Chen, T.T., Chong, K., Xue, Y.B., Liu, S.Q., and Wang, T. (2007). Proteomic identification of differentially expressed proteins associated with pollen germination and tube growth reveals characteristics of germinated *Oryza sativa* pollen. *Mol. Cell Proteomics.* **6**, 207–230.
- Dai, S.J., Li, L., Chen, T.T., Chong, K., Xue, Y.B., and Wang, T. (2006). Proteomic analyses of *Oryza sativa* mature pollen reveal novel proteins associated with pollen germination and tube growth. *Proteomics.* **6**, 2504–2529.
- Gu, J., and Lieber, M.R. (2008). Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining. *Genes Dev.* **22**, 411–415.
- Guan, Y., Dunham, M.J., and Troyanskaya, O.G. (2007). Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics.* **175**, 933–943.
- Holmes-Davis, R., Tanaka, C.K., Vensel, W.H., Hurkman, W.J., and McCormick, S. (2005). Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics.* **5**, 4864–4884.
- Imin, N., Kerim, T., Rolfe, B.G., and Weinman, J.J. (2004). Effect of early cold stress on the maturation of rice anthers. *Proteomics.* **4**, 1873–1882.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* **4**, 249–264.
- Jiang, S.Y., Jasmin, P.X., Ting, Y.Y., and Ramachandran, S. (2005). Genome-wide identification and molecular characterization of Ole_e_1, Allerg_1 and Allerg_2 domain-containing pollen-allergen-like genes in *Oryza sativa*. *DNA Res.* **12**, 167–179.
- Kerim, T., Imin, N., Weinman, J.J., and Rolfe, B.G. (2003a). Proteome analysis of male gametophyte development in rice anthers. *Proteomics.* **3**, 738–51.
- Kerim, T., Imin, N., Weinman, J.J., and Rolfe, B.G. (2003b). Proteomic analysis reveals developmentally expressed rice homologues of grass group II pollen allergens. *Functional Plant Biology.* **30**, 843–852.
- Li, C., and Wong, W.H. (2003). DNA-chip analyzer (dChip). In *The Analysis of Gene Expression Data: Methods and Software*, G. Parmigiani, Garrett, E.S., Irizarry, R. and S.L. Zeger, eds (New York: Springer), pp. 120–141.
- Ma, J.X., Devos, K.M., and Bennetzen, J.L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**, 860–869.
- Mohapatra, S., Lockey, R., and Shirley, S. (2005). Immunobiology of grass pollen allergens. *Curr. Allergy Asthma Rep.* **5**, 381–387.
- Nobuta, K., et al. (2007). An expression atlas of rice mRNAs and small RNAs. *Nat. Biotechnol.* **25**, 473–477.
- Noir, S., Brautigam, A., Colby, T., Schmidt, J., and Panstruga, R. (2005). A reference map of the *Arabidopsis thaliana* mature pollen proteome. *Biochem. Biophys. Res. Commun.* **337**, 1257–66.
- Oakeley, E.J., Podesta, A., and Jost, J.P. (1997). Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. *Proc. Natl Acad. Sci. U S A.* **94**, 11721–11725.
- Piriyapongsa, J., and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA.* **14**, 814–821.

- Radauer, C., and Breiteneder, H. (2006). Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J. Allergy Clin. Immunol.* **117**, 141–147.
- Sampedro, J., and Cosgrove, D. (2005). The expansin superfamily. *Genome Biol.* **6**, 242.1–11.
- Sampedro, J., Lee, Y., Carey, R.E., de Pamphilis, C., and Cosgrove, D.J. (2005). Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. *Plant J.* **44**, 409–419.
- Sen, M.M., Adhikari, A., Gupta-Bhattacharya, S., and Chanda, S. (2003). Airborne rice pollen and pollen allergen in an agricultural field: aerobiological and immunochemical evidence. *J. Environ. Monit.* **5**, 959–962.
- Sheoran, I.S., Ross, A.R.S., Olson, D.J.H., and Sawhney, V.K. (2007). Proteomic analysis of tomato (*Solanum lycopersicum*, formerly *Lycopersicon esculentum*) pollen. *J. Exp. Bot.* **58**, 3525–3535.
- Sheoran, I.S., Sproul, K.A., Olson, D.J.H., Ross, A.R.S., and Sawhney, V.K. (2006). Proteome profile and functional classification of proteins in *Arabidopsis thaliana* (Landsberg erecta) mature pollen. *Sex. Plant Reprod.* **19**, 185–196.
- Song, Z.P., Lu, B.R., and Chen, J.K. (2001). A study of pollen viability and longevity in *Oryza rufipogon*, *O. sativa*, and their hybrids. *International Rice Research Notes.* **26**, 31–32.
- Swoboda, I., et al. (2004). Molecular characterization of polygalacturonases as grass pollen-specific marker allergens: expulsion from pollen via submicronic respirable particles. *J. Immunol.* **172**, 6490–6500.
- Tsai, Y.T., Chen, S.H., Lin, K.L., and Hsieh, K.H. (1990). Rice pollen allergy in Taiwan. *Ann. Allergy* **65**, 459–462.
- Valdivia, E.R., Sampedro, J., Lamb, J.C., Chopra, S., and Cosgrove, D.J. (2007). Recent proliferation and translocation of pollen group 1 allergen genes in the maize genome. *Plant Physiol.* **143**, 1269–81.
- Valenta, R., Duchene, M., Pettenburger, K., Sillaber, C., Valent, P., Bettelheim, P., Breitenbach, M., Rumpold, H., Kraft, D., and Scheiner, O. (1991). Identification of profilin as a novel pollen allergen; IgE autoreactivity in sensitized individuals. *Science.* **253**, 557–560.
- van de Lagemaat, L.N., Gagnier, L., Medstrand, P., and Mager, D.L. (2005). Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res.* **15**, 1243–1249.
- Vitte, C., Panaud, O., and Quesneville, H. (2007). LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics.* **8**, 218.
- Wang, W., et al. (2006). High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell.* **18**, 1791–1802.
- Wang, X., Tang, H., Bowers, J.E., Feltus, F.A., and Paterson, A.H. (2007). Extensive concerted evolution of rice paralogs and the road to regaining independence. *Genetics.* **177**, 1753–1763.
- Wu, Z., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F., and Spencer, F. (2004). A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.* **99**, 909–917.
- Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science.* **319**, 1527–1530.
- Xu, H., Theerakulpisut, P., Goulding, N., Suphioglu, C., Singh, M.B., and Bhalla, P.L. (1995). Cloning, expression and immunological characterization of Ory s 1, the major allergen of rice pollen. *Gene.* **164**, 255–259.
- Xu, H.L., Goulding, N., Zhang, Y., Swoboda, I., Singh, M.B., and Bhalla, P.L. (1999). Promoter region of Ory s 1, the major rice pollen allergen gene. *Sex. Plant Reprod.* **12**, 125–126.

Page 1 of 2



Page 2 of 2

1. *Journal of the American Medical Association*, 2000; 284: 2689-2695.

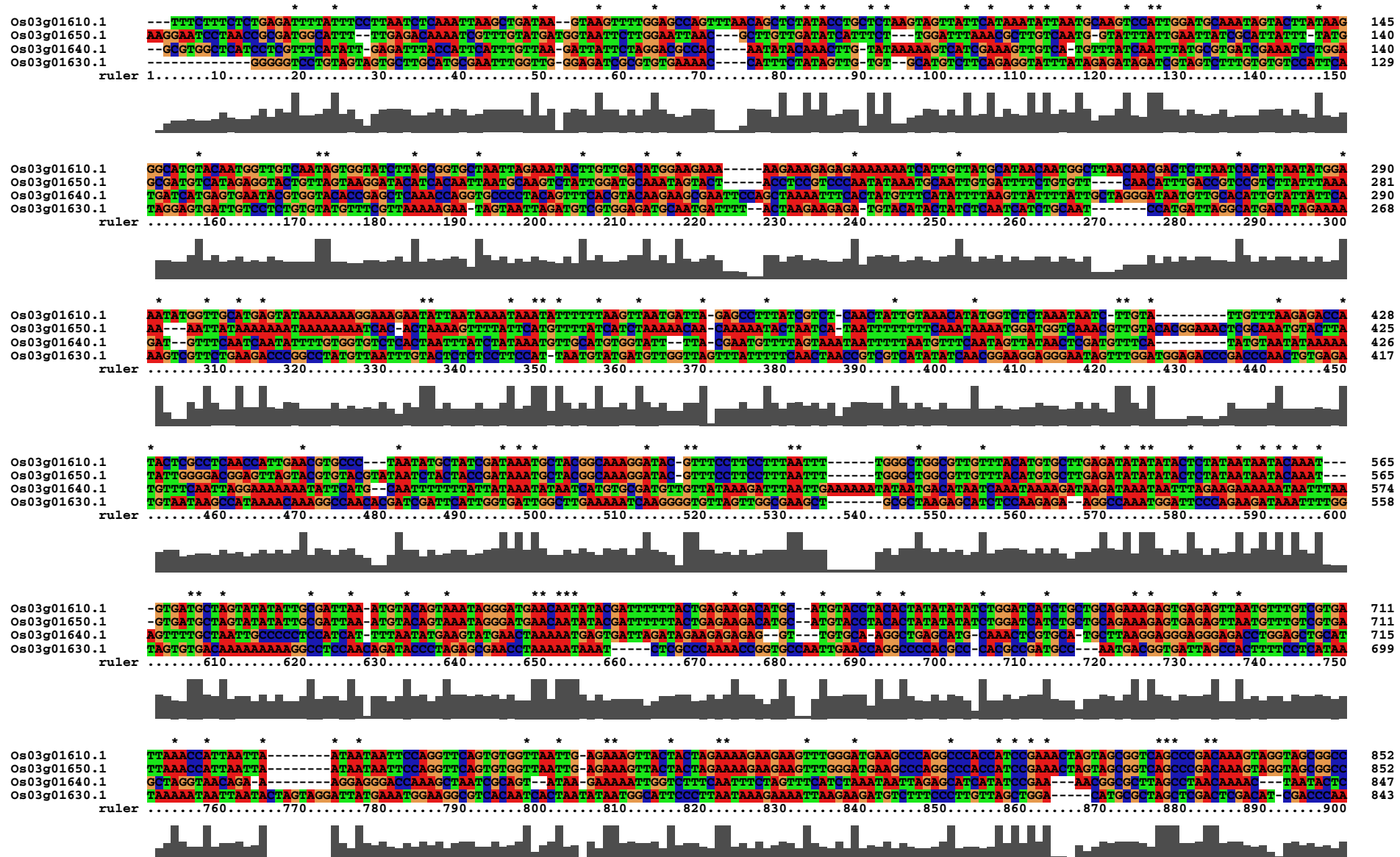
Year	Percentage of Population Aged 65 and Over
1950	7
1960	10
1970	12
1980	14
1990	16
2000	18
2010	19
2020	20
2030	21
2040	22
2050	22

ALIGNMENT OF AMINO ACID SEQUENCE OF RICE ORY S 1 ALLERGEN

[illegible][illegible]

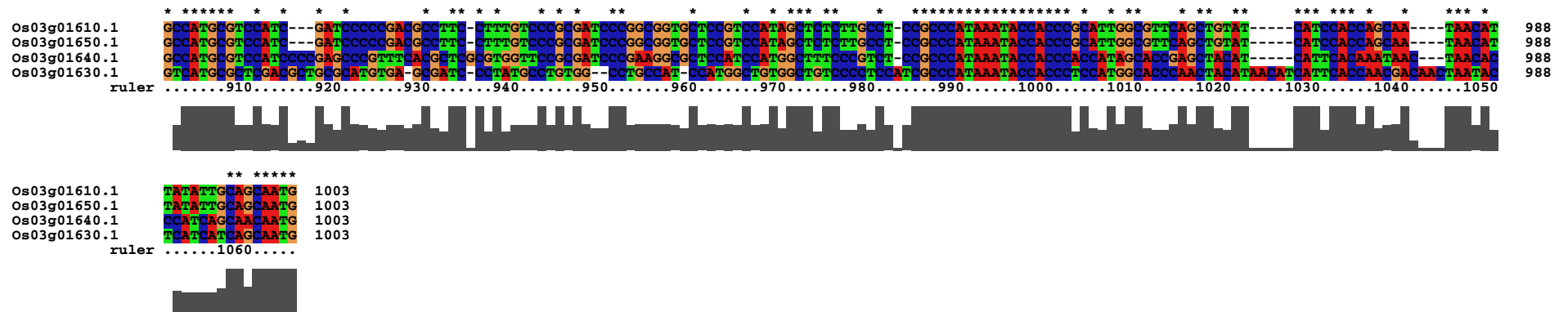
ALIGNMENT OF 1K UPSTREAM SEQUENCE OF ORY S 1

Page 1 of 2

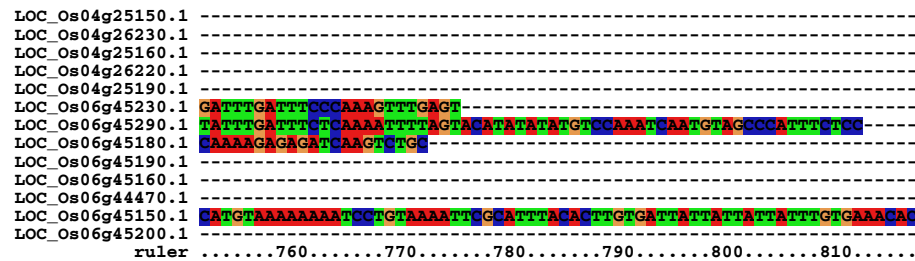
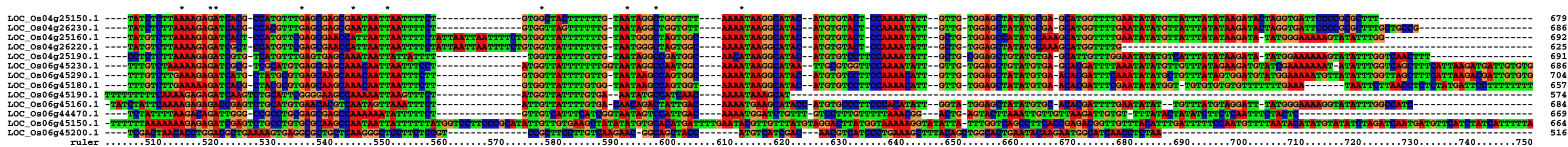
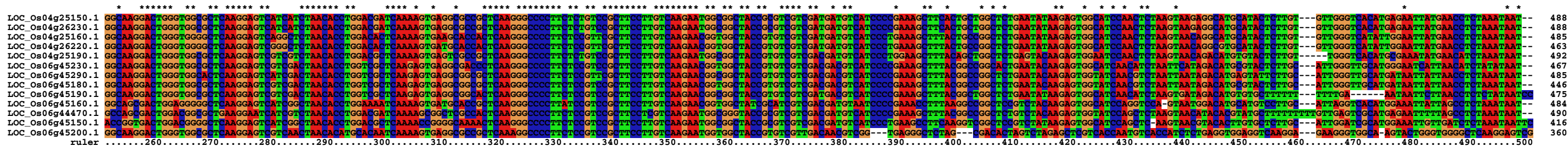
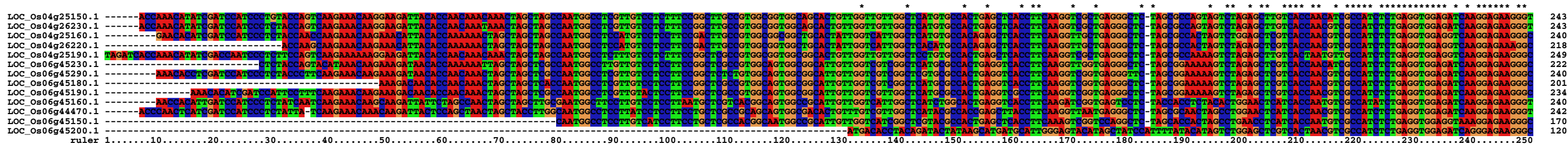


ALIGNMENT OF 1K UPSTREAM SEQUENCE OF ORY S 1

Page 2 of 2



Ory s 2 NUCLEOTIDE ALIGNMENT OF cDNA SEQUENCE

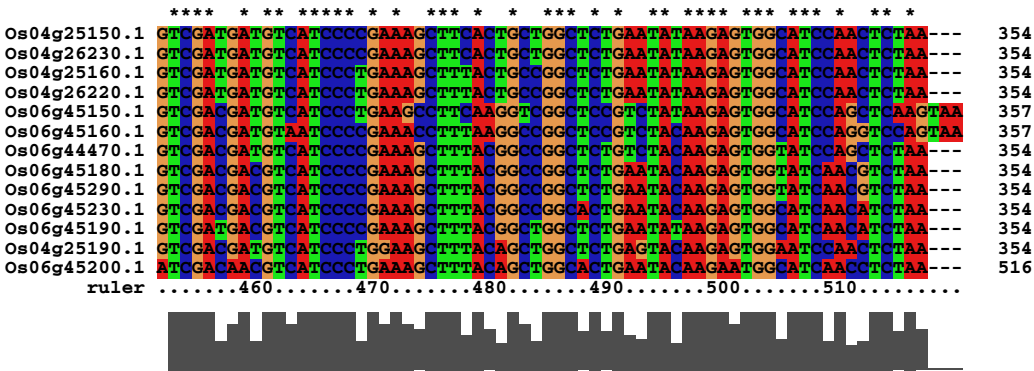


Page 1 of 2

[illegible]

1. The first part of the paper discusses the importance of the research and the need for a new approach to the study of the history of the world. It highlights the challenges faced by historians and the need for a more comprehensive and integrated approach to the study of the world's history.

ALIGNMENT OF PREDICTED AMINO ACID SEQUENCE OF ALLERGEN Ory s 2





250

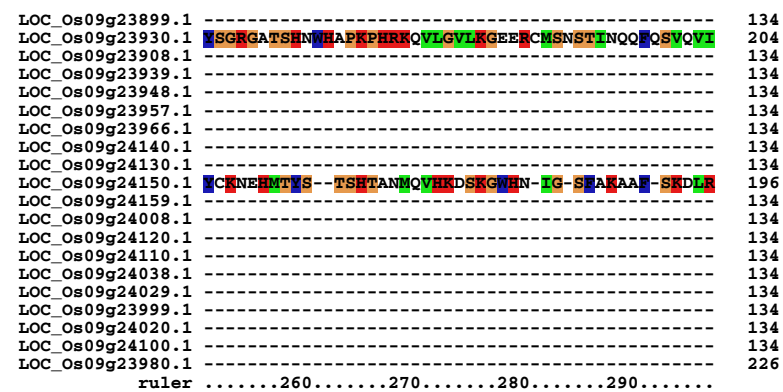
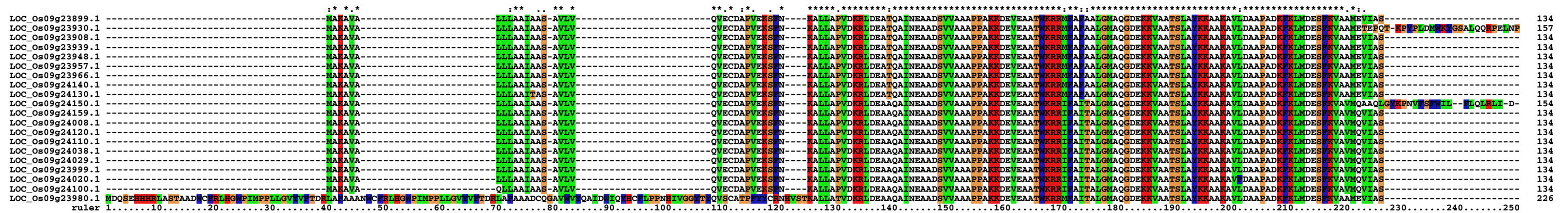
250

223

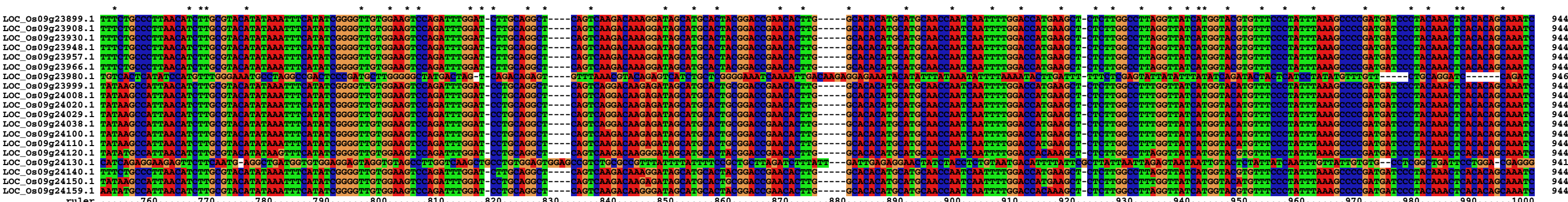
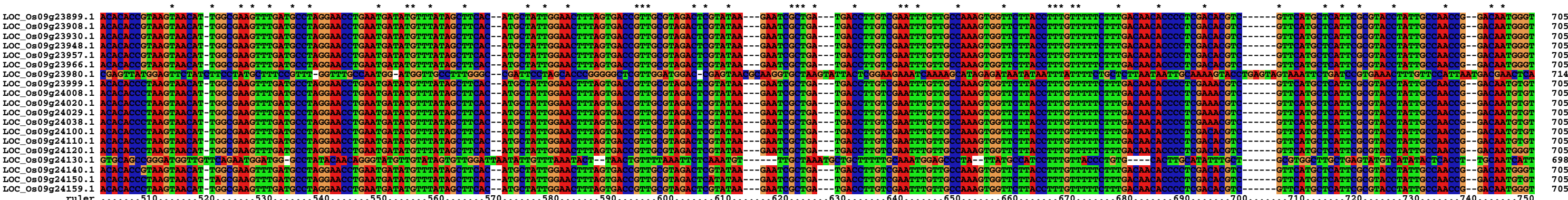
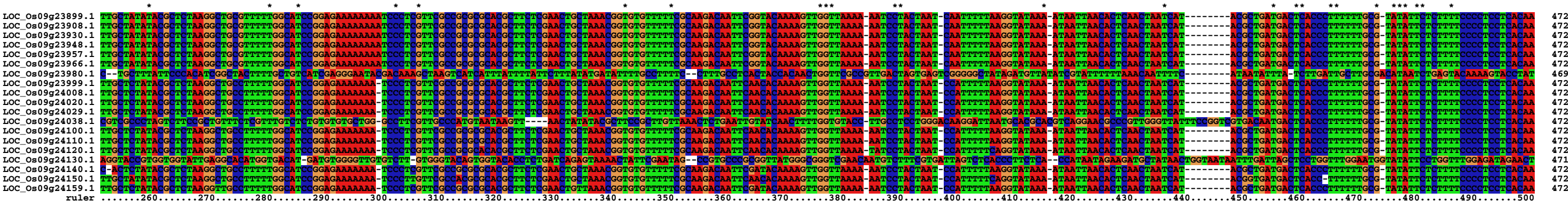
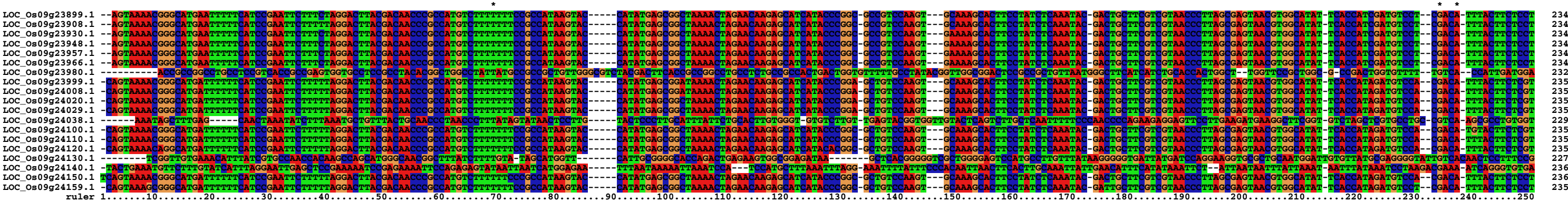
405

AMINO ACID SEQUENCE OF Ory s 23 PROTEIN

Page 1 of 1

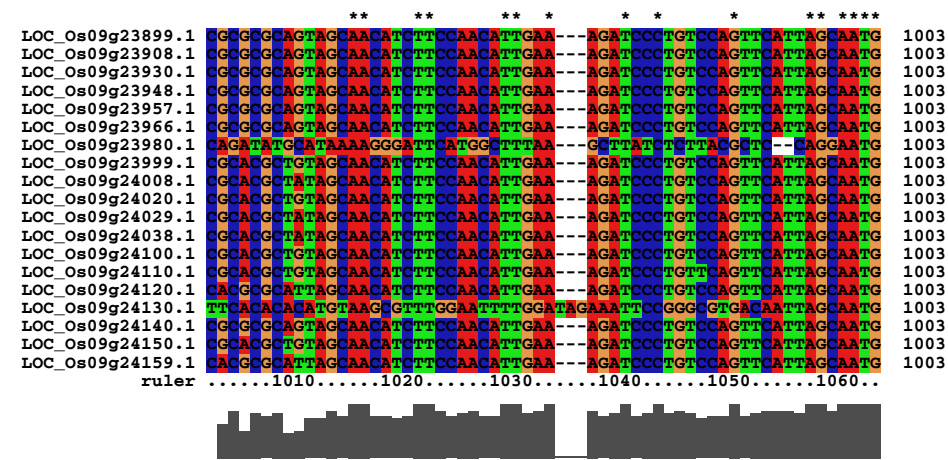


ALIGNMENT OF NUCLEOTIDES IN 1K UPSTREAM REGION of Ory s 23



ALIGNMENT OF NUCLEOTIDES IN 1K UPSTREAM REGION of Ory s 23

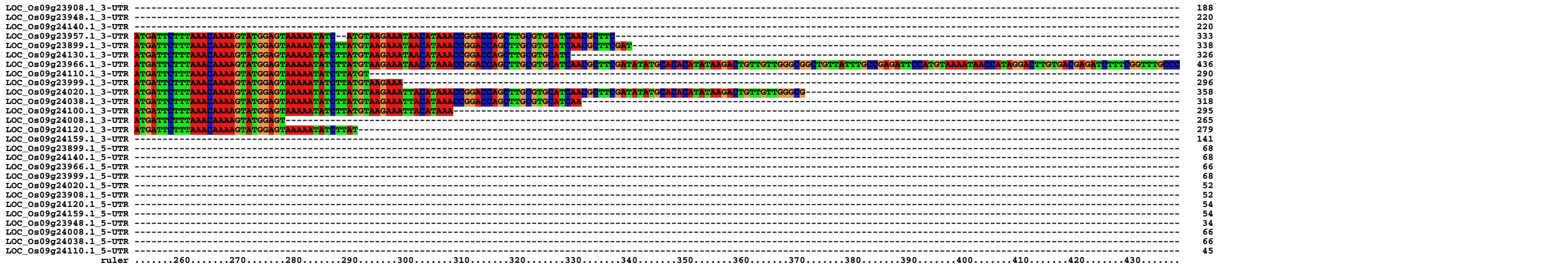
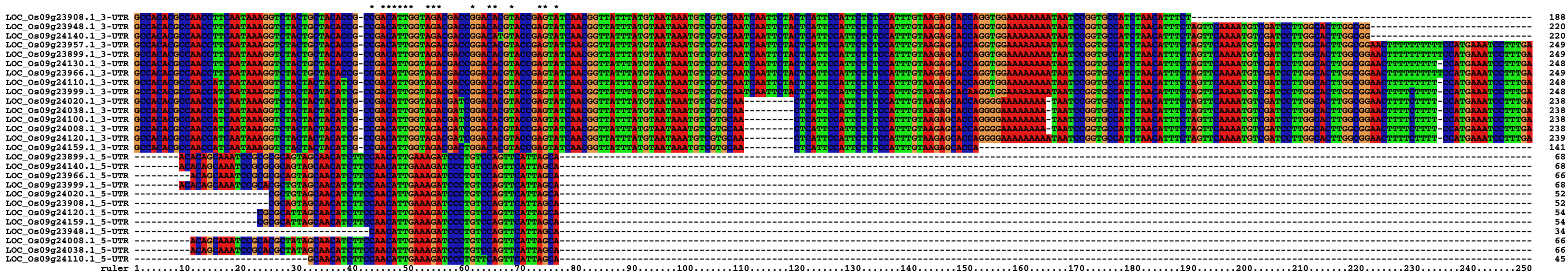
Page 2 of 2



SEQUENCE ALIGNMENT of UTRs of Ory s 23

File: utrs.ps

Date: Fri Feb 22 18:31:53 2008



Russell et al. (2008) Transcriptome-Based Examination of Putative Pollen Allergens o
Supplementary Spreadsheet S1. Transcriptome expression of allergen candidates (

Jiang designation	Affy 57K probe match	candidates	log2 intensity
OsOle1	OsAffx.23143.1.S1_x_at		2,62
OsOle1	OsAffx.23143.4.S1_s_at		2,66
OsOle2	Os.46013.1.S1_at		2,76
OsOle3	OsAffx.23143.3.S1_x_at		2,6
OsOle3	OsAffx.23143.5.S1_at		3,41
OsOle3	OsAffx.23143.5.S1_x_at		3,81
OsOle4	OsAffx.23143.3.S1_at		2,69
OsOle4	OsAffx.23143.3.S1_x_at		2,6
OsOle4	OsAffx.23143.4.S1_s_at		2,66
OsOle5	OsAffx.23143.1.S1_x_at		2,62
OsOle5	OsAffx.23143.3.S1_x_at		2,6
OsOle5	OsAffx.23143.4.S1_s_at		2,66
OsOle6	OsAffx.10949.1.S1_x_at		1,64
OsOle7	NO MATCH		
OsOle8	OsAffx.23143.1.S1_x_at		2,62
OsOle8	OsAffx.23143.3.S1_x_at		2,6
OsOle8	OsAffx.23143.4.S1_s_at		2,66
OsOle9	OsAffx.23143.1.S1_x_at		2,62
OsOle9	OsAffx.23143.3.S1_x_at		2,6
OsOle9	OsAffx.23143.4.S1_s_at		2,66
OsOle10	OsAffx.23143.1.S1_x_at		2,62
OsOle10	OsAffx.23143.4.S1_s_at		2,66
OsOle11	OsAffx.10948.1.S1_at		2,7
OsOle12	Os.28994.1.S1_at		2,54
OsOle13	Os.13994.1.S1_at		2,76
OsOle14	Os.5920.1.S1_at		4,35
OsOle15	OsAffx.24364.1.S1_at		3,99
OsOle16	Os.20298.1.S1_at		3,17
OsOle16	OsAffx.27050.1.A1_at		4,39
OsOle16	OsAffx.27050.1.S1_x_at		2,9
OsOle17	Os.17210.1.S1_at		10,18
OsOle17	Os.54406.1.S1_at		3,94
OsOle17	OsAffx.12988.1.S1_x_at		6,27
OsOle17	OsAffx.12988.2.S1_at		3,29
OsOle18	Os.51228.1.S1_at		6,08
OsOle18	Os.51228.2.A1_at		7,26
OsOle18	Os.51228.3.S1_at		11,06
OsOle18	Os.51228.3.S1_x_at		10,81
OsOle19	OsAffx.4002.1.A1_at		2,2
OsOle19	OsAffx.4002.1.S1_s_at		3,8
OsOle20	Os.280.1.S2_at		4,98
OsOle21	Os.9327.1.S1_at		4,63
OsOle22	OsAffx.27322.1.S1_at		2,08
OsOle22	OsAffx.27322.1.S1_s_at		2,95
OsOle23	OsAffx.14668.1.S1_at		2,38

OsOle24	Os.53412.1.S1_at	Ory s 11	8,86
OsOle25	NO MATCH		
OsOle26	Os.50018.1.S1_at		4,68
OsOle27	Os.55286.1.S1_at		3,17
OsOle28	Os.50339.1.S1_at	C13 precursor	7,38
OsOle29	Os.17295.1.S1_at		6,45
OsOle30	Os.52095.1.S1_at		4,77
OsOle30	Os.52095.2.S1_at		3,07
OsOle31	Os.8638.1.S1_at		7,38
OsOle31	Os.8638.1.S1_s_at		5,02
OsOle32	Os.10678.1.S1_at		6,46
OsOle32	Os.37820.1.S1_x_at		5,87
OsOle33	Os.37820.1.S1_x_at		5,87
OsOle33	Os.8385.2.S1_at		6,55
OsOle33	Os.8385.2.S1_x_at		5,58
OsOle34	Os.46081.1.S1_at		2,91
OsOle35	OsAffx.20155.1.S1_at		2,66
OsOle36	Os.3577.1.S1_x_at		6,81
OsOle36	Os.5297.1.S1_at		5,16
OsOle37	OsAffx.20155.1.S1_at		2,66
OsOle38	Os.39297.1.S1_at		4,02
OsOle38	Os.46709.1.S1_x_at		1,82
OsOle39	Os.37820.1.S1_x_at		5,87
OsOle39	Os.8385.2.S1_at		6,55
OsOle39	Os.8385.2.S1_x_at		5,58
OsOle40	Os.10678.1.S1_at		6,46
OsOle40	Os.37820.1.S1_x_at		5,87
OsOle41	Os.37820.1.S1_x_at		5,87
OsOle41	Os.8385.1.S1_s_at		7,18
OsOle42	Os.8203.1.S1_at		4,57
OsOle43	Os.46488.1.S1_at		3,61
OsOle44	Os.46081.1.S1_at		2,91
OsOle45	Os.7253.1.S1_at		4,83
OsAllerg1-1	OsAffx.22035.1.S1_x_at		2,19
OsAllerg1-2	Os.28773.1.S1_at		3,16
OsAllerg1-3	NO MATCH		
OsAllerg1-4	Os.4644.1.S1_at		5,14
OsAllerg1-5	OsAffx.12459.1.S1_at		4,39
OsAllerg1-6	Os.7602.1.S1_at		5,59
OsAllerg1-7	OsAffx.24716.1.S1_at		2,54
OsAllerg1-8	Os.4613.1.S1_at		4,77
OsAllerg1-9	Os.4613.1.S1_at		4,77
OsAllerg1-10	OsAffx.24284.2.S1_at		2,15
OsAllerg1-10	OsAffx.24284.5.A1_at		2,68
OsAllerg1-11	OsAffx.24284.5.A1_at		2,68
OsAllerg1-11	OsAffx.24284.5.S1_at		2,35
OsAllerg1-12	OsAffx.24284.3.S1_at		2,52
OsAllerg1-12	OsAffx.24284.5.A1_at		2,68
OsAllerg1-13	OsAffx.24284.4.S1_at		2,4
OsAllerg1-13	OsAffx.24284.4.S1_x_at		2,37

OsAllerg1-13	OsAffx.24284.5.A1_at		2,68
OsAllerg1-14	OsAffx.2643.1.S1_at		2,01
OsAllerg1-14	OsAffx.2643.1.S1_x_at		1,97
OsAllerg1-15	OsAffx.24284.4.S1_at		2,4
OsAllerg1-15	OsAffx.24284.4.S1_x_at		2,37
OsAllerg1-15	OsAffx.24284.5.A1_at		2,68
OsAllerg1-16	Os.21349.1.S1_at		4,73
OsAllerg1-17	Os.20535.1.S1_at		3,21
OsAllerg1-18	OsAffx.3144.1.S1_at		2,5
OsAllerg1-18	OsAffx.3144.1.S1_x_at		2,38
OsAllerg1-19	NO MATCH		
OsAllerg1-20	OsAffx.12748.2.S1_at		2,07
OsAllerg1-21	OsAffx.12748.2.S1_at		2,07
OsAllerg1-22	OsAffx.24987.1.S1_at		2,67
OsAllerg1-23	Os.57306.1.S1_at		2,19
OsAllerg1-24	NO MATCH		
OsAllerg1-25	Os.2368.1.S1_at		5,59
OsAllerg1-26	Os.2402.1.S1_at	Ory s 1	9,81
OsAllerg1-27	Os.2615.1.S1_at		2,43
OsAllerg1-28	Os.9312.1.S1_at		4,52
OsAllerg1-29	Os.2402.1.S1_at	Ory s 1	9,81
OsAllerg1-30	Os.2405.1.S1_at	Ory s 1	7,81
OsAllerg1-31	NO MATCH		
OsAllerg1-32	Os.6436.1.S1_at		3,85
OsAllerg1-33	Os.54957.1.S1_at		3,99
OsAllerg1-33	OsAffx.4092.1.S1_s_at		3,68
OsAllerg1-34	Os.2403.1.S1_at		2,27
OsAllerg1-35	NO MATCH		
OsAllerg1-36	Os.12319.1.S1_x_at	Ory s 2*	9,33
OsAllerg1-37	Os.9665.1.S1_at	Ory s 2	8,22
OsAllerg1-38	Os.2369.1.S1_a_at		7,61
OsAllerg1-39	Os.2696.1.S1_at		3,85
OsAllerg1-40	Os.54957.1.S1_at		3,99
OsAllerg1-40	OsAffx.4092.1.S1_s_at		3,68
OsAllerg1-41	Os.54019.1.S1_at		3,2
OsAllerg1-42	Os.410.1.S1_at		3,45
OsAllerg1-43	NO MATCH		
OsAllerg1-44	OsAffx.28114.2.A1_at	Ory s 2**	5,41
OsAllerg1-45	Os.54103.1.S1_at		3,4
OsAllerg1-46	Os.22908.1.S1_at		2,75
OsAllerg1-47	Os.47341.1.A1_at		3,19
OsAllerg1-48	Os.54103.1.S1_at		3,4
OsAllerg1-49	OsAffx.22609.1.S1_x_at		2,62
OsAllerg1-50	OsAffx.22303.1.S1_at		2,13
OsAllerg1-51	Os.5937.1.S1_at		3,06
OsAllerg1-52	Os.53024.1.S1_at	expansin	7,96
OsAllerg1-53	Os.12697.1.S1_at	expansin	9,69
OsAllerg1-54	Os.2373.1.S1_at		3,96
OsAllerg1-55	Os.6786.1.S1_a_at		6,07
OsAllerg1-56	Os.12234.1.S1_s_at		7,36

OsAllerg1-57	Os.2404.1.S1_at	4,46
OsAllerg1-58	OsAffx.18574.1.S1_x_at	2,3
OsAllerg1-59	NO MATCH	
OsAllerg1-60	OsAffx.8085.1.S1_at	2,44
OsAllerg1-60	OsAffx.8085.1.S1_x_at	2,48
OsAllerg1-61	Os.2938.1.S1_at	5,11
OsAllerg1-61	Os.2938.1.S1_x_at	5,31
OsAllerg1-62	Os.20191.1.S1_at	8,29
OsAllerg2-1	Os.1437.1.S1_at	7,56
OsAllerg2-2	OsAffx.24073.1.S1_at	4,57
OsAllerg2-3	Os.49442.2.S1_x_at	6,88
OsAllerg2-4	OsAffx.26038.1.S1_at	2,71
OsAllerg2-5	Os.23745.1.S1_at	5,51
OsAllerg2-6	Os.54631.1.A1_at	3,4
OsAllerg2-7	NO MATCH	

Jiang, S.Y., Jasmin, P.X., Ting, Y.Y., and Ramachandran, S. (2005). Genome-wide

*This is Ory s 2-B according to Kerim et al., 2003b

**This is Ory s 2-A according to Kerim et al., 2003b

of Rice (*Oryza sativa* ssp. *japonica*). Mol. Plant
described by Jiang, S.Y.(2005). This table relates each candidate to one or more 57k

log2-fold over seedling	p-value	TIGR v5 best match	IRGSP locus
0,28	2,94E-02	OS01G06960.1	
0,12	8,89E-03	OS01G06960.1	
0,17	6,73E-04	OS01G07040.1	
0,12	2,86E-02	OS01G06940.1	
1,5	2,25E-04	OS01G06940.1	
1,32	1,72E-03	OS01G06940.1	
0,08	1,84E-02	OS01G06980.1	
0,12	2,86E-02	OS01G06980.1	
0,12	8,89E-03	OS01G06980.1	
0,28	2,94E-02	OS01G07020.1	
0,12	2,86E-02	OS01G07020.1	
0,12	8,89E-03	OS01G07020.1	
0,01	4,52E-01	OS01G07060.1	
		OS01G07050.1	
0,28	2,94E-02	OS01G07020.1	
0,12	2,86E-02	OS01G07020.1	
0,12	8,89E-03	OS01G07020.1	
0,28	2,94E-02	OS01G06990.1	
0,12	2,86E-02	OS01G06990.1	
0,12	8,89E-03	OS01G06990.1	
0,28	2,94E-02	OS01G07010.1	
0,12	8,89E-03	OS01G07010.1	
0,16	3,37E-01	OS01G07030.1	
-0,17	3,77E-01	OS01G67390.1	Os01g0899700
0,01	7,97E-01	OS01G52650.1	Os01g0725900
-0,38	4,84E-02	OS01G52660.1	Os06g0507200
-2,71	2,59E-08	OS02G21280.1	Os02g0316900
0,1	6,45E-01	OS02G01190.1	Os02g0101900
2,12	1,12E-04	OS02G01190.1	
0,46	9,16E-02	OS02G01190.1	
3,4	1,02E-08	OS03G22250.1	Os02g0823300
-0,04	6,99E-01	OS03G22250.1	Os06g0139900
5,18	1,99E-08	OS03G22250.1	Os06g0613600
0,87	7,87E-03	OS03G22250.1	
5,02	1,56E-07	OS03G14130.1	Os04g0345400
-4,99	1,34E-09	OS03G14130.1	Os04g0345400
2,82	2,38E-05	OS03G14130.1	Os03g0245100
1,85	8,16E-05	OS03G14130.1	Os03g0245200
0,78	4,31E-02	OS04G35580.1	
-2,39	8,46E-07	OS04G35580.1	
0,43	9,32E-02	OS04G32680.1	Os04g0398600
-6,48	7,12E-08	OS05G45460.1	Os09g0453000
0,02	4,26E-01	OS05G45480.1	
-0,49	5,14E-02	OS05G45480.1	Os05g0530800
0,09	1,71E-03	OS05G13340.1	

12,29	3,18E-12	OS06G36240.1	Os05g0466200
		OS07G40100.1	
-7,58	2,42E-10	OS07G47750.1	Os05g0415400
-2,94	4,61E-07	OS07G47760.1	Os05g0164200
10,89	4,68E-08	OS09G39950.1	Os04g0413200
-7,81	4,27E-08	OS09G33450.2	Os09g0508300
0,08	7,37E-01	OS09G32988.1	
-0,91	1,26E-02	OS09G32988.1	
-8,68	7,25E-08	OS10G05950.1	Os07g0639600
-5,52	1,46E-08	OS10G05950.1	
-9,91	1,23E-11	OS10G05990.1	Os10g0150700
-10,44	3,80E-11	OS10G05990.1	Os10g0150400
-10,44	3,80E-11	OS10G06000.1	Os10g0150400
-10,01	3,70E-12	OS10G06000.1	Os05g0556700
-9,27	8,14E-12	OS10G06000.1	Os10g0150400
0,15	3,03E-03	OS10G13850.1	Os10g0205700
0,3	6,40E-02	OS10G39890.1	Os10g0550300
-4,38	8,15E-07	OS10G22590.1	Os06g0138700
-8,23	2,55E-11	OS10G22590.1	Os06g0248300
0,3	6,40E-02	OS10G39890.1	Os10g0550300
-2,99	1,16E-07	OS10G13914.1	
0,01	5,23E-01	OS10G13914.1	
-10,44	3,80E-11	OS10G06000.1	Os10g0150400
-10,01	3,70E-12	OS10G06000.1	Os05g0556700
-9,27	8,14E-12	OS10G06000.1	Os10g0150400
-9,91	1,23E-11	OS10G05990.1	Os10g0150700
-10,44	3,80E-11	OS10G05990.1	Os10g0150400
-10,44	3,80E-11	OS10G05970.1	Os10g0150400
-11,23	1,37E-11	OS10G05970.1	Os02g0647900
-6,65	2,07E-12	OS10G05750.1	Os12g0104700
0,54	1,03E-02	OS10G05790.1	Os10g0148700
0,15	3,03E-03	OS10G13850.1	Os10g0205700
-6,6	4,35E-10	OS12G28770.1	Os06g0470000
0,13	7,08E-02	OS01G14660.1	
0,08	1,24E-02	OS01G14650.1	Os01g0248900
		OS01G16770.1	
-5,53	4,27E-08	OS01G60770.1	Os01g0823100
0,72	1,27E-01	OS02G42650.1	
-6,24	9,48E-08	OS02G44108.1	Os08g0512700
0,18	9,97E-03	OS02G44106.1	
-4,47	5,41E-06	OS02G51040.1	Os02g0744200
-4,47	5,41E-06	OS02G51040.1	Os02g0744200
0,07	1,07E-02	OS02G16850.1	
0,19	2,64E-04	OS02G16850.1	Os02g0267700
0,19	2,64E-04	OS02G16839.1	Os02g0267700
0,08	4,22E-03	OS02G16839.1	Os02g0267700
-0,08	4,10E-01	OS02G16800.1	
0,19	2,64E-04	OS02G16800.1	Os02g0267700
0,14	5,06E-04	OS02G16780.1	Os02g0268600
0,1	2,61E-03	OS02G16780.1	

0,19	2,64E-04	OS02G16780.1	Os02g0267700
0,04	4,79E-02	OS02G16730.1	
0,02	2,47E-01	OS02G16730.1	
0,14	5,06E-04	OS02G16780.1	Os02g0268600
0,1	2,61E-03	OS02G16780.1	
0,19	2,64E-04	OS02G16780.1	Os02g0267700
-7,77	1,14E-07	OS03G04020.1	Os03g0132200
-0,55	1,75E-01	OS03G06000.1	Os03g0155300
0,08	4,32E-02	OS03G06010.1	
0,05	1,87E-01	OS03G06010.1	
		OS03G06040.1	
0,02	2,93E-01	OS03G06050.1	
0,02	2,93E-01	OS03G06060.1	
0,04	1,92E-01	OS03G06020.1	
0,07	1,33E-02	OS03G31480.1	Os12g0545900
		OS03G25990.1	
-4,22	6,53E-07	OS03G60720.1	Os03g0822000
13,19	5,29E-08	OS03G01650.1	Os03g0106500
-0,96	4,84E-05	OS03G01260.1	Os03g0102500
-6,42	1,77E-10	OS03G01270.2	Os08g0412200
13,19	5,29E-08	OS03G01650.1	Os03g0106500
12,45	4,49E-08	OS03G01640.1	Os03g0106800
		OS03G43130.1	
-4,17	4,70E-09	OS03G44290.2	Os03g0136200
-5,71	1,10E-10	OS04G44780.1	Os05g0518200
-0,29	4,07E-02	OS04G44780.1	
0,03	1,76E-01	OS04G46650.1	Os04g0552200
		OS04G46630.1	
11,3	2,02E-10	OS04G26230.1	
12,43	1,47E-10	OS04G25190.1	Os03g0139500
-4,54	4,74E-07	OS04G49410.1	Os04g0583500
-3,72	3,73E-08	OS04G15840.1	Os04g0228400
-5,71	1,10E-10	OS04G44780.2	Os05g0518200
-0,29	4,07E-02	OS04G44780.2	
0,07	7,61E-02	OS05G15690.1	Os07g0456900
-2,2	4,70E-04	OS05G19570.1	Os05g0276500
		OS08G22420.1	
6,08	6,24E-08	OS06G45290.1	
-2,58	7,74E-08	OS06G41700.1	Os08g0288400
-1,31	9,35E-06	OS06G50400.1	Os06g0718100
-1,98	2,18E-05	OS06G50960.1	Os05g0389600
-2,58	7,74E-08	OS06G41700.1	Os08g0288400
0,02	8,32E-01	OS06G01920.1	
0,05	6,29E-02	OS07G31390.1	
-1,24	1,80E-06	OS07G29290.1	
12,21	2,46E-10	OS08G44790.1	
13,17	9,62E-10	OS10G40090.1	Os10g0548600
-4,19	1,62E-10	OS10G40710.1	Os10g0555700
-8,44	8,23E-10	OS10G40720.1	Os07g0101300
-8,34	7,51E-08	OS10G40730.1	Os10g0556100

-6,04	2,82E-08	OS10G40700.1	Os10g0555600
0,49	1,28E-02	OS10G39110.1	
		OS10G30330.1	
0,05	8,02E-02	OS10G30340.1	
0,35	1,89E-01	OS10G30340.1	
-2,72	7,56E-04	OS10G39640.1	Os10g0542400
-2,64	1,28E-03	OS10G39640.1	Os10g0542400
9,91	2,31E-10	OS12G36040.1	Os12g0546800
-3,9	5,71E-06	OS01G06580.1	Os01g0159200
-2,35	1,51E-05	OS02G02850.1	
1,19	4,90E-03	OS03G44760.2	Os05g0387800
0,34	1,58E-03	OS04G15510.1	
-0,32	2,48E-01	OS04G13220.1	Os04g0209300
-1,67	2,28E-07	OS07G05420.1	Os02g0749300
		OS10G17020.1	

e identification and molecular characterization of Ole_e_I, Allerg_1 and Allerg_2

◁ microarray probe set, shows matches, TIGR and IGRSP loci and AGI homologs.

TIGR v5 corresponding description	# locus matches
proline-rich protein 1, putative, expressed	5
proline-rich protein 1, putative, expressed	6
proline-rich protein, putative	1
expressed protein	4
expressed protein	1
expressed protein	1
conserved hypothetical protein	1
conserved hypothetical protein	4
conserved hypothetical protein	6
arabinogalactan protein, putative	5
arabinogalactan protein, putative	4
arabinogalactan protein, putative	6
tyrosine-rich hydroxyproline-rich glycoprotein, put...	1
proline-rich protein, putative, expressed	
arabinogalactan protein, putative	5
arabinogalactan protein, putative	4
arabinogalactan protein, putative	6
pollen proteins Ole e I family protein	5
pollen proteins Ole e I family protein	4
pollen proteins Ole e I family protein	6
proline-rich protein 1, putative	5
proline-rich protein 1, putative	6
proline-rich protein, putative, expressed	1
extensin-1 precursor, putative, expressed	1
pollen proteins Ole e I family protein, expressed	2
pistil-specific extensin-like protein precursor, pu...	1
expressed protein	1
expressed protein	1
expressed protein	3
expressed protein	3
expressed protein	3
expressed protein	4
expressed protein	4
expressed protein	1
proline-rich protein, putative, expressed	2
proline-rich protein, putative, expressed	2
proline-rich protein, putative, expressed	2
proline-rich protein, putative, expressed	2
expressed protein	1
expressed protein	1
pollen-specific protein C13 precursor, putative, expressed	2
pistil-specific extensin-like protein precursor, pu...	1
arabinogalactan protein, putative, expressed	1
arabinogalactan protein, putative, expressed	1
conserved hypothetical protein	1

pollen allergen Phl p 11, putative, expressed	1
pollen-specific protein C13 precursor, putative, expressed	
expressed protein	1
phyloplanin precursor, putative, expressed	1
pollen-specific protein C13 precursor, putative, expressed	1
palmitoyltransferase ZDHHC9, putative, expressed	1
pollen proteins Ole e I family protein, expressed	1
pollen proteins Ole e I family protein, expressed	1
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	2
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	3
proline-rich protein, putative, expressed	3
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	1
expressed protein	1
pollen proteins Ole e I family protein, expressed	1
pollen-specific protein C13 precursor, putative, expressed	4
pollen-specific protein C13 precursor, putative, expressed	1
pollen proteins Ole e I family protein, expressed	1
expressed protein	1
expressed protein	1
proline-rich protein, putative, expressed	3
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	3
proline-rich protein, putative, expressed	3
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	1
proline-rich protein, putative, expressed	1
expressed protein	1
expressed protein	1
alpha-expansin 4 precursor, putative, expressed	1
alpha-expansin 15 precursor, putative, expressed	1
alpha-expansin 10 precursor, putative, expressed	
alpha-expansin 10 precursor, putative, expressed	1
beta-expansin 3 precursor, putative, expressed	1
beta-expansin 4 precursor, putative, expressed	1
beta-expansin 4 precursor, putative	1
alpha-expansin 15 precursor, putative, expressed	1
alpha-expansin 15 precursor, putative, expressed	1
alpha-expansin 3 precursor, putative, expressed	1
alpha-expansin 3 precursor, putative, expressed	5
alpha-expansin 3 precursor, putative	5
alpha-expansin 3 precursor, putative	2
alpha-expansin 10 precursor, putative, expressed	1
alpha-expansin 10 precursor, putative, expressed	5
alpha-expansin 1 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	1

alpha-expansin 1 precursor, putative, expressed	5
alpha-expansin 1 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	5
expansin-like 3 precursor, putative, expressed	1
alpha-expansin 17 precursor, putative, expressed	1
alpha-expansin 17 precursor, putative, expressed	1
alpha-expansin 17 precursor, putative, expressed	1
alpha-expansin 15 precursor, putative, expressed	
alpha-expansin 15 precursor, putative, expressed	2
alpha-expansin 15 precursor, putative, expressed	2
alpha-expansin 17 precursor, putative, expressed	1
alpha-expansin 15 precursor, putative, expressed	1
alpha-expansin 15 precursor, putative, expressed	
alpha-expansin 6 precursor, putative, expressed	1
major pollen allergen Ory s 1 precursor, putative, ...	3
beta-expansin 1a precursor, putative, expressed	1
beta-expansin 1a precursor, putative, expressed	1
major pollen allergen Ory s 1 precursor, putative, ...	3
major pollen allergen Ory s 1 precursor, putative, ...	1
F-box domain containing protein	
beta-expansin 4 precursor, putative, expressed	1
rare lipoprotein A like double-psi beta-barrel cont...	1
rare lipoprotein A like double-psi beta-barrel cont...	1
beta-expansin 2 precursor, putative, expressed	1
beta-expansin 2 precursor, putative, expressed	
pollen allergen Phl p 2 precursor, putative, expressed	2
pollen allergen Phl p 2 precursor, putative, expressed	2
alpha-expansin 13 precursor, putative, expressed	2
alpha-expansin 1 precursor, putative, expressed	1
rare lipoprotein A like double-psi beta-barrel cont...	1
rare lipoprotein A like double-psi beta-barrel cont...	1
beta-expansin 2 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	1
retrotransposon protein, putative, unclassified	
pollen allergen Phl p 2 precursor, putative, expressed	2
alpha-expansin 20 precursor, putative, expressed	1
alpha-expansin 1 precursor, putative, expressed	1
expansin-like 3 precursor, putative, expressed	1
alpha-expansin 20 precursor, putative, expressed	1
alpha-expansin 7 precursor, putative, expressed	1
expansin-related protein 1 precursor, putative, exp...	1
expansin-like 3 precursor, putative, expressed	1
alpha-expansin 3 precursor, putative, expressed	1
beta-expansin 1a precursor, putative, expressed	1
beta-expansin 1a precursor, putative, expressed	1
beta-expansin 1a precursor, putative, expressed	1
beta-expansin 1a precursor, putative, expressed	1

beta-expansin 1a precursor, putative, expressed	1
alpha-expansin 7 precursor, putative	1
alpha-expansin 1 precursor, putative, expressed	
alpha-expansin 17 precursor, putative, expressed	1
alpha-expansin 17 precursor, putative, expressed	1
expansin-like 3 precursor, putative, expressed	1
expansin-like 3 precursor, putative, expressed	1
alpha-expansin 9 precursor, putative, expressed	1
fasciclin-like arabinogalactan protein 7 precursor,...	1
C-1-tetrahydrofolate synthase, cytoplasmic, putativ...	1
SWI1, putative, expressed	1
retrotransposon protein, putative, unclassified	1
multidrug resistance-associated protein 4, putative...	1
anthocyanidin 3-O-glucosyltransferase, putative, expressed	1
retrotransposon protein, putative, unclassified	

2 domain-containing pollen-allergen-like genes in *Oryza sativa*. DNA Res 12, 167-179.

[illegible]

Pollen allergen Phl p 11, putative, expressed	LOC_Os06g36240.1 (11)
Expressed protein	LOC_Os07g47750.1 (11)
Phylloplanin precursor, putative, expressed	LOC_Os07g47760.1 (11)
Pollen-specific protein C13 precursor, putative, expressed	LOC_Os09g39950.1 (11)
	LOC_Os09g33450.1 (11); LOC_Os09g33450.2 (11)
Pollen proteins Ole e I family protein, expressed	LOC_Os09g32988.1 (11)
Pollen proteins Ole e I family protein, expressed	LOC_Os09g32988.1 (9)
Proline-rich protein, putative, expressed	LOC_Os10g05950.1 (11)
[LOC_Os07g30154, Expressed protein],[LOC_Os10g05950.1 (11); LOC_Os10g05950.1 (11)	LOC_Os07g30154.1 (1); LOC_Os10g05950.1 (11)
	LOC_Os10g05990.1 (11); LOC_Os10g05990.2 (11)
[LOC_Os10g05970, Proline-rich protein, putative, expressed	LOC_Os10g05970.1 (11); LOC_Os10g05970.2 (11);
[LOC_Os10g05970, Proline-rich protein, putative, expressed	LOC_Os10g05970.1 (11); LOC_Os10g05970.2 (11);
Proline-rich protein, putative, expressed	LOC_Os10g06000.1 (11)
Proline-rich protein, putative, expressed	LOC_Os10g06000.1 (11)
Expressed protein	LOC_Os10g13850.1 (11)
Pollen proteins Ole e I family protein, expressed	LOC_Os10g39890.1 (11)
[LOC_Os06g04690, F-box domain containing protein, expressed	LOC_Os06g04690.1 (8); LOC_Os06g04710.1 (7); LOC_Os06g04710.2 (7);
Pollen-specific protein C13 precursor, putative, expressed	LOC_Os10g22590.1 (11)
Pollen proteins Ole e I family protein, expressed	LOC_Os10g39890.1 (11)
Expressed protein	LOC_Os10g13914.1 (11)
Expressed protein	LOC_Os10g13914.1 (11)
[LOC_Os10g05970, Proline-rich protein, putative, expressed	LOC_Os10g05970.1 (11); LOC_Os10g05970.2 (11);
Proline-rich protein, putative, expressed	LOC_Os10g06000.1 (11)
Proline-rich protein, putative, expressed	LOC_Os10g06000.1 (11)
	LOC_Os10g05990.1 (11); LOC_Os10g05990.2 (11)
[LOC_Os10g05970, Proline-rich protein, putative, expressed	LOC_Os10g05970.1 (11); LOC_Os10g05970.2 (11);
[LOC_Os10g05970, Proline-rich protein, putative, expressed	LOC_Os10g05970.1 (11); LOC_Os10g05970.2 (11);
	LOC_Os10g05970.1 (11); LOC_Os10g05970.2 (11)
Proline-rich protein, putative, expressed	LOC_Os10g05750.1 (11)
Proline-rich protein, putative, expressed	LOC_Os10g05790.1 (11)
Expressed protein	LOC_Os10g13850.1 (11)
Expressed protein	LOC_Os12g28770.1 (11)
Alpha-expansin 4 precursor, putative, expressed	LOC_Os01g14660.1 (11)
Alpha-expansin 15 precursor, putative, expressed	LOC_Os01g14650.1 (11)
Alpha-expansin 10 precursor, putative, expressed	LOC_Os01g60770.1 (10)
Beta-expansin 3 precursor, putative, expressed	LOC_Os02g42650.1 (11)
Beta-expansin 4 precursor, putative, expressed	LOC_Os02g44108.1 (11)
Beta-expansin 4 precursor, putative	LOC_Os02g44106.1 (11)
Alpha-expansin 15 precursor, putative, expressed	LOC_Os02g51040.1 (11)
Alpha-expansin 15 precursor, putative, expressed	LOC_Os02g51040.1 (11)
Alpha-expansin 3 precursor, putative, expressed	LOC_Os02g16850.1 (11)
[LOC_Os02g16780, Alpha-expansin 1 precursor, putative, expressed	LOC_Os02g16780.1 (2); LOC_Os02g16800.1 (1); LOC_Os02g16800.2 (1);
[LOC_Os02g16780, Alpha-expansin 1 precursor, putative, expressed	LOC_Os02g16780.1 (2); LOC_Os02g16800.1 (1); LOC_Os02g16800.2 (1);
[LOC_Os02g16809, Alpha-expansin 3 precursor, putative, expressed	LOC_Os02g16809.1 (8); LOC_Os02g16839.1 (8)
Alpha-expansin 10 precursor, putative, expressed	LOC_Os02g16800.1 (11)
[LOC_Os02g16780, Alpha-expansin 1 precursor, putative, expressed	LOC_Os02g16780.1 (2); LOC_Os02g16800.1 (1); LOC_Os02g16800.2 (1);
Alpha-expansin 1 precursor, putative, expressed	LOC_Os02g16780.1 (11)
Alpha-expansin 1 precursor, putative, expressed	LOC_Os02g16780.1 (11)

[LOC_Os02g16780, Alpha-expansin 1 precursor, putative, LOC_Os02g16780.1 (2); LOC_Os02g16800.1 (1); LOC_Os02g16730.1 (11); LOC_Os02g16730.1 (11); LOC_Os02g16780.1 (11); LOC_Os02g16780.1 (11); LOC_Os02g16780.1 (11); LOC_Os02g16780.1 (2); LOC_Os02g16800.1 (1); LOC_Os03g04020.1 (11); LOC_Os03g04020.2 (11); LOC_Os03g06000.1 (11); LOC_Os03g06010.1 (11); LOC_Os03g06010.1 (11); LOC_Os03g06050, Alpha-expansin 15 precursor, putative, LOC_Os03g06050.1 (5); LOC_Os03g06060.1 (11); LOC_Os03g06050, Alpha-expansin 15 precursor, putative, LOC_Os03g06050.1 (5); LOC_Os03g06060.1 (11); LOC_Os03g06020.1 (11); LOC_Os03g31480.1 (11); LOC_Os03g60720.1 (11); LOC_Os03g01610, Major pollen allergen Ory s 1 precursor, putative, LOC_Os03g01610.1 (11); LOC_Os03g01630.1 (3); LOC_Os03g01260.1 (11); LOC_Os03g01270.1 (11); LOC_Os03g01270.2 (10); LOC_Os03g01610, Major pollen allergen Ory s 1 precursor, putative, LOC_Os03g01610.1 (11); LOC_Os03g01630.1 (3); LOC_Os03g01640.1 (11); LOC_Os03g44290.1 (11); LOC_Os03g44290.2 (11); LOC_Os04g44780.1 (11); LOC_Os04g44780.2 (11); LOC_Os04g44780.1 (11); LOC_Os04g44780.2 (11); LOC_Os04g46650.1 (10); LOC_Os04g25150, Pollen allergen Phl p 2 precursor, putative, LOC_Os04g25150.1 (9); LOC_Os04g26230.1 (11); LOC_Os04g25190, Pollen allergen Phl p 2 precursor, putative, LOC_Os04g25190.1 (11); LOC_Os04g26190.1 (5); LOC_Os04g49410, Alpha-expansin 13 precursor, putative, LOC_Os04g49410.1 (11); LOC_Os04g49420.1 (11); LOC_Os04g15840.1 (11); LOC_Os04g44780.1 (11); LOC_Os04g44780.2 (11); LOC_Os04g44780.1 (11); LOC_Os04g44780.2 (11); LOC_Os05g15690.1 (11); LOC_Os05g19570.1 (10); LOC_Os06g45180, Pollen allergen Phl p 2 precursor, putative, LOC_Os06g45180.1 (3); LOC_Os06g45290.1 (11); LOC_Os06g41700.1 (11); LOC_Os06g50400.1 (11); LOC_Os06g50960.1 (11); LOC_Os06g41700.1 (11); LOC_Os06g01920.1 (11); LOC_Os07g31390.1 (11); LOC_Os07g29290.1 (10); LOC_Os08g44790.1 (11); LOC_Os10g40090.1 (11); LOC_Os10g40710.1 (11); LOC_Os10g40720.1 (11); LOC_Os10g40730.1 (11);

Beta-expansin 1a precursor, putative, expressed	LOC_Os10g40700.1 (11)
Alpha-expansin 7 precursor, putative	LOC_Os10g39110.1 (11)
Alpha-expansin 17 precursor, putative, expressed	LOC_Os10g30340.1 (11)
Alpha-expansin 17 precursor, putative, expressed	LOC_Os10g30340.1 (11)
Expansin-like 3 precursor, putative, expressed	LOC_Os10g39640.1 (11)
Expansin-like 3 precursor, putative, expressed	LOC_Os10g39640.1 (11)
Alpha-expansin 9 precursor, putative, expressed	LOC_Os12g36040.1 (11)
Fasciclin-like arabinogalactan protein 7 precursor, putative	LOC_Os01g06580.1 (11)
C-1-tetrahydrofolate synthase, cytoplasmic, putative, expressed	LOC_Os02g02850.1 (11)
	LOC_Os03g44760.1 (11); LOC_Os03g44760.2 (11);
Retrotransposon protein, putative, unclassified	LOC_Os04g15510.1 (11)
Multidrug resistance-associated protein 4, putative, expressed	LOC_Os04g13220.1 (11)
Anthocyanidin 3-O-glucosyltransferase, putative, expressed	LOC_Os07g05420.1 (8)

ESTs	Pfam matches	CPSG loci	AGI closest homolog	tBLASTx exp to TAIR7
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT2G33790.1				9,00E-07
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT5G05500.1				1,00E-07
0 PF01190		AT5G05500.1		1,00E-09
[LOC_Os0:[LOC_Os01g06990, PF01190],[LOC_Os01g07020, PF01190]				
0				
0				
0				
[LOC_Os0:[LOC_Os01g06990, PF01190],[LOC_Os01g07020, PF01190]				
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT5G05500.1				1,00E-07
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT2G33790.1				9,00E-07
[LOC_Os0:[LOC_Os01g06990, PF01190],[LOC_Os01g07020, PF01190]				
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT5G05500.1				1,00E-07
0 PF01190		AT5G05500.1		3,00E-23
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT2G33790.1				9,00E-07
[LOC_Os0:[LOC_Os01g06990, PF01190],[LOC_Os01g07020, PF01190]				
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT5G05500.1				1,00E-07
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT2G33790.1				9,00E-07
[LOC_Os0:[LOC_Os01g06960, PF01190],[LOC_Os01g0AT5G05500.1				1,00E-07
3 PF01190				
3 PF01190		AT1G76930.2		2,00E-05
[LOC_Os0:[LOC_Os01g52640, PF00400[x3]],[LOC_Os01g52650, PF01190]				
10 PF01190				
15		AT5G54855.1		1,00E-25
[LOC_Os02g01190.1,4],[LOC_Os02g01190.2,19]		AT2G16630.1		6,00E-10
[LOC_Os02g01190.1,4],[LOC_Os02g01190.2,19],[LOC_Os04g33540.1,0],[LOC_Os05g27970.1,0]				
[LOC_Os02g01190.1,4],[LOC_Os02g01190.2,19],[LOC_AT2G16630.1				1,00E-13
[LOC_Os02g57750.1,64],[LOC_Os02c	CPSG	AT5G04460.1		3,00E-32
[LOC_Os0:[LOC_Os02g57760, PF00891]		AT3G53140.1		8,00E-57
[LOC_Os0:[LOC_Os02g57760, PF00891]		AT3G53140.1		2,00E-29
0	CPSG			
[LOC_Os0:[LOC_Os03g14120, PF01113,PF05173],[LOCAT3G59890			other annotation	
[LOC_Os0:[LOC_Os03g14120, PF01113,PF05173],[LOCAT3G19020.1				2,00E-09
[LOC_Os0:[LOC_Os03g14120, PF01113,PF05173],[LOCAT3G59890			other annotation	
[LOC_Os0:[LOC_Os03g14120, PF01113,PF05173],[LOCAT3G59890.1				8,00E-08
1		AT5G03710.1		1,00E-15
1				
[LOC_Os0:[LOC_Os04g32670, PF03016],[LOC_Os04g3AT4G08685.1				1,00E-06
37 PF01190		AT1G28290	other annotation	
7 PF01190				
7 PF01190				
0		AT1G29140.1		5,00E-05

9 PF01190	AT1G29140	other annotation
7	AT3G16660.1	4,00E-05
3	AT3G16670.1	1,00E-09
15 PF01190	AT4G08685.1	2,00E-20
[LOC_Os09g33450.1,79],[LOC_Os09g33450.2,20]	AT3G26935	other annotation
12 PF01190	AT5G13140	other annotation
12 PF01190	AT5G13140	other annotation
10 PF06735	AT4G01985.1	2,00E-06
[LOC_Os0:[LOC_Os10g05950, PF06735]	AT4G01985.1	1,00E-06
[LOC_Os10g05990.1,0],[LOC_Os10g05990.2,0]		
[LOC_Os1([LOC_Os10g05990, PF06735],[LOC_Os10g0AT5G56330.1		6,00E-10
[LOC_Os1([LOC_Os10g05990, PF06735],[LOC_Os10g0AT5G56330.1		6,00E-10
0 PF06735		
0 PF06735	AT5G56330.1	1,00E-11
32	AT2G30560.1	8,00E-19
8 PF01190	AT3G09925.1	2,00E-16
[LOC_Os0([LOC_Os06g04690, PF00646],[LOC_Os06g04980, PF00646],[LOC_Os10g22590, PF01190]		
235 PF01190 CPSG	AT4G08685.1	7,00E-23
8 PF01190	AT3G09925.1	2,00E-16
4	AT2G40113.1	3,00E-11
4		
[LOC_Os1([LOC_Os10g05990, PF06735],[LOC_Os10g0AT5G56330.1		6,00E-10
0 PF06735		
0 PF06735	AT5G56330.1	1,00E-11
[LOC_Os10g05990.1,0],[LOC_Os10g05990.2,0]		
[LOC_Os1([LOC_Os10g05990, PF06735],[LOC_Os10g0AT5G56330.1		6,00E-10
[LOC_Os1([LOC_Os10g05990, PF06735],[LOC_Os10g0AT5G56330.1		6,00E-10
[LOC_Os10g05970.1,411],[LOC_Os10g05970.2,229]		
183 PF06735	AT5G03710.1	1,00E-15
6 PF06735		
32	AT2G30560.1	8,00E-19
6	AT5G15780	other annotation
2 PF03330 PF01357	AT1G20190.1	3,00E-67
3 PF03330 PF01357	AT1G69530.3	2,00E-48
43 PF03330 PF01357		
3 PF01357 PF03330	AT4G28250.1	3,00E-76
28 PF03330 PF01357		
0 PF03330 PF01357	AT2G45110.1	1,00E-57
64 PF03330 PF01357		
64 PF03330 PF01357		
1 PF03330 PF01357	AT1G20190.1	2,00E-45
[LOC_Os0:[LOC_Os02g16780, PF01357,PF03330],[LOCAT1G20190.1		2,00E-67
[LOC_Os0:[LOC_Os02g16780, PF01357,PF03330],[LOCAT1G20190.1		2,00E-67
[LOC_Os0:[LOC_Os02g16809, PF01357,PF03330],[LOCAT2G40610.1		3,00E-33
1 PF03330 PF01357	AT2G40610.1	4,00E-39
[LOC_Os0:[LOC_Os02g16780, PF01357,PF03330],[LOCAT1G20190.1		2,00E-67
1 PF03330 PF01357	AT1G20190.1	5,00E-58
1 PF03330 PF01357	AT1G20190.1	5,00E-58

[LOC_Os0:[LOC_Os02g16780, PF01357,PF03330],[LOCAT1G20190.1	2,00E-67
14 PF03330 PF01357 AT1G20190.1	4,00E-80
14 PF03330 PF01357 AT1G20190.1	4,00E-80
1 PF03330 PF01357 AT1G20190.1	5,00E-58
1 PF03330 PF01357 AT1G20190.1	5,00E-58
[LOC_Os0:[LOC_Os02g16780, PF01357,PF03330],[LOCAT1G20190.1	2,00E-67
[LOC_Os03g04020.1,80],[LOC_Os03g04020.2,49] AT4G38400.1	5,00E-44
1 PF03330 PF01357 AT1G20190.1	8,00E-46
1 PF03330 PF01357 AT1G20190.1	2,00E-58
1 PF03330 PF01357 AT1G20190.1	6,00E-74
[LOC_Os0:[LOC_Os03g06050, PF01357,PF03330],[LOCAT1G20190.1	7,00E-51
[LOC_Os0:[LOC_Os03g06050, PF01357,PF03330],[LOCAT1G20190.1	7,00E-51
1 PF03330 PF01357 AT1G20190.1	1,00E-73
2 PF03330 PF01357 AT1G20190.1	7,00E-19
96 PF03330 PF01357 AT2G39700.1	2,00E-23
[LOC_Os0:[LOC_Os03g01610, PF01357,PF03330],[LOCAT1G65680.1	7,00E-10
27 PF01357 PF03330 AT1G65680.1	7,00E-22
[LOC_Os03g01270.1,51],[LOC_Os03g01270.2,43] AT1G65680.1	7,00E-27
[LOC_Os0:[LOC_Os03g01610, PF01357,PF03330],[LOCAT1G65680.1	7,00E-10
32 PF01357 PF03330 AT1G65680.1	6,00E-25
[LOC_Os03g44290.1,46],[LOC_Os03g44290.2,41],[LOC,AT2G45110.1	2,00E-19
[LOC_Os04g44780.1,1],[LOC_Os04g44780.2,20] AT4G28250	other annotation
[LOC_Os04g44780.1,1],[LOC_Os04g44780.2,20] AT4G28250.1	3,00E-73
1 PF01357 PF03330 AT1G65680.1	4,00E-15
[LOC_Os0:[LOC_Os04g25150, PF01357],[LOC_Os04g26230, PF01357]	
[LOC_Os0:[LOC_Os04g25190, PF01357],[LOC_Os04g26190, PF01357]	
[LOC_Os0:[LOC_Os04g49410, PF01357,PF03330],[LOCAT3G03220.1	1,00E-29
41 PF01357 PF03330 AT1G20190.1	9,00E-42
[LOC_Os04g44780.1,1],[LOC_Os04g44780.2,20] AT4G28250	other annotation
[LOC_Os04g44780.1,1],[LOC_Os04g44780.2,20] AT4G28250.1	3,00E-73
18 PF01357 PF03330 AT1G65680.1	3,00E-17
7 PF03330 PF01357 AT3G55500.1	9,00E-33
[LOC_Os0:[LOC_Os06g45180, PF01357],[LOC_Os06g45290, PF01357]	
15 PF01357 PF03330 AT4G38210.1	3,00E-52
9 PF03330 PF01357 AT1G20190.1	5,00E-15
10 PF01357 PF03330 AT4G38400.1	1,00E-10
15 PF01357 PF03330 AT4G38210.1	3,00E-52
1 PF03330 PF01357 AT1G12560.1	3,00E-51
2 PF03330 AT4G38400.1	2,00E-37
88 PF03330 PF01357 AT4G38400.1	2,00E-18
18 PF03330 PF01357 AT5G05290.1	1,00E-45
177 PF01357 PF03330 AT1G65680.1	4,00E-16
49 PF03330 PF01357 AT1G65680.1	4,00E-10
78 PF01357 PF03330	
324 PF01357 PF03330	

237 PF01357 PF03330		
0 PF03330 PF01357	AT1G62980.1	2,00E-61
1 PF03330 PF01357	AT1G20190.1	1,00E-69
1 PF03330 PF01357	AT1G20190.1	1,00E-69
59 PF03330 PF01357	AT3G45960.1	3,00E-26
59 PF03330 PF01357	AT3G45960.1	3,00E-26
7 PF01357 PF03330	AT5G02260.1	1,00E-39
13 PF02469	AT5G03710.1	1,00E-08
7 PF02882 PF00763	AT4G00620.1	1,00E-106
[LOC_Os03g44760.1,24],[LOC_Os03g44760.2,24],[LOC_	AT5G51330	other annotation
0		
8 PF00005 PF00005 PF00664 PF00664	AT2G47800.1	7,00E-38
13 PF00201	AT5G17050	other annotation

TAIR7 annotation	Jiang sequence reference
Pollen Ole e 1 allergen protein contai	OsOle1 gb AP003209 (-) 106334-108170 length 1836 bp
Pollen Ole e 1 allergen and extensin i	OsOle1 gb AP003209 (-) 106334-108170 length 1836 bp
Pollen Ole e 1 allergen and extensin i	OsOle2 gb AP003209 (-) 131631-132194 length 563 bp
	OsOle3 gb AP003209 (-) 104952-105519 length 567 bp
	OsOle3 gb AP003209 (-) 104952-105519 length 567 bp
	OsOle3 gb AP003209 (-) 104952-105519 length 567 bp
	OsOle4 gb AP003209 (-) 111253-111841 length 588 bp
	OsOle4 gb AP003209 (-) 111253-111841 length 588 bp
Pollen Ole e 1 allergen and extensin i	OsOle4 gb AP003209 (-) 111253-111841 length 588 bp
Pollen Ole e 1 allergen protein contai	OsOle5 gb AP003209 (-) 117071-117639 length 568 bp
	OsOle5 gb AP003209 (-) 117071-117639 length 568 bp
Pollen Ole e 1 allergen and extensin i	OsOle5 gb AP003209 (-) 117071-117639 length 568 bp
Pollen Ole e 1 allergen and extensin i	OsOle6 gb AP003209 (-) 137080-137664 length 584 bp
	OsOle7 gb AP003209 (-) 134183-134740 length 557 bp
Pollen Ole e 1 allergen protein contai	OsOle8 gb AP003209 (-) 122948-123523 length 575 bp
	OsOle8 gb AP003209 (-) 122948-123523 length 575 bp
Pollen Ole e 1 allergen and extensin i	OsOle8 gb AP003209 (-) 122948-123523 length 575 bp
Pollen Ole e 1 allergen protein contai	OsOle9 gb AP003209 (-) 114479-115039 length 560 bp
	OsOle9 gb AP003209 (-) 114479-115039 length 560 bp
Pollen Ole e 1 allergen and extensin i	OsOle9 gb AP003209 (-) 114479-115039 length 560 bp
Pollen Ole e 1 allergen protein contai	OsOle10 gb AP003209 (-) 120152-120721 length 569 bp
Pollen Ole e 1 allergen and extensin i	OsOle10 gb AP003209 (-) 120152-120721 length 569 bp
	OsOle11 gb AP003209 (-) 124482-129835 length 5353 bp
Proline-rich extensin-like family protei	OsOle12 gb AP003313 (-) 6456-8289 length 1833 bp
	OsOle13 gb AP003343 (-) 45286-45899 length 613 bp
	OsOle14 gb AP003343 (-) 54852-55593 length 741 bp
Pollen Ole e 1 allergen and extensin i	OsOle15 gb AP004024 (-) 26461-29589 length 3128 bp
Proline-rich family protein, contains p	OsOle16 gb AP004187 (+) 59823-61240 length 1417 bp
	OsOle16 gb AP004187 (+) 59823-61240 length 1417 bp
Proline-rich family protein, contains p	OsOle16 gb AP004187 (+) 59823-61240 length 1417 bp
Expressed protein chr5:1260010-12	OsOle17 gb AC121491 (-) 91777-92127 length 350 bp
O-diphenol-O-methyl transferase, put	OsOle17 gb AC121491 (-) 91777-92127 length 350 bp
O-diphenol-O-methyl transferase, put	OsOle17 gb AC121491 (-) 91777-92127 length 350 bp
	OsOle17 gb AC121491 (-) 91777-92127 length 350 bp
Dihydrodipicolinate reductase family j	OsOle18 gb AC129008 (+) 68620-69699 length 1079 bp
Leucine-rich repeat family protein / e	OsOle18 gb AC129008 (+) 68620-69699 length 1079 bp
Dihydrodipicolinate reductase family j	OsOle18 gb AC129008 (+) 68620-69699 length 1079 bp
Dihydrodipicolinate reductase family j	OsOle18 gb AC129008 (+) 68620-69699 length 1079 bp
Expressed protein chr5:971106-971	OsOle19 gb AL606592 (-) 12562-16609 length 4047 bp
	OsOle19 gb AL606592 (-) 12562-16609 length 4047 bp
Pollen Ole e 1 allergen and extensin i	OsOle20 gb AL731591 (-) 152866-153549 length 683 bp
Pollen Ole e 1 allergen and extensin i	OsOle21 gb AC121365 (+) 15921-16810 length 889 bp
	OsOle22 gb AC121365 (+) 21630-22652 length 1022 bp
	OsOle22 gb AC121365 (+) 21630-22652 length 1022 bp
Pollen Ole e 1 allergen and extensin i	OsOle23 gb AC135422 (-) 8763-9267 length 504 bp

Pollen Ole e 1 allergen and extensin iOsOle24 gb AP003613 (+) 129618-130232 length 614 bp
OsOle25 gb AP003823 (+) 2544-3050 length 506 bp
Expressed protein | chr3:5676835-56 OsOle26 gb AP004343 (+) 35895-36408 length 513 bp
Expressed protein | chr3:5681425-56 OsOle27 gb AP004343 (+) 38286-39153 length 867 bp
Pollen Ole e 1 allergen and extensin iOsOle28 gb AP006450 (+) 49596-50051 length 455 bp
Zinc finger (DHHC type) family proteiOsOle29 gb AC108759 (-) 127264-131332 length 4068 bp
Expressed protein | chr5:4170563-41 OsOle30 gb AP005706 (+) 149917-151817 length 1900 bp
Expressed protein | chr5:4170563-41 OsOle30 gb AP005706 (+) 149917-151817 length 1900 bp
Expressed protein | chr4:866236-868 OsOle31 gb AE017056 (-) 271367-272931 length 1564 bp
Expressed protein | chr4:866236-868 OsOle31 gb AE017056 (-) 271367-272931 length 1564 bp
OsOle32 gb AE017056 (+) 292512-293553 length 1041 bp
Carbonic anhydrase family protein, α OsOle32 gb AE017056 (+) 292512-293553 length 1041 bp
Carbonic anhydrase family protein, α OsOle33 gb AE017057 (+) 3453-4257 length 804 bp
OsOle33 gb AE017057 (+) 3453-4257 length 804 bp
Carbonic anhydrase family protein, α OsOle33 gb AE017057 (+) 3453-4257 length 804 bp
Glycine-rich protein | chr2:13024424- OsOle34 gb AE017071 (-) 52324-53264 length 940 bp
Pollen Ole e 1 allergen and extensin iOsOle35 gb AE017117 (+) 36771-37889 length 1118 bp
OsOle36 gb AC107314 (+) 75956-78563 length 2607 bp
Pollen Ole e 1 allergen and extensin iOsOle36 gb AC107314 (+) 75956-78563 length 2607 bp
Pollen Ole e 1 allergen and extensin iOsOle37 gb AC025783 (+) 81533-82651 length 1118 bp
Expressed protein | chr2:16757428-1 OsOle38 gb AC092172 (+) 4965-11153 length 6188 bp
OsOle38 gb AC092172 (+) 4965-11153 length 6188 bp
Carbonic anhydrase family protein, α OsOle39 gb AC084884 (+) 99107-99911 length 804 bp
OsOle39 gb AC084884 (+) 99107-99911 length 804 bp
Carbonic anhydrase family protein, α OsOle39 gb AC084884 (+) 99107-99911 length 804 bp
OsOle40 gb AC084884 (+) 88197-89238 length 1041 bp
Carbonic anhydrase family protein, α OsOle40 gb AC084884 (+) 88197-89238 length 1041 bp
Carbonic anhydrase family protein, α OsOle41 gb AC084884 (-) 80529-81317 length 788 bp
OsOle41 gb AC084884 (-) 80529-81317 length 788 bp
Expressed protein | chr5:971106-971 OsOle42 gb AC091665 (-) 57939-59044 length 1105 bp
OsOle43 gb AC091665 (-) 82065-83666 length 1601 bp
Glycine-rich protein | chr2:13024424- OsOle44 gb AC099325 (+) 14403-15343 length 940 bp
Pollen Ole e 1 allergen and extensin iOsOle45 gb AL713941 (+) 94988-97173 length 2185 bp

Expansin, putative (EXP11), similar tOsAllerg1-1 gb AP002865 (-) 118640-119502 length 862 bp
Expansin, putative (EXP1), identical tOsAllerg1-2 gb AP002865 (-) 112929-113805 length 876 bp
OsAllerg1-3 gb AP000837 (+) 58583-59329 length 746 bp
OsAllerg1-4 gb AP003348 (+) 146021-147244 length 1223 bp
Beta-expansin, putative (EXPB3), sirOsAllerg1-5 gb AP004768 (+) 73690-80153 length 6463 bp
OsAllerg1-6 gb AP005289 (+) 101596-102842 length 1246 bp
Beta-expansin, putative (EXPB4), sirOsAllerg1-7 gb AP005289 (+) 91278-92642 length 1364 bp
OsAllerg1-8 gb AP005297 (+) 118982-120047 length 1065 bp
OsAllerg1-9 gb AP004114 (+) 3153-10072 length 6919 bp
Expansin, putative (EXP11), similar tOsAllerg1-10 gb AP004128 (+) 1974-2930 length 956 bp
Expansin, putative (EXP11), similar tOsAllerg1-10 gb AP004128 (+) 1974-2930 length 956 bp
Expansin, putative (EXP11), similar tOsAllerg1-11 gb AP004128 (-) 21523-22434 length 911 bp
Expansin, putative (EXP8), similar to OsAllerg1-11 gb AP004128 (-) 21523-22434 length 911 bp
Expansin, putative (EXP8), similar to OsAllerg1-12 gb AP004128 (-) 25407-26339 length 932 bp
Expansin, putative (EXP11), similar tOsAllerg1-12 gb AP004128 (-) 25407-26339 length 932 bp
Expansin, putative (EXP11), similar tOsAllerg1-13 gb AP004128 (-) 30318-31215 length 897 bp
Expansin, putative (EXP11), similar tOsAllerg1-13 gb AP004128 (-) 30318-31215 length 897 bp

Expansin, putative (EXP11), similar to OsAllerg1-13 gb AP004128 (-) 30318-31215 length 897 bp
 Expansin, putative (EXP11), similar to OsAllerg1-14 gb AP004128 (-) 49042-50270 length 1228 bp
 Expansin, putative (EXP11), similar to OsAllerg1-14 gb AP004128 (-) 49042-50270 length 1228 bp
 Expansin, putative (EXP11), similar to OsAllerg1-15 gb AP005428 (+) 32993-33890 length 897 bp
 Expansin, putative (EXP11), similar to OsAllerg1-15 gb AP005428 (+) 32993-33890 length 897 bp
 Expansin, putative (EXP11), similar to OsAllerg1-15 gb AP005428 (+) 32993-33890 length 897 bp
 Expansin family protein (EXPL2), conOsAllerg1-16 gb AC098693 (-) 21866-23289 length 1423 bp
 Expansin, putative (EXP11), similar to OsAllerg1-17 gb AC134233 (+) 118069-119000 length 931 bp
 Expansin, putative (EXP11), similar to OsAllerg1-18 gb AC134233 (-) 122626-123738 length 1112 bp
 Expansin, putative (EXP11), similar to OsAllerg1-18 gb AC134233 (-) 122626-123738 length 1112 bp
 OsAllerg1-19 gb AC134233 (+) 140429-141535 length 1106 bp
 Expansin, putative (EXP11), similar to OsAllerg1-20 gb AC134233 (+) 145699-146670 length 971 bp
 Expansin, putative (EXP11), similar to OsAllerg1-21 gb AC134233 (+) 156296-157052 length 756 bp
 Expansin, putative (EXP11), similar to OsAllerg1-22 gb AC105730 (+) 107233-108234 length 1001 bp
 Expansin, putative (EXP11), similar to OsAllerg1-23 gb AC133398 (-) 36141-36905 length 764 bp
 OsAllerg1-24 gb AC084766 (+) 71888-73141 length 1253 bp
 Expansin, putative (EXP4), similar to OsAllerg1-25 gb AC093018 (+) 127826-129986 length 2160 bp
 Similar to beta-expansin, putative (E)OsAllerg1-26 gb AC107224 (-) 107332-108135 length 803 bp
 Similar to beta-expansin, putative (E)OsAllerg1-27 gb AC125411 (-) 91657-92487 length 830 bp
 Similar to beta-expansin, putative (E)OsAllerg1-28 gb AC125411 (+) 101900-103430 length 1530 bp
 Similar to beta-expansin, putative (E)OsAllerg1-29 gb AC107224 (+) 92783-93586 length 803 bp
 Similar to beta-expansin, putative (E)OsAllerg1-30 gb AC107224 (+) 103250-104053 length 803 bp
 OsAllerg1-31 gb AC125496 (-) 103172-104065 length 893 bp
 Beta-expansin, putative (EXPB4), similar to OsAllerg1-32 gb AL713930 (+) 78600-81523 length 2923 bp
 Beta-expansin, putative (EXPB3), similar to OsAllerg1-33 gb AL662984 (+) 27876-31792 length 3916 bp
 Beta-expansin, putative (EXPB3), similar to OsAllerg1-33 gb AL662984 (+) 27876-31792 length 3916 bp
 Similar to beta-expansin, putative (E)OsAllerg1-34 gb AL606633 (+) 59616-60718 length 1102 bp
 OsAllerg1-35 gb AL606633 (+) 49610-51094 length 1484 bp
 OsAllerg1-36 gb AL731608 (+) 140724-141077 length 353 bp
 OsAllerg1-37 gb AL662972 (+) 632-1322 length 690 bp
 Expansin, putative (EXP13), similar to OsAllerg1-38 gb AL662987 (+) 131472-134499 length 3027 bp
 Expansin, putative (EXP11), similar to OsAllerg1-39 gb AL731598 (+) 17466-18351 length 885 bp
 Beta-expansin, putative (EXPB3), similar to OsAllerg1-40 gb AL662984 (+) 29083-32208 length 3125 bp
 Beta-expansin, putative (EXPB3), similar to OsAllerg1-40 gb AL662984 (+) 29083-32208 length 3125 bp
 Similar to beta-expansin, putative (E)OsAllerg1-41 gb AC119290 (-) 102796-103924 length 1128 bp
 Expansin, putative (EXP16), similar to OsAllerg1-42 gb AC130731 (-) 66267-67702 length 1435 bp
 OsAllerg1-43 gb AC144736 (-) 37899-39249 length 1350 bp
 OsAllerg1-44 gb AP004744 (-) 118156-118509 length 353 bp
 Expansin, putative (EXP20), similar to OsAllerg1-45 gb AP003539 (+) 82016-86001 length 3985 bp
 Expansin, putative (EXP11), similar to OsAllerg1-46 gb AP003769 (+) 80512-81388 length 876 bp
 Expansin family protein (EXPL2), conOsAllerg1-47 gb AP003685 (+) 77095-78266 length 1171 bp
 Expansin, putative (EXP20), similar to OsAllerg1-48 gb AP003938 (-) 117272-119874 length 2602 bp
 Expansin, putative (EXP7), similar to OsAllerg1-49 gb AP000616 (-) 56364-57313 length 949 bp
 Expansin family protein (EXPL2), conOsAllerg1-50 gb AP003956 (+) 73393-74676 length 1283 bp
 Expansin family protein (EXPL2), conOsAllerg1-51 gb AP004315 (-) 62167-65086 length 2919 bp
 Expansin, putative (EXP2), identical to OsAllerg1-52 gb AP005544 (+) 127880-128764 length 884 bp
 Similar to beta-expansin, putative (E)OsAllerg1-53 gb AC020666 (-) 117022-118492 length 1470 bp
 Similar to beta-expansin, putative (E)OsAllerg1-54 gb AC037426 (+) 15951-17196 length 1245 bp
 OsAllerg1-55 gb AC037426 (+) 3921-5311 length 1390 bp
 OsAllerg1-56 gb AC069300 (+) 14989-16770 length 1781 bp

OsAllerg1-57 gb AC037426 (+) 21967-23058 length 1091 bp
 Expansin, putative (EXP18), identical OsAllerg1-58 gb AC092697 (+) 42611-43835 length 1224 bp
 OsAllerg1-59 gb AC087550 (+) 91512-92802 length 1290 bp
 Expansin, putative (EXP11), similar to OsAllerg1-60 gb AC087550 (-) 94406-95419 length 1013 bp
 Expansin, putative (EXP11), similar to OsAllerg1-60 gb AC087550 (-) 94406-95419 length 1013 bp
 Similar to expansin family protein (EXOsAllerg1-61 gb AC051633 (+) 62082-63446 length 1364 bp
 Similar to expansin family protein (EXOsAllerg1-61 gb AC051633 (+) 62082-63446 length 1364 bp
 Expansin, putative (EXP9), similar to OsAllerg1-62 gb AL731742 (-) 127680-129794 length 2114 bp

 Expressed protein | chr5:971106-971 OsAllerg2-1 gb AP002521 (-) 40381-41190 length 809 bp
 Tetrahydrofolate dehydrogenase/cycl OsAllerg2-2 gb AP004078 (-) 39274-42095 length 2821 bp
 Meiosis protein-related (DYAD) (SWI OsAllerg2-3 gb AC145381 (+) 64536-69700 length 5164 bp
 OsAllerg2-4 gb AL662993 (+) 4772-6934 length 2162 bp
 Glutathione-conjugate transporter (MIOsAllerg2-5 gb AL663001 (-) 11965-14969 length 3004 bp
 UDP-glucuronosyl/UDP-glucosyl tran: OsAllerg2-6 gb AP004263 (+) 85986-87419 length 1433 bp
 OsAllerg2-7 gb AC119149 (-) 39538-41199 length 1661 bp

IRGSP matches to Affymetrix probes

chr01|11971:3225885-3226584 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3311803-3312411 (-) // 96.19 //
chr01|11971:3286506-3288343 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3297243-3297812 (-) // 100.0 // /// chr01|11971:3287762-3303696 (-) // 94.29 // /// chr01|11971:
chr01|11971:3286506-3288343 (-) // 100.0 //
chr01|11971:3286506-3288343 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3225885-3226584 (-) // 100.0 //
chr01|11971:3286506-3288343 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3283704-3283997 (-) // 100.0 //

chr01|11971:3225885-3226584 (-) // 100.0 //
chr01|11971:3286506-3288343 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3225885-3226584 (-) // 100.0 //
chr01|11971:3286506-3288343 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3225885-3226584 (-) // 100.0 //
chr01|11971:3291425-3292014 (-) // 100.0 // /// chr01|11971:3300324-3303696 (-) // 90.67 //
chr01|11971:3283704-3283997 (-) // 100.0 //
chr01|11971:39463785-39466229 (-) // 99.75 //
chr01|11971:30590229-30591228 (-) // 100.0 //
chr06|11976:18071224-18071883 (-) // 86.13 //
chr02|11972:12567315-12580766 (+) // 100.0 //
chr02|11972:110805-112358 (+) // 98.25 //
chr05|11975:16205033-16205498 (+) // 100.0 //
chr05|11975:16206605-16207526 (-) // 100.0 //
chr02|11972:35357005-35363032 (+) // 99.9 //
chr06|11976:2090289-2093661 (-) // 99.22 //
chr06|11976:24531076-24532738 (-) // 100.0 //
chr03|11973:12708080-12710295 (+) // 100.0 //
chr04|11974:16232832-16237500 (+) // 99.88 //
chr04|11974:16232832-16237500 (+) // 99.88 //
chr03|11973:7650735-7655907 (+) // 99.91 //
chr03|11973:7654817-7656178 (-) // 99.9 //
chr04|11974:20854042-20855739 (-) // 100.0 //
chr10|11980:22623028-22624138 (-) // 100.0 //
chr04|11974:19503154-19507279 (+) // 99.93 //
chr09|11979:16986887-16989405 (-) // 99.8 //
chr05|11975:26245354-26246307 (+) // 100.0 //
chr05|11975:26273187-26275243 (-) // 100.0 //
chr05|11975:7291603-7292057 (+) // 100.0 //

chr05|11975:22795521-22798595 (-) // 99.93 //
 chr05|11975:20225331-20227088 (-) // 100.0 //
 chr05|11975:3750418-3751115 (+) // 99.86 //
 chr04|11974:20224927-20227854 (+) // 99.87 //
 chr09|11979:19701630-19706208 (-) // 99.71 //
 chr09|11979:18943433-18944272 (-) // 100.0 //
 chr09|11979:21260875-21261421 (-) // 100.0 //
 chr07|11977:26616960-26620328 (-) // 99.84 //
 chr11|11981:5006218-5008656 (-) // 97.93 //
 chr10|11980:3004848-3006285 (+) // 99.72 //
 chr10|11980:2996985-2998097 (-) // 99.9 //
 chr10|11980:2996985-2998097 (-) // 99.9 //
 chr05|11975:27603454-27605138 (+) // 99.92 //
 chr10|11980:3015766-3016621 (+) // 71.63 //
 chr10|11980:7294502-7295642 (-) // 99.71 //
 chr10|11980:21255321-21257922 (-) // 99.72 //
 chr06|11976:2195235-2197149 (+) // 78.39 // /// chr06|11976:2042908-2047320 (-) // 94.65 // /// chr06|11976
 chr06|11976:7693324-7695943 (-) // 99.95 //
 chr10|11980:21255321-21257922 (-) // 99.72 //
 chr10|11980:7322724-7329236 (+) // 99.55 //
 chr10|11980:7322754-7323577 (+) // 91.5 //
 chr10|11980:2996985-2998097 (-) // 99.9 //
 chr05|11975:27603454-27605138 (+) // 99.92 //
 chr10|11980:3015766-3016621 (+) // 71.63 //
 chr10|11980:3004848-3006285 (+) // 99.72 //
 chr10|11980:2996985-2998097 (-) // 99.9 //
 chr10|11980:2996985-2998097 (-) // 99.9 //
 chr02|11972:26079331-26079668 (+) // 96.45 //
 chr11|11981:226411-230967 (+) // 98.19 // /// chr12|11982:241967-246389 (+) // 99.83 //
 chr10|11980:2905913-2907826 (-) // 99.84 //
 chr10|11980:7294502-7295642 (-) // 99.71 //
 chr06|11976:15600604-15601843 (+) // 99.78 //
 chr07|11977:11455839-11456511 (+) // 100.0 //
 chr01|11971:8192884-8193954 (-) // 100.0 //
 chr01|11971:35472365-35474187 (+) // 99.78 //
 chr02|11972:25498777-25500043 (-) // 100.0 //
 chr08|11978:25294211-25298656 (+) // 99.9 //
 chr02|11972:26538182-26538521 (+) // 96.46 // /// chr10|11980:6622460-6622799 (+) // 100.0 // /// chr10|11980:6622460-6622799 (+) // 100.0 //
 chr02|11972:31214698-31216303 (+) // 98.26 //
 chr02|11972:31214698-31216303 (+) // 98.26 //
 chr02|11972:9555584-9555815 (+) // 100.0 //
 chr02|11972:9569674-9570572 (+) // 100.0 //
 chr02|11972:9569674-9570572 (+) // 100.0 //
 chr02|11972:9569674-9570572 (+) // 100.0 //
 chr02|11972:9562674-9563847 (+) // 100.0 //
 chr02|11972:9569674-9570572 (+) // 100.0 //
 chr02|11972:9600302-9601259 (-) // 100.0 //
 chr02|11972:9574550-9575483 (+) // 100.0 //

chr02|11972:9569674-9570572 (+) // 100.0 //
chr04|11974:24758111-24758762 (-) // 100.0 //
chr04|11974:24931264-24933778 (+) // 100.0 //
chr02|11972:9600302-9601259 (-) // 100.0 //
chr02|11972:9574550-9575483 (+) // 100.0 //
chr02|11972:9569674-9570572 (+) // 100.0 //
chr03|11973:1809674-1811385 (+) // 97.58 //
chr03|11973:2977965-2979110 (+) // 92.86 //
chr11|11981:24939518-24941477 (-) // 100.0 //
chr11|11981:24975673-24977449 (-) // 100.0 //

chr03|11973:2934943-2935369 (-) // 100.0 //
chr03|11973:2934943-2935369 (-) // 100.0 //
chr03|11973:2946505-2947084 (+) // 100.0 //
chr12|11982:22014243-22020753 (+) // 99.88 //

chr03|11973:34445315-34450213 (-) // 99.93 //
chr03|11973:365219-366444 (+) // 96.23 // /// chr03|11973:379414-380638 (-) // 96.15 //
chr03|11973:157565-158553 (-) // 96.67 //
chr08|11978:19588423-19593153 (-) // 99.91 //
chr03|11973:365219-366444 (+) // 96.23 // /// chr03|11973:379414-380638 (-) // 96.15 //
chr03|11973:375685-376785 (+) // 99.91 //

chr03|11973:1982366-1984095 (+) // 98.57 //
chr05|11975:25594387-25599283 (+) // 99.97 //
chr04|11974:26285141-26287968 (+) // 100.0 //
chr04|11974:27436572-27437986 (+) // 98.1 //

chr04|11974:15267833-15268519 (+) // 100.0 // /// chr04|11974:14510064-14510750 (-) // 99.13 //
chr03|11973:2138447-2140844 (-) // 98.89 //
chr04|11974:29266052-29269509 (+) // 98.2 //
chr04|11974:8578855-8580134 (+) // 98.32 //
chr05|11975:25594387-25599283 (+) // 99.97 //
chr04|11974:26285141-26287968 (+) // 100.0 //
chr07|11977:15897736-15900452 (-) // 99.96 //
chr05|11975:11372416-11374199 (-) // 99.19 //

chr06|11976:27325493-27327706 (-) // 100.0 //
chr08|11978:11454035-11456300 (+) // 99.86 //
chr06|11976:30519196-30520348 (+) // 99.44 //
chr05|11975:18781090-18784140 (+) // 85.16 //
chr08|11978:11454035-11456300 (+) // 99.86 //
chr06|11976:532474-534605 (-) // 100.0 //
chr07|11977:18640562-18641173 (+) // 100.0 //
chr04|11974:33872603-33876349 (+) // 99.34 //
chr05|11975:17712860-17713988 (-) // 99.91 //
chr10|11980:21135322-21137161 (+) // 98.4 //
chr10|11980:21506244-21507964 (-) // 99.53 //
chr07|11977:74408-80730 (+) // 100.0 //
chr10|11980:21529165-21531620 (+) // 99.8 //

chr10|11980:21500470-21502007 (-) // 98.67 //
chr10|11980:20314719-20315200 (-) // 100.0 // /// chr12|11982:3081060-3081536 (-) // 90.73 //

chr10|11980:15652940-15658420 (+) // 100.0 //
chr01|11971:23326787-23327638 (-) // 100.0 //
chr10|11980:20849866-20851619 (+) // 99.59 //
chr10|11980:20849866-20851619 (+) // 99.59 //
chr12|11982:22055980-22058413 (+) // 99.66 //

chr01|11971:3094149-3095324 (-) // 99.66 //
chr02|11972:975622-976077 (-) // 100.0 //
chr05|11975:18680508-18684040 (-) // 95.23 //
chr04|11974:8081058-8081946 (-) // 100.0 //
chr04|11974:7297067-7305626 (-) // 99.96 //
chr02|11972:31485828-31488480 (+) // 99.85 //

:3294651-3300894 (-) // 91.05 //

2056783-2058121 (-) // 73.83 //

80:8930817-8931156 (-) // 96.17

Contents of Supplementary Data:

1. Pipeline-Annotated “Pollen Allergens” Not Specifically Expressed in Rice Pollen
2. Ole e 1-related allergens (as characterized by Jiang et al., 2005)
3. Ory s 1 (Further Details)
4. Ory s 2 (Further Details)
5. Additional Potential Group I, II/III Allergens (Further Details)
6. Profilin A (Ory s 12) (Further Details)
7. Ory s 23 (Further Details)
8. Other Suspected Allergens (Further Details)
 - a. EF Hand
 - b. Pectate Lyase
 - c. Glycoside hydrolase family 28, Exopolygalacturonase
 - d. FAD Binding

1. Pipeline-Annotated “Pollen Allergens” Not Specifically Expressed in Rice Pollen

Pipeline annotation of the rice genome includes loci annotated in the rice genome as “pollen allergens” particularly those from dicots (annotated as being similar to “Alt a 7”, “Ole e 8”, and “Bet v 1-D/H”) and more distant gymnosperms (“Jun a 1”) were poor candidates and were not found in rice. “Bet v 1-D/H”-like proteins, abundantly expressed in seeds and sporophytic tissues (supported by microarray data, digital Northern, and ESTs), are potential allergens, but not pollen allergens. Not abundant, but nonetheless produced mainly in pollen were putative Ory s 4, (pipeline annotated as “pollen allergen Phl p 4,” LOC_Os05g28490) and a putative Ory s 3 (“Group 3 pollen allergen, putative, expressed,” LOC_Os06g45210).

2. Ole e 1-related allergens (as characterized by Jiang et al., 2005)

Numerous Ole e I related sites were documented in the genomic assay of Jiang et al. (2005), but as is the case for other dicot allergens, in fact only two loci were highly activated and upregulated in pollen grains—LOC_Os06g36240.1, pollen allergen Ory s 11 (pipeline annotated as Phl p 11), and LOC_Os09g39950.1, pollen-specific protein C13 precursor (Supplementary Table S1). The Ory s 11 locus encodes a protein with a signal

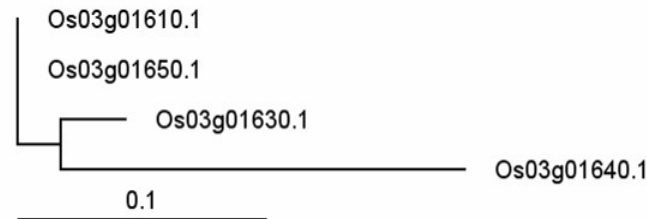
peptide (27 AVA-TA) with 40% identity and 64% positives with an *Arabidopsis* gene (At1g29140, pollen Ole e 1 allergen and extensin family protein, $p < 9e^{-29}$); this protein is also expressed in MPSS at an intensity rank of 246 and has proteome support (Dai et al., 2007), indicating translation and reasonable abundance. The pollen-specific protein C13 precursor locus also encodes a protein with a signal peptide (27 AVA-TA) and has 38% identity and 53% positives with an *Arabidopsis* gene (At4g08685, SAH7, $p < 2e^{-20}$); it is 63rd in intensity among MPSS signatures and is represented in three matches to rice proteins in two proteome studies, reflecting high abundance in mature pollen (Dai et al., 2006, 2007). Both contain a predicted transmembrane domain enclosing a signal peptide region at the C-end, suggesting a hydrophobic terminus. *Arabidopsis* homologues both reflect annotation in TAIR8 as pollen Ole e 1 allergens and extensin family proteins, indicating retention of allergic activity across dicot and monocot classes; these are presumed to share a similar role in cell wall modification.

Supplementary Table S1. Pollen expression microarray data on sequences with “Ole e 1” motif. (Predicted MWs 18.86 to 17.73 KDa, respectively).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (with Pfams, orientation and introns)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
45.	Os.53412.1.S1_at	Os06g36240.1 (11)	LOC_Os06g36240 Pollen allergen Ory s 11, putative, expressed (Pfam:PF01190) (+ with 1 intron)	8.858	12.285	3.18E-12	ESTs: LOC_Os06g36240 (9)	GATCTGTTTTGTAC GTA 95 GATCTGCGGGGGGC GCC 661	9b Dai et al., 2007
200.	Os.50339.1.S1_at	Os09g39950.1 (11)	LOC_Os09g39950 Pollen-specific protein C13 precursor, putative, expressed (Pfam:PF01190) (+ with no introns)	7.375	10.885	4.68E-08	ESTs: LOC_Os09g39950 (15)	GATCGGCGGAGGG GACG 2307	BAD54680, 10b, 11b Dai et al., 2006, 2007

3. Ory s 1 (Further Details)

Outside of the coding region, there are minor differences in the different isoforms of LOC_Os03g01610 and LOC_Os03g01650, including a slightly shorter 5'-UTR region for LOC_Os03g01650 (22 bp vs. 35 bp) and a slightly longer 3'UTR region (400 versus 388 bp); both are attributed with ESTs. The most similar gene to the former two is LOC_Os03g01630, which possesses 96% identity (before removal of its 117 bp intron), whereas LOC_Os03g01640 is the least similar, displaying a 82% identity (Supplementary Figure S1). The presence of an intron in LOC_Os03g01630 is unusual. Genomic information from *Oryza sativa* ssp. *indica* indicates similar sequence patterns that are conserved in other *Oryza* species.



Supplementary Figure S1. Phylogenetic diagram of relationships of Ory s 1 loci using sequence of putative proteins. LOC_Os03g01610.1 is identical to LOC_Os03g01650.1 though the latter is on the (-) strand, whereas the former is in opposite orientation on the (+) strand. Note grouping of loci in chromosome 4 match locus on (+) strand with locus on (-) strand. Chromosome 6 sequences are also largely grouped in triplets. Note significant divergence of amino acid sequences.

According to probe set Os.2402.1.S1_at (which matches LOC_Os03g01610 and LOC_Os03g01650), Ory s 1 was the second most upregulated transcript in rice pollen, displaying 9.18 log₂ intensity of label, upregulated over the seedling control by 13.185 log₂ (>×8000). Probe set Os.2405.1.S1_at ranked 30th among activated probes (full match for LOC_Os03g01640), displaying an intensity of 7.805 log₂ upregulated by 12.447 log₂ over the seedling control (Supplementary Table S2). Abundance of Ory s 1 transcripts in pollen is also strongly supported by MPSS (<http://mpss.udel.edu/rice/>, dataset NPO; Nobuta et al., 2007) with Ory s 1-related signatures in the top 85 signatures in mature pollen. Proteome support indicates abundant translation products, coinciding most closely with LOC_Os03g01610 and LOC_Os03g01650 sequences, and accounting for nine reported proteins from five proteome studies to date on rice (Kerim et al., 2003a, 2003b; Imin et al., 2004; Dai et al., 2006, 2007). Ory s 1 members appear to be exclusively expressed in pollen, so the number of recovered ESTs has been relatively low; pollen has not been a direct target of EST studies, unlike MPSS and proteome studies. As indicated in the text, the four Ory s 1 loci are phylogenetically closely related to one another.

The upstream promoter region of these four genes had several areas of local sequence conservation, with the most prevalent area of conservation (-68 to -51) coinciding with the TAAATA sequence noted as a putative TATA box by Xu et al. (1999). Typical pollen vegetative cell promoter motifs, for example those of LAT52 (Bate and Twell, 1998) and g10 (Rogers et al., 2001) promoter motifs were not noted within positions coinciding with promoter activity in those prior studies. Sequence information and alignment are in another supplementary data file.

Supplementary Table S2. Pollen expression microarray data on Ory s 1. (Predicted molecular weights 24.63 to 28.94 KDa, measured weights have varied from 30-34 KDa).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (with Pfams, strand orientation and introns)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
2	Os.2402.1.S1_at	LOC_Os03g01610 (11) LOC_Os03g01630 (3) LOC_Os03g01650 (11)	[LOC_Os03g01610 ¹ , Major pollen allergen Ory s 1 precursor, putative, expressed (Pfam:PF01357, PF03330)] (+ with no intron), [LOC_Os03g01630 ² , Major pollen allergen Ory s 1 precursor, putative (Pfam:PF01357,PF03330) (+ with 1 intron)], [LOC_Os03g01650 ³ , Major pollen allergen Ory s 1 precursor, putative, expressed (Pfam:PF01357, PF03330)] (- with no introns)	9.814	13.185	5.29E-08	Xu et al., 1995, 1999 ESTs: Os03g01610 (28) Os03g01630 (0) Os03g01650 (18)	GATCAAGTGCTC CAAGC ¹⁻³ 77 GATCACCTTCCAC ATCG ¹⁻³ 52 GATCATCGCCGA GGACG ¹⁻³ 1097 GATCCGGCGCGG CTGCC ^{1,3} 1772	^{1,3} Kerim et al., 2003a, 2003b; Imin et al., 2004; Dai et al., 2006, 2007 ² LOC_Os03g01630 Kerim et al., 2003b; Dai et al., 2006
30	Os.2405.1.S1_at	LOC_Os03g01640 (11)	LOC_Os03g01640 Major pollen allergen Ory s 1 precursor, putative, expressed (Pfam:PF01357,PF03330) (+ with no introns)	7.805	12.447	4.49E-08	Xu et al., 1995, 1999 ESTs: Os03g01640 (32)	GATCCAAGTCGA CTAAG 367 GATCGAGTCTTC AATTC 76 GATCAGAAAATT TTGAT 18	Kerim et al., 2003b; Dai et al., 2006, 2007

4. Ory s 2 (Further Details)

MPSS supports that these Phl p 2 pollen allergen homologues are transcriptionally very highly upregulated. There are ten MPSS signatures—seven, corresponding to the loci listed below, are in the top 25 MPSS for rice pollen.

7. LOC_Os04g26220 (Ory s 2-C)
8. LOC_Os04g26230 (Ory s 2-B)
10. LOC_Os06g45290 (Ory s 2-A)
12. LOC_Os06g45180 (Ory s 2-A)
17. LOC_Os04g25160 (Ory s 2-C)
18. LOC_Os06g44470 (Ory s 2)
23. LOC_Os04g25150 (Ory s 2-B)

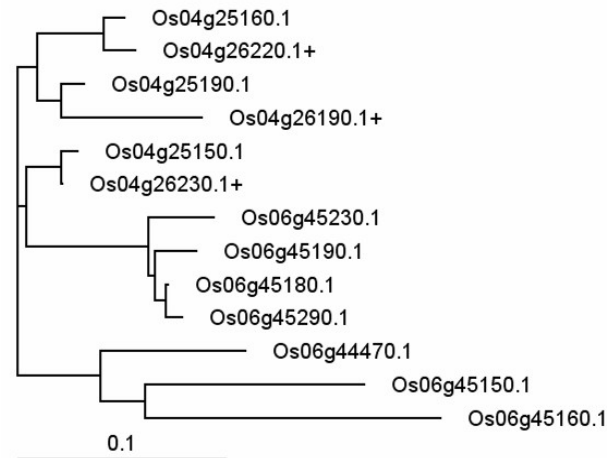
Less abundant transcripts are LOC_Os06g45150 (Ory s 2), which is the 78th most highly upregulated locus, and LOC_Os06g45160 (Ory s 2), which is 104th. LOC_Os06g45190 (Ory s 2) is present at lower levels but is clearly also transcribed (<http://mpss.udel.edu/rice/>, dataset NPO; Nobuta et al., 2007).

Abundance of translated products is supported by proteome data (Kerim et al., 2003b; Dai et al., 2006, 2007), the latter studies supporting loci LOC_Os04g25160, LOC_Os04g25190, LOC_Os06g44470, and LOC_Os06g45180. Thus, high expression of Ory s 2 in rice pollen is supported by transcriptome and proteome data. In Kerim et al. (2003b), sequences corresponding to these were designated Ory s 2-C.

The genomic arrangement of the Ory s 2 family suggests that its loci natively lack introns; intron-bearing loci appear to represent altered copies that may be functionless pseudogenes (LOC_Os04g26190 and LOC_Os06g45200). There appear to be four major groups of Ory s 2 genes. Two groups are located on chromosome 4 and two on chromosome 6, with triplicate pairings a common pattern. On chromosome 6, all copies were on the (-) strand. Of these, LOC_Os06g45190, LOC_Os06g45180, LOC_Os06g45230, and LOC_Os06g45290 were most similar; LOC_Os06g44470, LOC_Os06g45150, LOC_Os06g45160 formed a second grouping that was nested within nearer to some chromosome 4 sequences (see Figure 2, Supplementary Table S3).

In contrast, loci on chromosome 4 are found in groups of three and each of these shares strong sequence similarity with loci on the opposite strand and oriented in the opposite direction. The (-) strand of chromosome 4 has three loci, LOC_Os04g25150, LOC_Os04g25160, and LOC_Os04g25190. These sequences are most closely paired with LOC_Os04g26230, LOC_Os04g26220 and LOC_Os04g26190, respectively. Thus, this sequence appears to represent a reversed insertion of a locus triplet on the reverse strand that also seems to include interstitial loci as well (Fig. 2).

The remaining seven related sequences are found on chromosome 6 in tandem and non-tandem groupings and all on this chromosome are coded on the (-) strand. Predicated peptides display sufficient variability that less than half of the coded amino acid sequences are identical (51 of the 117), with slightly greater conservation of sequence in the “pollen allergen group I” motif region. The most divergent genes are LOC_Os06g45200, which has two introns, and LOC_Os04g26190, which is a truncated version. Both of these excise the signal peptide region and have not been recovered. LOC_Os04g26190 encodes essentially only the core pollen allergen group I motif (PF01357) from position 2 to 75, a 90 amino acid sequence that is a close match for Lol p 2-A in organization and sequence, though it seems to have arisen from Ory s 2 (see Supplementary Figure S2). Kerim et al. (2003b) named Ory s 2-A based on SWISS PROT P83466 (LOC_Os06g45180 and LOC_Os06g45290), Ory s 2-B based on CAD40510 (LOC_Os04g26230) and Ory s 2-C based on CAD40509 (LOC_Os04g25160).



Supplementary Figure S2. Phylogenetic diagram of relationships of Ory s 2 loci using sequence of putative proteins (omitting LOC_Os06g45200, which is the most divergent). Note grouping of loci in chromosome 4 match locus on (+) strand with locus on (-) strand. Chromosome 6 sequences are also largely grouped in triplets. Note significant divergence of amino acid sequences.

Supplementary Table S3. Pollen expression microarray data on Ory s 2. (Predicted molecular weights 12.30 to 12.46 KDa, near measured weight on 2D).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (with Pfams, orientation and introns)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
32	Os.9665.1.S1_at	LOC_Os04g25190.1 (11); LOC_Os04g26190.1 (5)	[LOC_Os04g25190 ¹ , Pollen allergen Ory s 2 precursor, putative, expressed (Pfam:PF01357) (- with no introns)], [LOC_Os04g26190 ² , Pollen allergen Lol p 2-A, putative (Pfam:PF01357)] (+ with 1 intron)	8.222	12.427	1.47E-10	ESTs: LOC_Os04g25190 (16) LOC_Os04g26190 (0)	¹² GATCGTGTTCATGTTGA (11946)	XP_471811 Dai et al. 2006, 2007
75	Os.7428.1.S1_at	LOC_Os06g44470.1 (11)	LOC_Os06g44470 Pollen allergen Ory s 2 precursor, putative, expressed (Pfam:PF01357) (- with no introns)	7.103	12.011	6.69E-06	ESTs: LOC_Os06g44470 (20)	GATCTGTTTGTCTTTG (6635)	BAD37571 Dai et al. 2006, 2007
116	OsAffx.15933.1.S1_s_at	LOC_Os06g45150.1 (11)	LOC_Os06g45150 Pollen allergen Ory s 2 precursor, putative, expressed (Pfam:PF01357) (- with no introns)	7.718	11.718	3.02E-07	ESTs: LOC_Os06g45150 (5)	GATCAATGATGTTTCATC (244) GATCGAGTCTGCCTGTG (943)	
140	Os.25407.1.S1_at	LOC_Os04g25160.1 (10); LOC_Os04g26220.1 (2)	[LOC_Os04g25160 ¹ , Pollen allergen Ory s 2-C precursor, putative, expressed (Pfam:PF01357) (- with no introns)], [LOC_Os04g26220 ² , Pollen allergen Ory s 2-C precursor, putative, expressed (Pfam:PF01357) (+ with no introns)]	7.994	11.358	2.11E-08	ESTs: LOC_Os04g25160 (7) LOC_Os04g26220 (4)	¹ GATCACTCCATGTTCTGA (953) ² GATCGCTCCATGTTCTGA (15264)	CAD40508 Dai et al. 2006, 2007 Kerim et al., 2003b

148	Os.22589.1.S1_at	LOC_Os06g45160.1 (11)	LOC_Os06g45160 Pollen allergen Ory s 2 precursor, putative, expressed (Pfam:PF01357) (- with no introns)	7.998	11.303	2.14E-09	ESTs: (7)	GATCAAGGAGAAGGGTG (3203)	
149	Os.12319.1.S1_x_at	LOC_Os04g25150.1 (9); LOC_Os04g26230.1 (11)	[LOC_Os04g25150, Pollen allergen Ory s 2-B precursor, putative, expressed (Pfam:PF01357) (- with no introns)], [LOC_Os04g26230, Pollen allergen Ory s 2-B precursor, putative, expressed (Pfam:PF01357) (+ with no introns)]	9.334	11.303	2.02E-10	ESTs: LOC_Os04g25150 (5) LOC_Os04g26230 (29)	¹ GATCACGCCATGTTGA (5449) ² GATCACGCCACGTTGA (13806) ¹² GATCAAGGAGAAGGGTG (3203)	CAD40509 Kerim et al., 2003b
906	OsAffx.28114.3.A1_at	LOC_Os06g45180.1 (1); LOC_Os06g45190.1 (11)	[LOC_Os06g45180 ¹ , Pollen allergen Ory s 2-A precursor, putative, expressed (Pfam:PF01357) (- with no introns)], [LOC_Os06g45190 ² , Pollen allergen Ory s 2 precursor, putative, expressed (Pfam:PF01357) (- with no introns)]	5.204	6.463	1.06E-05	ESTs: LOC_Os06g45180 (6) LOC_Os06g45190 (3)	¹ GATCAAGTCTGCATTG (6545) ¹ GATCACGCTACGCGTGA (8787) ² GATCAAGTCTGCATTG (2191)	¹ BAD45861 Dai et al. 2006, 2007 Kerim et al., 2003b
1010	OsAffx.28114.2.A1_at	LOC_Os06g45180.1 (3); LOC_Os06g45290.1 (11)	[LOC_Os06g45180 ¹ , Pollen allergen Ory s 2-A precursor, putative, expressed (Pfam:PF01357) (- with no introns)], [LOC_Os06g45290 ² , Pollen allergen Ory s 2-A precursor, putative, expressed (Pfam:PF01357) (- with no introns)]	5.413	6.082	6.24E-08	ESTs: LOC_Os06g45180 (6) LOC_Os06g45290 (24)	¹ GATCAAGTCTGCATTG (6545) ¹ GATCACGCTACGCGTGA (8787) ² GATCATGCTATGCGTGA (10885)	¹ BAD45861 Dai et al. 2006, 2007 Kerim et al., 2003b

5. Additional Potential Group I, II/III Allergens (Further Details)

The first additional allergen is LOC_Os10g40090, which is annotated as a putative, expressed β -expansin-1a precursor with a signal peptide at position 24 (GSC-GP), an RlpA-like double-psi beta-barrel (DPBB_1) at positions 80 to 159 and “pollen_allerg_1” motif at positions 172 to 253). This is ranked 3rd among transcripts in the 57K microarray, the 34th most highly upregulated sequence in MPSS and has proteome support. The second is LOC_Os08g44790, which is a putative, expressed α -expansin-3 precursor with signal peptide at position 25 (GDA-AP) and a leading (mostly cleaved) transmembrane domain, a DPBB_1 motif at positions 77 to 164 and “pollen_allerg_1” motif at positions 175 to 253. This is ranked 54th among transcripts (5th in dChip) in the 57K microarray and is the 70th most abundant transcript sequence in MPSS. The sequence has many EST matches, but no principal matches among reported proteome products. The third is LOC_Os12g36040, which is a putative, expressed α -expansin-9 precursor with signal peptide at position 29 (SFA-AD), a DPBB_1 motif at positions 116 to 186 and “pollen_allerg_1” motif at positions 197 to 275. This is ranked 301st among transcripts in the 57K microarray (Supplementary Table S4).

Supplementary Table S4. Pollen expression microarray data on highly pollen-expressed expansins. (Predicted MWs 28.78, 28.99, 30.05 KDa).

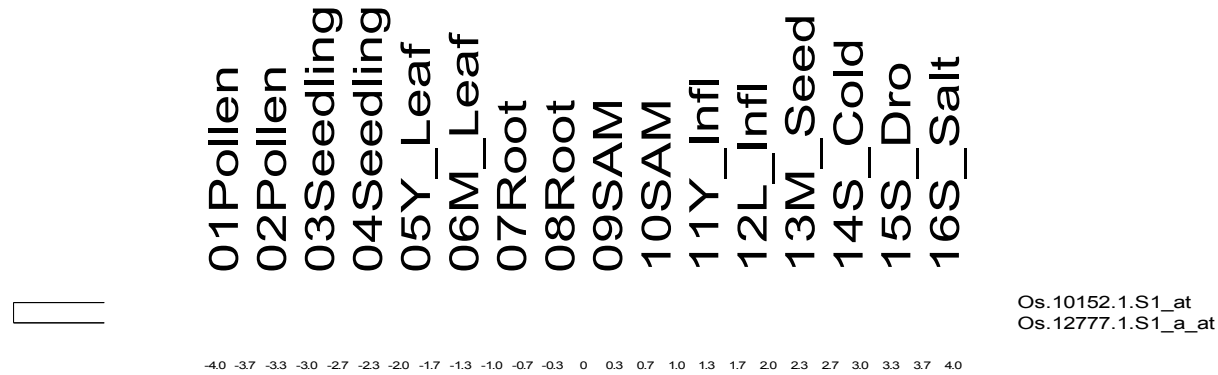
Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (and Pfams)	Intensity (log ₂)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
3.	Os.12697.1.S1_at	Os10g40090.1 (11)	Beta-expansin 1a precursor, putative, expressed (Pfam:PF01357,PF03330) (+ strand with 1 intron)	9.693	13.172	9.62E-10	ESTs: 177	GATCGCTCACCACCAAC (56) GATCAAGAACGTCAACC (18) GATCTCCTGCGGCAACG (243) GATCCTCCTAATTATT(2107)	14a, 15b Dai et al. 2007
54	Os.53024.1.S1_at	Os08g44790.1 (11)	Alpha-expansin 3 precursor, putative, expressed (Pfam:PF01357,PF03330) (+ strand with 2 introns)	7.964	12.214	2.46E-10	ESTs: 18	GATCAGGTACACGATAA (49) GATCATACTAGCTACAA (135)	
301	Os.20191.1.S1_at	Os12g36040.1 (11)	Alpha-expansin 9 precursor, putative, expressed (Pfam:PF01357,PF03330) (+ strand with 1 intron)	8.293	9.908	2.31E-10	ESTs: 7	GATCGATTTTTTCAAC (113)	

6. Profilin A (Ory s 12) (Further Details)

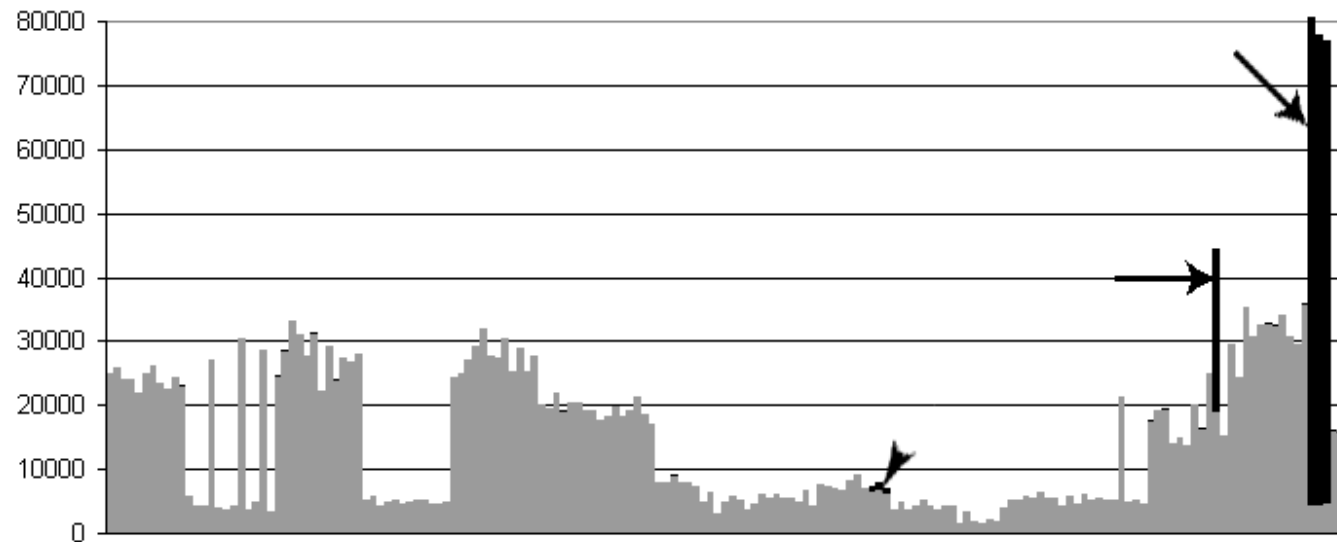
In rice pollen, profilin A transcripts appear to be highly expressed (intensity 8.120 log₂) and upregulated by 11.74 log₂ times (>×2000) above seedlings, according to microarray results. MPSS indicates that transcript abundance is among the top 30 signatures, of which two loci were specific for pollen-expressed profilin. In rice, the two loci active exclusively in pollen, LOC_Os10g17660 and LOC_Os10g17680, are on opposite strands and each of the loci contains a short 5'-UTR (116 bp for LOC_Os10g17660 and 104 bp for LOC_Os10g17680) and a longer 3'-UTR (371 bp for LOC_Os10g17660 and 181 bp for LOC_Os10g17680) with similar overall construction (Supplementary Figure S3) and high sequence similarity. Although six nucleotides differ in their coding sequences, they encode identical polypeptide sequences. Their 1k upstream region displays less conservation than other inverted near duplicates, with the exception of the 0 to -100 presumed promoter region. Strong proteome support exists for translation in pollen grains (Kerim et al., 2003a; Dai et al., 2006) and post-germination pollen tubes (Dai et al., 2007).



Supplementary Figure S3. Diagrammatic representation of TIGR gene models of LOC_Os10g17660 and LOC_Os10g17680 show mirror image organizations and very high sequence similarities. These include extensive similarities in non-coding regulatory sequences.



Supplementary Figure S4. Heat map of profilin (Pfam: PF00235) based on dChip (Li and Wong, 2003; <http://www.biostat.harvard.edu/complab/dchip/>) analysis of CEL data with pollen and seedling data supplemented by GEO data. Genomic data on rice indicate two major types of profilin: two pollen loci of profilin A and a single locus of profilin-2 apparently expressed in all other tissues.



Supplementary Figure S5. Expression of profilin using GEO data from the Affymetrix 57K rice genomic microarray platform. Stacked bars show profilin A in black (probe set Os.12777.1.S1_a_at, representing paired loci LOC_Os10g17660 and LOC_Os10g17680) and profilin-2 in gray (probe set Os.10152.1.S1_at, representing the largely sporophytically-expressed LOC_Os06g05880). Profilin A is detectable in mature inflorescences (arrowhead), and abundantly in collections of anthers at anthesis (horizontal arrow) and especially pollen (diagonal arrow).

In the rice genome there are only three loci devoted to profilin, of which two seem only to be expressed in pollen (Supplementary Figure S4; Supplementary Table S5). All other sporophytic expression seems to correspond to a remaining profilin-2 (LOC_Os06g05880, Supplementary Figure S5). Profilins are typically involved in the sequestration of actin monomers and are highly expressed in pollen, presumably as a protein that modulates polymerization and the dynamic nature of the elaborate actin cytoskeleton of elongating pollen tubes.

Supplementary Table S5. Pollen expression microarray data on Ory s 12 (predicted molecular weight 14.25 KDa).

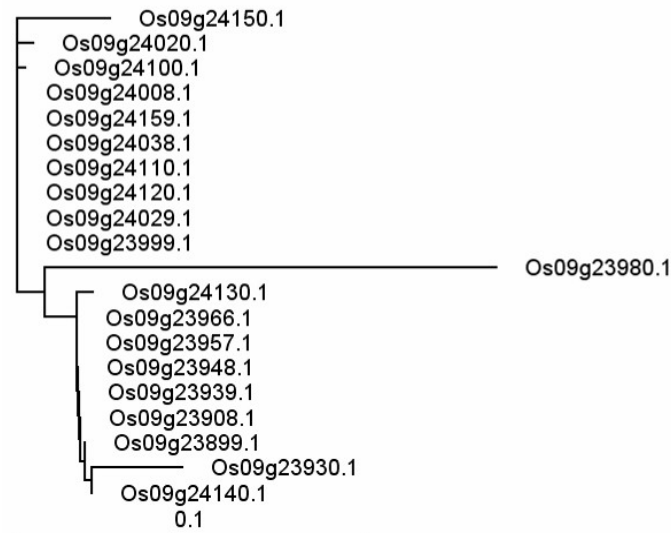
Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (and Pfams)	Intensity (log ₂)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
113	Os.12777.1.S1_a_at	LOC_Os10g17660.1 (11); LOC_Os10g17680.1 (11)	[LOC_Os10g17660 ¹ , Profilin A, putative, expressed (Pfam: PF00235) (+ strand with 2 introns)], [LOC_Os10g17680 ² , Profilin A, putative, expressed (Pfam:PF00235) (- strand with 2 introns)]	8.120	11.737	7.42E-10	ESTs: LOC_Os10g17660 (16) LOC_Os10g17660 (13)	¹ GATCACCGTCGCCATCA (16) ¹ GATCACCTCTGTACGT (172) ¹ GATCCCATTTCTTCATG (5760) ² GATCACTCCTGCAGTG (73) ² GATCACATCCATTGTCG (63) ² GATCAGGAGGCATCAC A (5)	^{1,2} NP_920667, 38b Dai et al. 2006, 2007 Q9FUD1 Kirim et al. 2003

7. Ory s 23 (Further Details)

The peptide sequences of Ory s 23 consist of two distinct forms, differing at positions 48, 79, 81, 82, 128, and 130. The two alternative peptide sequences are highlighted as yellow and light blue in Figure 3. Six single nucleotide polymorphisms are present that code for three single amino acid substitutions in LOC_Os09g24020, LOC_Os09g24100, and LOC_Os09g24130, each of which is annotated as an expressed product. The divergence is narrow (Supplementary Figure S6). All are represented by one probe set which is highly upregulated (Supplementary Table S6).

Interestingly, sequence synonymy also extends to upstream sequences, expanding sometimes to the full 1.4 kbp distance between coding regions of adjacent loci; these typically differ by only a few basepairs despite that this is a non-coding region, with the exception of three upstream sequences that have apparently been displaced. Alignments of coding and non-coding regions are presented in supplementary data. The low sequence variability expressed by Ory s 23 suggests either a very short history for this gene at its current copy number, or a mechanism that has extraordinarily strict sequence conservation. Given that non-coding upstream and UTR sequences provide an almost complete match, with

relatively little variability except along two motifs, the suggestion is that this has a short history and that a faithful copy mechanism that can capture non-coding regions seems a likely explanation. The grouping of different sequence motifs as well as the distance between coding sequences suggests that multiple loci may have been inserted at the same time, potentially in pairs.



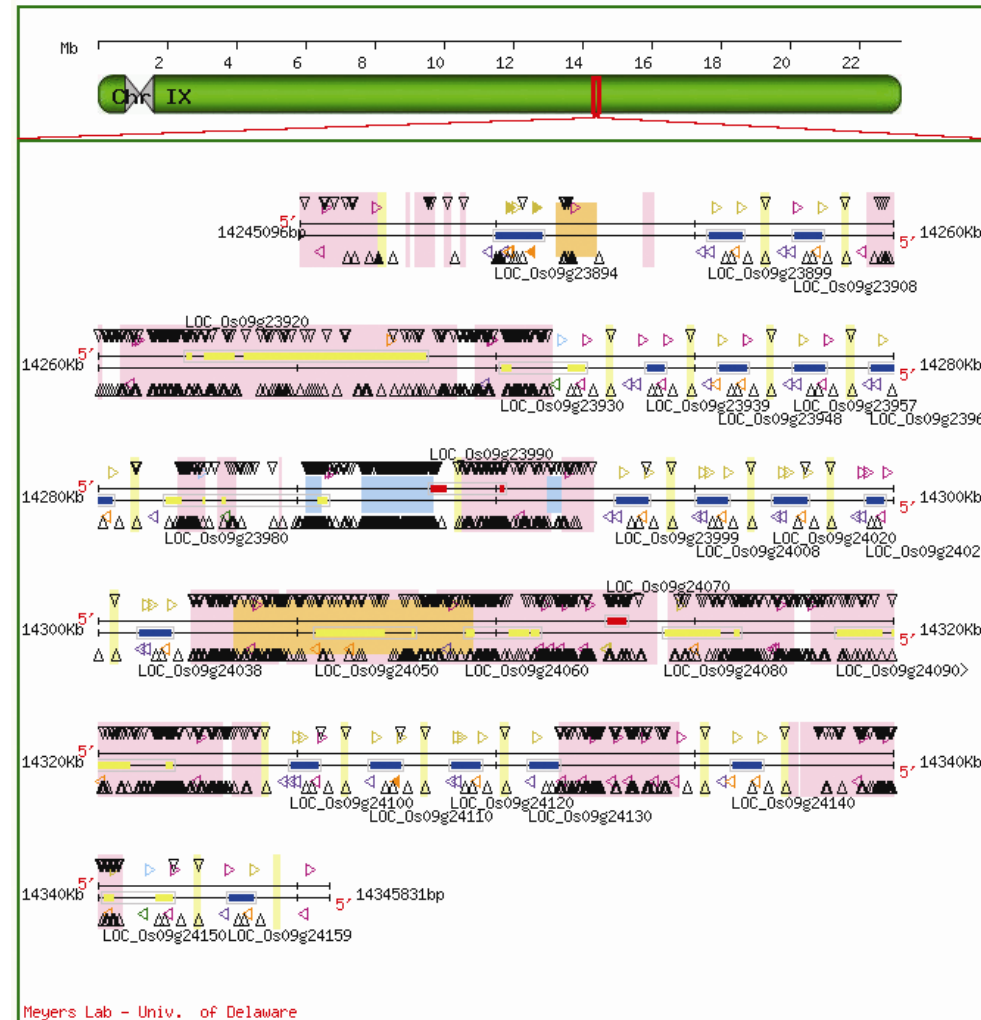
Supplementary Figure S6. Relationship of coding sequences of Ory s 23 variants. LOC_Os09g23980, LOC_Os09g23930, and LOC_Os09g24150 contain core sequences of Ory s 23 that are linked to repeat elements of TEs that may render them to be functionless pseudogenes, especially as upstream regulatory elements. Note the minor degree of divergence of most sequences. Top and bottom groups of non-TE linked sequences differ by only 8 nucleotides of 305 bp.

Transposable element (TE) motifs are linked to three additional nearly complete Ory s 23 sequences (Supplementary Figure S7). One open reading frame has a long leading retrotransposon sequence that constitutes a properly terminated Ory s 23 sequence at LOC_Os09g23980. Two shorter, trailing TE motifs that appear anomalously terminated are connected to reading frames at LOC_Os09g24150, with a 14 bp overlapping transposon motif, and LOC_Os09g23930, with a 17 bp overlapping retrotransposon motif at the end of an otherwise normal Ory s 23 sequence. The anomalous flanking sequences have high homology with TE sequences located on other chromosomes, suggesting multiple independent insertions of the TE sequences, which may have potentially arisen as an inadvertent consequence of TE activity itself (Supplementary Figure S7).

The presence of multiple repetitive elements in broad intervals between tandem Ory s 23 sequences, and short repetitive elements between Ory s 23 sequences in non coding regions suggests that this region may be a target of TE insertion. The alternative of DNA repair genes selecting this region as a target remains a possibility but the presence of multiple TE loci between Ory s 23 tandem repeats implicates action by TEs in their origin. The presence of such sequences may represent pseudogenes which may have been inserted by retrotransposable elements. Wang et al. (2006), for example, unexpectedly observed that retroposition was involved in generating large numbers of intronless genes in rice and provided strong evidence for involvement of RTEs as a mechanism of insertion. RTEs, representing a “copy and paste” transposition, can rapidly increase gene copy numbers and plant genome size, as well as providing, over time, a source of paralog formation (Bennetzen, 2007). The rapid and highly related sequences of Ory s 23 noted here, along with LTR-repeat linked variants make RTE insertion an attractive mechanism for amplification of transcription of this sequence pair.

Supplementary Table S6. Pollen expression microarray data on Ory s 23. (Predicted molecular weight 13.99 KDa).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (with Pfams, orientation) (all are on – strand and intronless)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
6.	Os.17958.1.S1_at	LOC_Os09g23899.1 (10); LOC_Os09g23908.1 (3); LOC_Os09g23930.1 (1); LOC_Os09g23939.1 (1); LOC_Os09g23948.1 (7); LOC_Os09g23957.1 (1); LOC_Os09g23966.1 (8); LOC_Os09g23980.1 (1); LOC_Os09g23999.1 (8); LOC_Os09g24008.1 (8); LOC_Os09g24020.1 (5); LOC_Os09g24029.1 (1); LOC_Os09g24038.1 (6); LOC_Os09g24100.1 (1); LOC_Os09g24110.1 (8); LOC_Os09g24120.1 (8); LOC_Os09g24130.1 (1); LOC_Os09g24140.1 (4); LOC_Os09g24150.1 (1); LOC_Os09g24159.1 (1)	[LOC_Os09g23899, LOC_Os09g23939, LOC_Os09g23948, LOC_Os09g24029, Pollen allergen Ory s 23, putative], [LOC_Os09g23908, LOC_Os09g23957, LOC_Os09g23966, LOC_Os09g23999, LOC_Os09g24008, LOC_Os09g24020, LOC_Os09g24038, LOC_Os09g24100, LOC_Os09g24110, LOC_Os09g24120, LOC_Os09g24130, LOC_Os09g24140, LOC_Os09g24159, Pollen allergen Ory s 23, putative, expressed], [LOC_Os09g23930, Retrotransposon protein, putative, unclassified], [LOC_Os09g23980, Retrotransposon protein, putative, unclassified (Pfam:PF03578[x2])], [LOC_Os09g24150, Transposon protein, putative, unclassified]	9.749	13.068	1.87E-07	ESTs: LOC_Os09g23899 (0) LOC_Os09g23908 (3) LOC_Os09g23930 (0) LOC_Os09g23939 (0) LOC_Os09g23948 (0) LOC_Os09g23957 (2) LOC_Os09g23966 (26) LOC_Os09g23980 (0) LOC_Os09g23999 (5) LOC_Os09g24008 (2) LOC_Os09g24020 (5) LOC_Os09g24029 (0) LOC_Os09g24038 (8) LOC_Os09g24100 (2) LOC_Os09g24110 (5) LOC_Os09g24120 (18) LOC_Os09g24130 (10) LOC_Os09g24140 (9) LOC_Os09g24150 (0) LOC_Os09g24159 (8)	GATCCTTGCACTTGGC (30302) (all) GATCCCTGTCCAGTTCA (16697) (all) GATCGGACACGTACCGA (2346) (blue group)	LOC_Os09g24100: BAD36285 Dai et al. 2006 LOC_Os09g24120: 59b, 60b Dai et al. 2006, 2007 LOC_Os09g24130: BAD36288 Dai et al. 2006



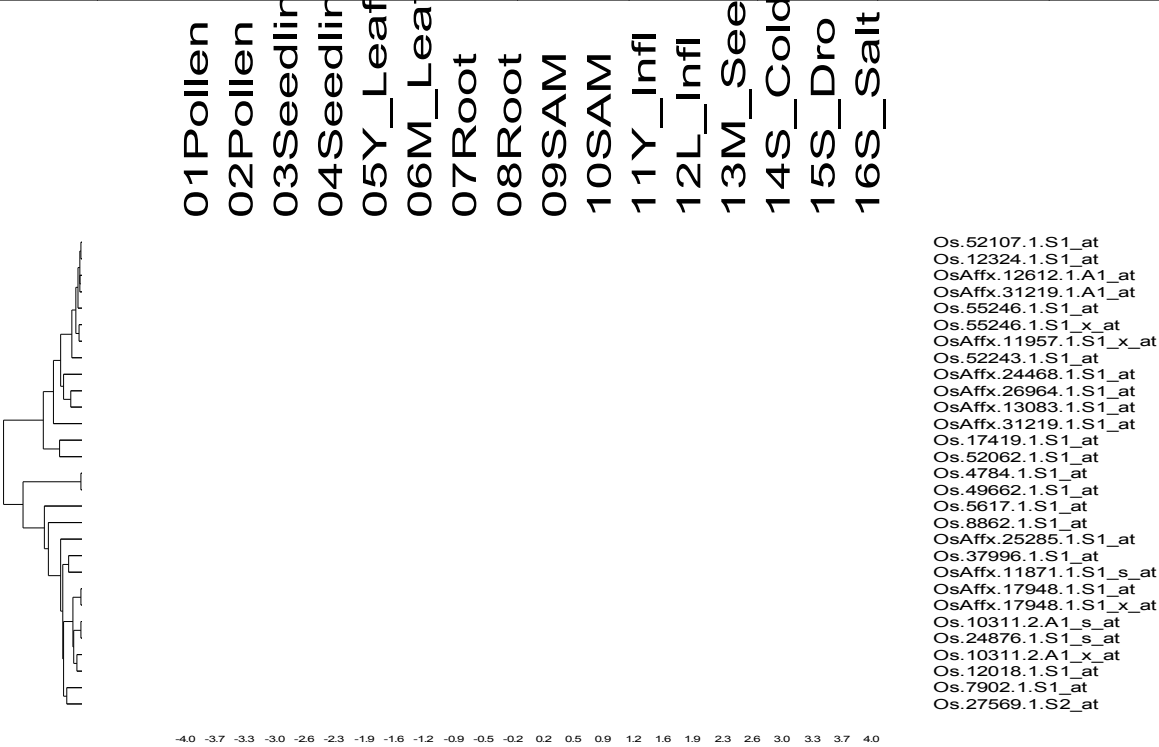
Supplementary Figure S7. Data from Rice Massively Parallel Signature Sequencing (MPSS) and TIGR sequences (Nobuta et al., 2007), showing segment of chromosome 9 containing all 20 copies of the Ory s 23 sequence. Blue regions on the Crick strand represent Ory s 23 loci. Note that Ory s 23 sequences are interspersed with equally spaced light yellow repetitive sequences. Transposable elements are indicated in yellow and orange shaded regions. Upward and downward pointing arrows represent small RNA signatures. (Image downloaded and modified from Meyers Lab site, <http://www.mpss.udel.edu/rice/>).

8. Other Suspected Allergens (Further Details)

a. EF hand family protein Ory s 7

Supplementary Table S7. Pollen expression microarray data on EF hand family protein Ory s 7 (predicted MWs 31.44, 20.09, 18.58 KDa, respectively).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (and Pfams)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
21	Os.12324.1.S1_at	LOC_Os08g44660.1 (4)	EF hand family protein, expressed (Pfam:PF00036 [x2]) (+ strand with 2 introns)	7.894	12.526	2.53E-08	ESTs: 0	GATCGACACCGACGGCG (152)	
331	Os.52107.1.S1_at	LOC_Os12g12730.1 (11)	EF hand family protein, expressed (Pfam:PF00036 [x4]) (- strand and no introns)	7.538	9.671	7.42E-08	ESTs: 2	GATCAGCATGGTGGACG (327)	
4597	Os.52243.1.S1_at	LOC_Os04g45180.1 (11)	EF hand family protein, expressed (Pfam:PF00036, PF02809) (+ strand w/5 introns)	8.664	2.087	6.57E-06	ESTs: 15		

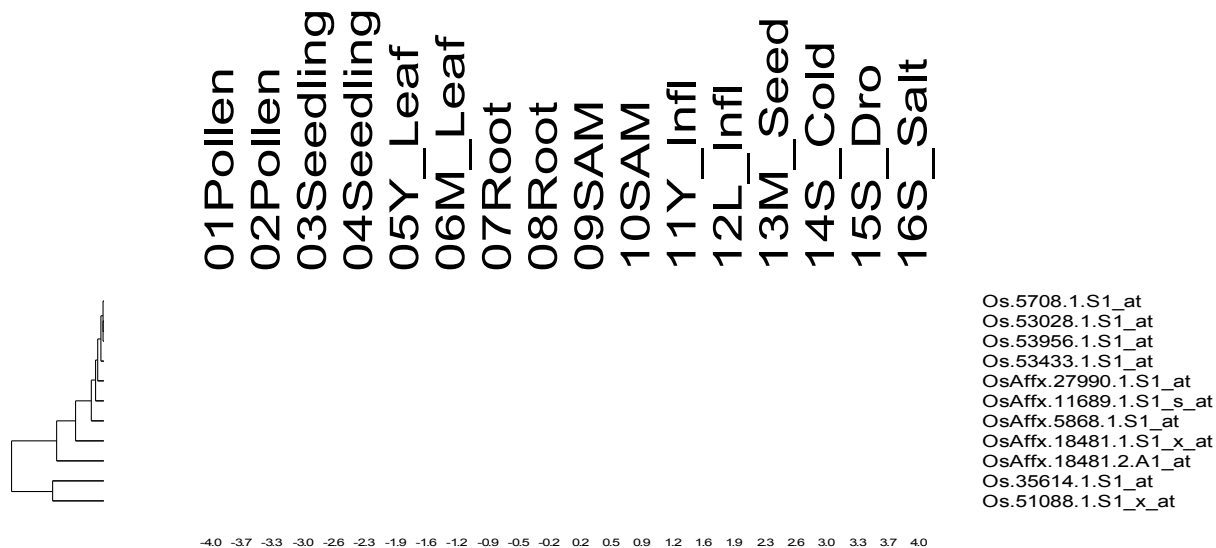


Supplementary Figure S8. Heat map of EF hand-related motifs (Pfam: PF00036) generated based on dChip analysis supplemented by GEO data.

b. Pectate lyase

Supplementary Table S8. Pollen expression microarray data on Pectate lyase (predicted molecular weight 49.23, 50.32, 50.22, 39.07, 50.27, 50.27 KDa, respectively).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (and Pfams)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
85	Os.18224.1.S1_at	LOC_Os06g38510.1 (11)	Pectate lyase precursor, putative, expressed (Pfam: PF00544, PF04431) (+ strand with 2 introns)	8.135	11.923	2.51E-08	ESTs: 87	GATCATCAGCCAGGGGA (2284) GATCAGAAAGAGCCAA (1211)	
209	Os.53956.1.S1_at	LOC_Os02g12300.1 (11)	Pectate lyase precursor, putative, expressed (Pfam: PF00544, PF04431) (+ strand with 3 introns)	8.336	10.754	6.87E-11	ESTs: 55	GATACCAATGACCAAA (76) GATCCGCCGGCGAGATG (14) GATCACCGTGGCGTTCA (32)	
219	Os.53028.1.S1_at	LOC_Os06g05260.1 (11)	Pectate lyase precursor, putative, expressed (PF00544, PF04431) (+ strand with 1 intron)	7.870	10.673	6.86E-09	ESTs: 31	GATCATCAGCCAGGGGA (2284)	
230	Os.5708.1.S1_at	LOC_Os06g05209.1 (11); LOC_Os06g05272.1 (9)	[LOC_Os06g05209, Pectate lyase precursor, putative, expressed (Pfam: PF00544, PF04431)] (- with 1 intron), [Os06g05272, Pectate lyase precursor, putative (Pfam: PF00544)] (+ strand with 1 intron)	7.375	10.536	2.01E-07	ESTs: Os06g05209 (14); Os06g05272 (0)	¹ GATCCGCGACTCCAAGC (20) ¹ GATCGTGAACCACAACA (29) ² GATCCGCGACTCCAAGC (16) ² GATCACCGTGGCGTTCA (32) ¹² GATCATCAGCCAGGGGA (2284)	

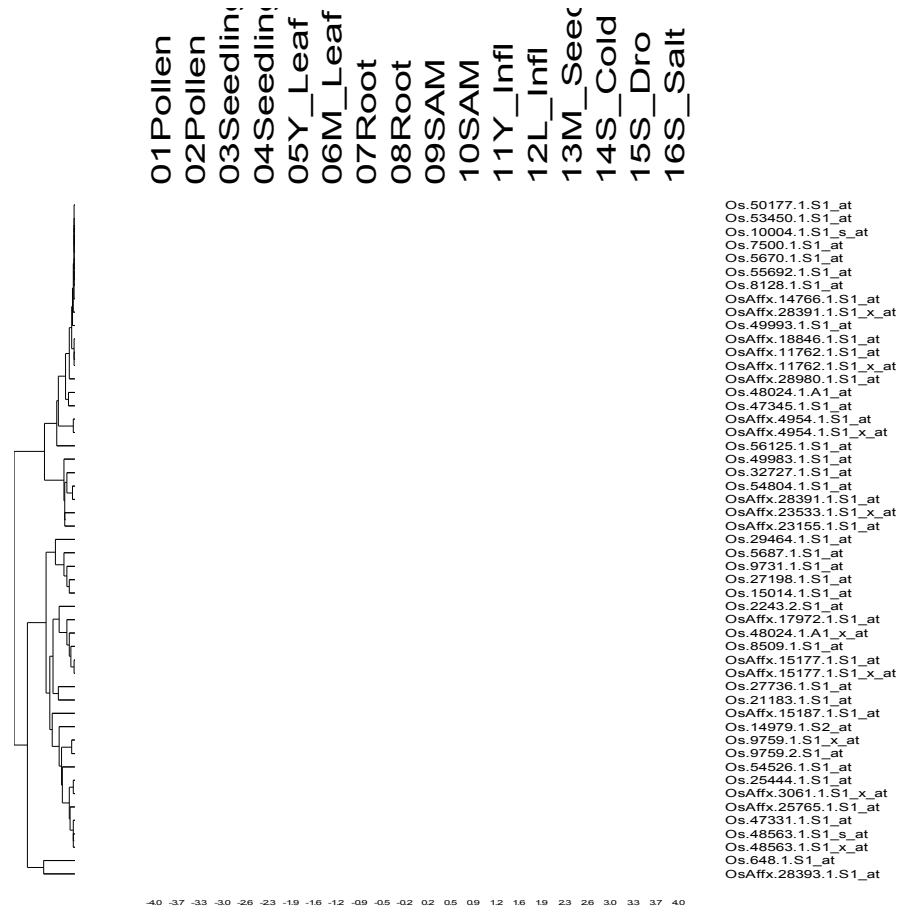


Supplementary Figure S9. Heat map of pectate lyase-related motifs (Pfam: PF00544, PF04431) based on dChip analysis supplemented by GEO data.

c. Glycoside hydrolase family 28: Exopolygalacturonase

Supplementary Table S9. Pollen expression microarray data on Glycoside hydrolase family 28, Exopolygalacturonase (predicted MWs 43.00, 46.34, 35.12, 43.12, 43.12. 44.61, 43.50 KDa, respectively).

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (and Pfams)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
35	Os.7500.1.S1_at	LOC_Os06g35300.1 (2); LOC_Os06g35320.1 (11); LOC_Os06g35370.1 (4)	[LOC_Os06g35300 ¹ , Exopolygalacturonase precursor, putative (Pfam:PF00295)] (- w/1 intron), [LOC_Os06g35320 ² , Exopolygalacturonase precursor, putative, expressed (Pfam:PF00295)] (- w/1 intron), [LOC_Os06g35370 ³ , Exopolygalacturonase precursor, putative, expressed (Pfam:PF00295)] (- w/1 intron)	8.672	12.387	3.21E-09	ESTs: LOC_Os06g35300 (0) LOC_Os06g35320 (27) LOC_Os06g35370 (93)	^{2,3} GATCTGCACTGCCAACG (59) ^{2,3} GATCAACCCACGCTACA (4351) ^{2,3} GATCAAGTCCTATGAGG (40)	1b, 2b Dai et al., 2007
63	Os.10004.1.S1_s_at	LOC_Os02g10300.1 (11)	Exopolygalacturonase precursor, putative, expressed (Pfam:PF00295) (+ w/2 introns)	8.431	12.099	7.03E-08	ESTs: LOC_Os02g10300 (9)	GATCGACAAGGTGACCA (43) GATCAAGTCGTACGAGG (6) GATCATCATCGACCAGA (4194)	XP_464471 13a, 14a Dai et al., 2006, 2007
121	Os.8128.1.S1_at	LOC_Os01g33300.1 (11)	Exopolygalacturonase precursor, putative, expressed (Pfam:PF00295) (- w/2 introns)	7.986	11.691	2.98E-09	ESTs: LOC_Os01g33300 (10)	GATCCATCCATCGTCGA (771) GATCAGCGTGGGGTGCC (11) GATCCCCAAGGGAGAAT (50)	
125	Os.53450.1.S1_at	LOC_Os06g40890.1 (11)	Exopolygalacturonase precursor, putative, expressed (Pfam:PF00295) (+ w/1 intron)	8.577	11.557	1.02E-08	ESTs: LOC_Os06g40890 (21)	GATCAGCGCGTGACCA(20) GATCACCATCGCCGCCA (12) GATCAAGTCGTACGAGG (6)	
800	Os.5670.1.S1_at	LOC_Os08g23790.1 (11)	Exopolygalacturonase precursor, putative, expressed (Pfam:PF00295) (+ w/4 introns)	5.775	6.822	3.13E-06	ESTs: LOC_Os08g23790 (77)	GATCGTTAATTTTTTGC (433)	BAD03446 Dai et al., 2006

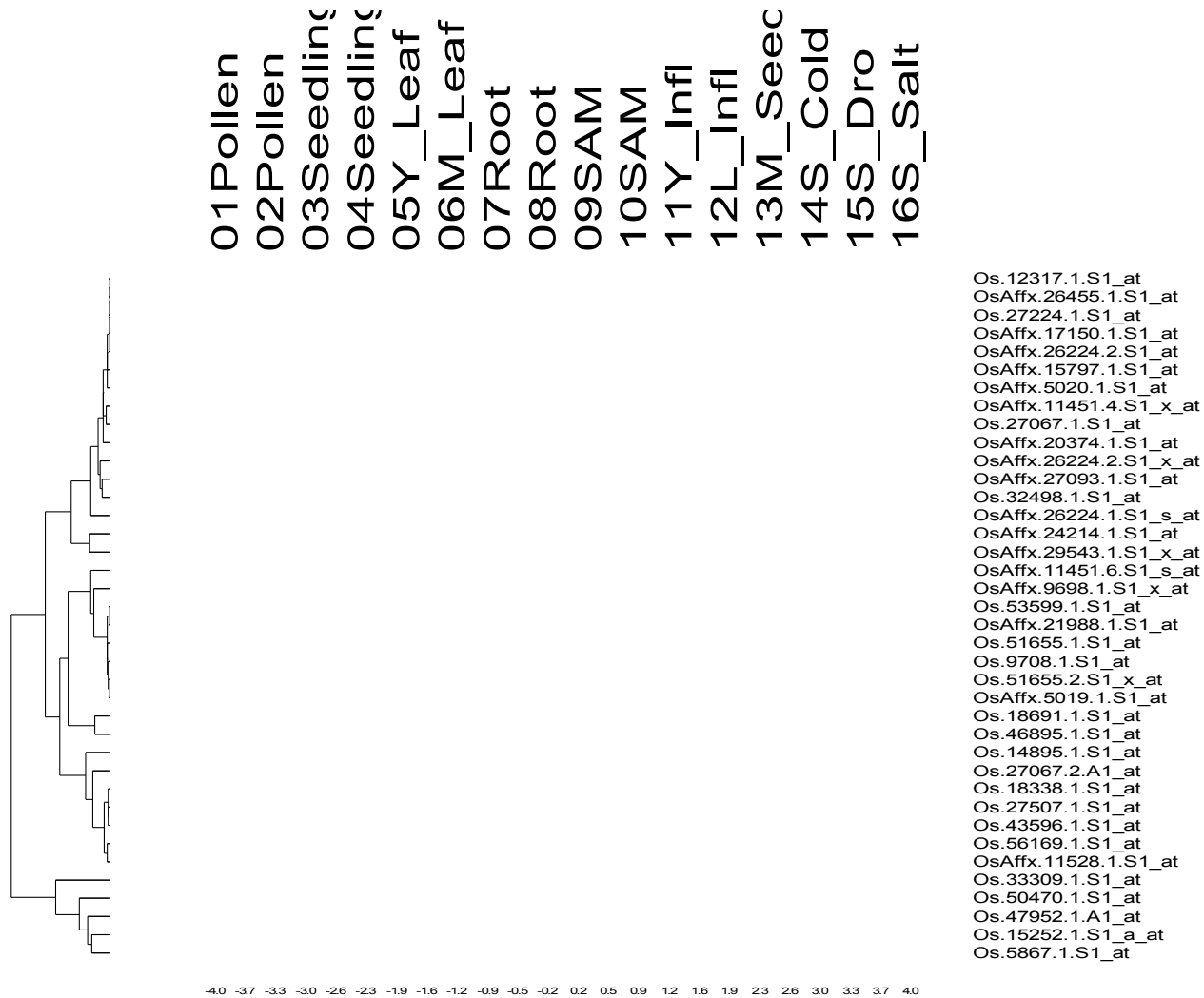


Supplementary Figure S10. Heat map of glycosyl hydrolase family 28 proteins (PF00295), most abundantly represented in rice by exopolysaccharide-related motifs. Image based on dChip (Li and Wong, 2003; <http://www.biostat.harvard.edu/complab/dchip/>) analysis of CEL data with pollen and seedling data supplemented by GEO data.

d. FAD-binding motif (PF01565)

Supplementary Table S10. Pollen expression microarray data on FAD-binding motif (PF01565) (predicted molecular weight 58.43 KDa),

Rank	Probe Set ID	TIGR v5 locus (probe match)	Description (and Pfams)	Intensity (log2)	Log ₂ upreg >seedling	p datasets	Experimental confirmation		
							Gene	MPSS (TPM)	Proteome
33	Os.12317.1.S1_at	LOC_Os06g35590.1 (11)	Reticuline oxidase precursor, putative, expressed (Pfam:PF01565,PF08031) (- strand with no introns)	8.508	12.414	1.94E-08	ESTs LOC_Os06g35590 (12)	GATCAACCAACTAA CGT (804)	BAD54133 Dai et al., 2006



Supplementary Figure S11. Heat map of FAD-binding motif (PF01565), based on dChip (Li and Wong, 2003; <http://www.biostat.harvard.edu/complab/dchip/>) analysis of CEL data with pollen and seedling data supplemented by GEO data.

REFERENCES (ONLY APPEARING IN THIS DATA SUPPLEMENT)

- Bate, N., and Twell, D.** (1998). Functional architecture of a late pollen promoter: Pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Mol. Biol.* **37**, 885-896.
- Rogers, H.J., Bate, N., Combe, J., Sullivan, J., Sweetman, J., Swan, C., Lonsdale, D.M., and Twell, D.** (2001). Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. *Plant Mol. Biol.* **45**, 577-585.