Recent Proliferation and Translocation of Pollen Group 1 Allergen Genes in the Maize Genome[1][W][OA]

Elene R. Valdivia[2,3], Javier Sampedro[2]*, Jonathan C. Lamb, Surinder Chopra, and Daniel J. Cosgrove

Department of Biology (E.R.V., J.S., D.J.C.) and Department of Crop and Soil Sciences (S.C.), Penn State University, University Park, Pennsylvania 16802; and Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211 (J.C.L.)

The dominant allergenic components of grass pollen are known by immunologists as group 1 allergens. These constitute a set of closely related proteins from the β-expansin family and have been shown to have cell wall-loosening activity. Group 1 allergens may facilitate the penetration of pollen tubes through the grass stigma and style. In maize (*Zea mays*), group 1 allergens are divided into two classes, A and B. We have identified 15 genes encoding group 1 allergens in maize, 11 genes in class A and four genes in class B, as well as seven pseudogenes. The genes in class A can be divided by sequence relatedness into two complexes, whereas the genes in class B constitute a single complex. Most of the genes identified are represented in pollen-specific expressed sequence tag libraries and are under purifying selection, despite the presence of multiple copies that are nearly identical. Group 1 allergen genes are clustered in at least six different genomic locations. The single class B location and one of the class A locations show synteny with the rice (*Oryza sativa*) regions where orthologous genes are found. Both classes are expressed at high levels in mature pollen but at low levels in immature flowers. The set of genes encoding maize group 1 allergens is more complex than originally anticipated. If this situation is common in grasses, it may account for the large number of protein variants, or group 1 isoallergens, identified previously in turf grass pollen by immunologists.

During the spring and summer, grasses release large amounts of wind-dispersed pollen, causing hay fever and seasonal asthma in more than 25% of the population in temperate regions (Knox and Suphioglu, 1996a). Allergy symptoms are triggered by the release of allergens when grass pollen comes into contact with the moist surface of the human respiratory tract (Knox and Suphioglu, 1996b). The allergenic proteins of grass pollen have been classified into different groups on the basis of antibody cross-reactivity, physicochemical properties, and sequence relatedness (Knox and Suphioglu, 1996a). Group 1 allergens, along with group 5 allergens, are probably the most widespread inducers of human allergic disease due to the frequency and strength of their immunoglobulin E reactivity (Suphioglu, 2000). It has been estimated that more than 95% of people who suffer from grass pollen allergies have antibodies that recognize group 1 allergens (Bhalla, 2003).

Group 1 grass pollen allergens are now considered part of a superfamily of plant cell wall proteins called expansins. In this nomenclature, group 1 allergens are part of the β-expansin EXPB family, one of four expansin families (Kende et al., 2004). Thus, they may also be called pollen β-expansins. Expansins have the ability to induce rapid extension of isolated cell walls and enhance stress relaxation (Sampedro and Cosgrove, 2005). The first expansins were purified from young cucumber (*Cucumis sativus*) seedlings on the basis of their capacity to induce cell wall extension (McQueen-Mason et al., 1992). When the first expansin sequences were obtained, it was recognized that they shared distant homology to grass group 1 allergens (Shcherban et al., 1995). It was later shown that group 1 allergens from maize (*Zea mays*) indeed had cell wall-loosening activity (Cosgrove et al., 1997).

Based on the analysis of the rice (*Oryza sativa*) genome, it has been proposed that the lineage that gave rise to group 1 allergens (represented in this species by *OsEXPB1*, *OsEXPB9*, *OsEXPB10*, and *OsEXPB13*) originated in monocots through short-range translocation (Sampedro et al., 2005). On the basis of phylogenetic analysis, group 1 allergens have been divided into two rather divergent classes, A and B (approximately 60% identical at the protein level), both of which are present in many different grass species (Li et al., 2003). The origin and subsequent divergence of the two classes

[1] This work was supported by the National Institutes of Health (grant no. GM060397) and the Department of Energy (grant no. DE–FG02–84ER13179 to D.J.C.); the National Science Foundation Division of Biological Infrastructure (grant no. 0423898 to Dr. James Birchler, University of Missouri); and E.R.V. was supported in part by the National Science Foundation (Graduate Research Training grant no. DGE–9354969 to D.J.C.).

[2] These authors contributed equally to the paper.

[3] Present address: Division of Biological Sciences, 304 Life Sciences Center, University of Missouri, Columbia, MO 65211.

* Corresponding author; e-mail jus16@psu.edu; fax 814–865–9131.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Daniel J. Cosgrove (dcosgrove@psu.edu).

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.106.092544

has been linked in rice to an ancient whole-genome duplication event shared by most grasses and predating the divergence of rice and maize (Sampedro et al., 2005). Class A and class B allergens from maize have similar wall extension activity, showing clear specificity for grass cell walls (Li et al., 2003).

Expression of group 1 allergen genes in maize is predominantly restricted to pollen grains (Broadwater et al., 1993; Wu et al., 2001). Mature grass pollen contains high amounts of group 1 allergens and it has been suggested that the corresponding genes are probably expressed late in pollen development, coinciding with the accumulation of storage proteins and carbohydrates (Knox and Suphioglu, 1996b). However, in maize, transcripts for group 1 pollen allergens were detected in very low amounts in uninucleate microspores and subsequently increased greatly after the first pollen mitosis (Broadwater et al., 1993).

On the basis of its pattern of expression and its activity, it was proposed that grass group 1 pollen allergens loosen the walls of the stigma and style to help the elongating pollen tube reach the ovary (Cosgrove et al., 1997). Such a role could explain the unusual abundance, atypical pH optimum (5.5), and high solubility of group 1 allergens, as well as their tendency to cause cell wall breakage, characteristics not seen in expansins involved in other growth processes (Li et al., 2003). Based on the crystal structure of maize group 1 pollen allergens, it has been suggested that these proteins may act by detaching arabinoxylan chains from cellulose microfibrils (Yennawar et al., 2006).

The maize genome is the product of a polyploidization event that happened between 5 and 12 million years ago (Mya), shortly after the split from the sorghum lineage, and was followed by extensive diploidization (Swigonova et al., 2004). Since the duplication event, it has been estimated that at least 50% of duplicate genes have been lost, obscuring microsynteny in the process (Lai et al., 2004). The maize lineage diverged from rice, the only fully sequenced grass genome, about 50 Mya during the early history of the Poaceae family and, although extensive gene collinearity can still be detected, considerable chromosomal rearrangements appear to have taken place (Salse et al., 2004). The maize genome ($n = 10$) contains around 2.4 Gbp, 6 times the size of rice, and at least 60% of the genome is composed of retrotransposons (Paterson et al., 2005). A number of sequencing projects, including sequencing of bacterial artificial chromosome (BAC) ends, methyl filtration, and high-Cot selection libraries, as well as selected entire BACs, in addition to more than one-quarter of a million cDNA sequences, have already provided good representation of maize gene content (Messing and Dooner, 2006). In parallel, a detailed physical map has been built and anchored to the genetic map (Gardiner et al., 2004). In its latest release, July 19, 2005, this map covers 91% of the genome in 721 contigs (http://www.genome.arizona.edu/fpc/maize/#top). The inbred line selected for genomic sequencing as well as the physical map is B73, which was developed by Iowa State University in 1972 (Troyer, 1999).

In this study, we combine the results of our own gene sequencing and localization with publicly available sequence data to construct a census of maize genes encoding group 1 allergens. Whereas all class B genes are highly similar, class A genes in maize can be classified into two differentiated complexes. We also show that maize pollen group 1 allergens are encoded by a large set of genes that has grown through extensive tandem duplication and transpositions throughout the genome.

RESULTS

Maize Contains a Large Number of Group 1 Allergen Genes

To elaborate a census of maize pollen β-expansin genes, we analyzed the genomic and cDNA sequences deposited in GenBank. Due to evidence of variation among cultivars, detailed analysis of maize cDNA sequences was limited to those from libraries obtained from inbred line B73 because they can be directly compared with the available genomic sequences. Also included in the analysis were two genomic sequences obtained by us in the course of this work (GenBank accession nos. DQ525684 and DQ525685). Because B73 is highly inbred and has an extremely low level of heterozygosity (Gethi et al., 2002), we took all confirmed sequence differences to represent separate genes. We ignored, however, all the sequence mismatches that could be explained by sequencing errors or amplification artifacts. With these considerations in mind, we were able to identify a total of 15 genes, 11 of which fall into class A and four into class B (Table I). All of the class B genes and nine of the class A genes are represented in pollen-specific expressed sequence tag (EST) collections. The assembled cDNA sequences cover the entire coding region for all the expressed genes and the same is true for all but one of the genomic sequences (*EXPB10d*). Taking into account that over 200 B73 cDNA sequences are group 1 allergen fragments and that 11 of the 13 expressed genes are also represented in genomic sequences, it is unlikely that many more highly expressed genes remain to be discovered. In addition to active or potentially active genes, we also found genomic sequences for seven pollen β-expansin pseudogenes; that is, genes whose coding region contains stop codons or frame-shifting insertions or deletions. The details of which sequences correspond to each of the genes or pseudogenes can be found in Supplemental Tables S1 to S3 and a detailed explanation of how they were grouped is presented in the Supplemental Text.

In a previous study, four maize pollen β-expansin genes were named on the basis of cDNA sequences (Li et al., 2003). Two of these sequences (class B genes *EXPB1* and *EXPB9*) were obtained from inbred line Kentucky 21. The two other sequences (class A

Table I. *Group 1 allergen genes identified in the maize B73 cultivar*

For each gene, the availability of genomic sequence (asterisks indicate partial sequences), the number of diagnostic ESTs, and the chromosome location, when known, are indicated (see Fig. 3 for detailed localization).

Complex	Gene	Genome	ESTs	Chromosome
Class A				
EXPB10	*EXPB10a*	Yes	5	9
	EXPB10b		2	
	EXPB10c	Yes		3
	EXPB10d	Yes *		3
	ψEXPB10e	Yes		9
EXPB11	*EXPB11a*		14	
	EXPB11b	Yes	11	5
	EXPB11c	Yes	34	5
	EXPB11d	Yes	6	5
	EXPB11e	Yes	19	5
	EXPB11f	Yes	2	
	EXPB11g	Yes	6	5
	ψEXPB11h	Yes		5
	ψEXPB11i	Yes*		5
	ψEXPB11j	Yes*		2
	ψEXPB11k	Yes		
	ψEXPB11l	Yes*		
Class B				
EXPB1	*EXPB1a*	Yes	10	9
	EXPB1b	Yes	15	9
	EXPB1c	Yes	41	9
	EXPB9a	Yes	9	9
	ψEXPB1d	Yes*		

EXPB10 and *EXPB11*) were assemblies of ESTs from several cultivars. In this work, we have adopted a conservative gene-naming system due to the problems raised by differences between maize cultivars, evidence of gene conversion, and incompleteness of the maize genome sequence. In phylogenetic analyses, each of the genes we have identified in inbred line B73 can be confidently placed as a sister sequence to one of the four previously described genes, with the closest B73 genes showing from 0 to 3 nucleotide mismatches with these reference sequences in the coding region (see Supplemental Text). Consequently, we have decided to maintain the four existing names and consider them as lineages. To distinguish individual genes within each lineage, we added a letter to the name (e.g. *EXPB10a*) as we had previously done for two identical rice pollen β-expansin genes (*OsEXPB1a* and *OsEXPB1b*).

The sequence tables show that class A genes are very poorly represented in both high Cot and methyl filtration genomic libraries in comparison with class B genes (Supplemental Tables S1–S3). In the case of the high Cot libraries, which select for unique sequences, this difference could simply be due to the larger number of class A genes. The underrepresentation in methyl filtration libraries could indicate that the regions containing class A genes have high levels of methylation, but this possibility needs to be investigated further. The available genomic sequences also indicate that none of the maize class A genes contain any introns, as had been previously found in rice. On the other hand, maize class B genes have the three introns ancestral to clade EXPB-I, referred to as A, B, and C (Sampedro et al., 2005). In the single class B gene known in rice (*OsEXPB9*), only introns A and C are present (Lee et al., 2001). The most parsimonious explanation for this pattern is that the ancestral pollen β-expansin contained A, B, and C introns, all of which were lost in the class A branch before the splitting of the rice and maize lineages. The loss of intron B in rice *OsEXPB9* happened later, after this lineage had diverged from maize.

A phylogenetic analysis of all the full-length protein sequences of maize pollen β-expansin genes can be seen in Figure 1 (see Supplemental Fig. S1 for DNA trees and pseudogene phylogeny). Whereas all class B genes are very similar (97.7% or higher DNA identity), class A genes can be clearly separated into two distinct branches where the similarity between the genes in separate branches (85.3%–87.2% DNA identity) is considerably lower than that inside each branch (93.7% and 97.7%). The two branches correspond to the previously identified *EXPB10* and *EXPB11* genes. For this reason, we will refer to them in this work as the EXPB10 and EXPB11 gene complexes, whereas all class B genes can be considered members of the EXPB1 complex, as indicated by the shading in Figure 1. In addition to rice, we included in our analyses sorghum, a species closely related to maize that has a large number of pollen cDNA sequences available. There are more than 200 sorghum ESTs for pollen β-expansin genes in GenBank, thus providing a good sampling of the gene diversity in this species. The phylogenetic tree in Figure 1 includes all the sorghum full-length assemblies found in the PlantGDB database (Dong et al., 2005). It clearly suggests that the EXPB10 and EXPB11 complexes diverged very recently, sometime after the maize lineage had split from sorghum, around 12 Mya (Swigonova et al., 2004). The remaining partial sorghum sequences are very similar to the full-length assemblies (94% or more identity) and do not change this conclusion (see Supplemental Fig. S1 and Supplemental Text for details).

One of the most surprising results of the sequence analysis is the large number of nearly identical genes that seem to be under purifying selection (i.e. selection to preserve the protein sequence). This is particularly the case for the EXPB11 complex, where five genes (*EXPB11a–EXPB11e*) produce identical mature proteins and show only synonymous changes outside the signal peptide (Fig. 2A). A similar pattern of purifying selection is seen in the other complexes (Supplemental Figs. S2 and S3). The ratio of pairwise nonsynonymous to synonymous substitutions, excluding the signal peptide, ranges from 0 to 0.15 for class A genes and from 0 to 0.22 for class B genes (Supplemental Fig. S4). In addition to the large number of active genes, the presence of several pseudogenes in the EXPB11 complex is further evidence of the high rate of gene duplication in this lineage. All pseudogenes seem to

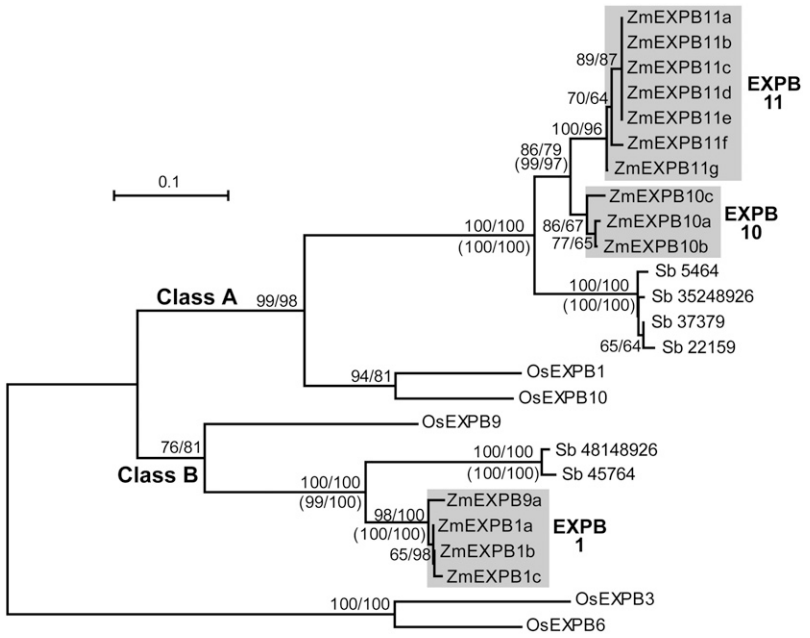

Figure 1. Phylogeny of maize group 1 pollen allergens. The tree shown was obtained by neighbor joining, using the protein sequences, without their signal peptides, of all the full-length group 1 pollen allergens in maize, sorghum, and rice, as well as the two closest rice proteins as an outgroup. A maximum-likelihood tree was identical for all resolved nodes (bootstrap >60 for any method). Bootstrap values are given next to resolved nodes, first for neighbor joining, followed by maximum likelihood. For selected nodes, bootstrap values from the DNA tree in Supplemental Figure S1B are indicated in parenthesis. Sorghum sequences (Sb) are identified by numbers assigned by PlantGDB *Sorghum_bicolor* EST assembly version 154a. Classes A and B are labeled at their base. Maize gene complexes are boxed in gray.

be fairly recent, showing 91.7% to 96.7% DNA identity with expressed genes. The oldest pollen β-expansin pseudogene seems to be *ψEXPB10e*, with 82.6% identity and a large portion of the sequence deleted. In addition to premature stop codons and/or frame-shifting insertions or deletions, most pseudogenes show evidence of relaxed selection (Supplemental Fig. S4).

Another important result of the sequence analysis is evidence of a gene conversion event in the case of *EXPB1c*. An interlocus conversion happens when the sequence of one gene, or part of it, is substituted for that of another related gene. In this case, the conversion tract in *EXPB1c* encompasses between 504 and 767 nucleotides and covers most of the coding sequence. This region is now identical to both *EXPB1a* and *EXPB1b*, the likely donor genes, whereas both the 5′ and 3′ ends show numerous mismatches (Supplemental Fig. S3). An analysis using GENECONV (Padidam et al., 1999) strongly suggests that this pattern could be a product of gene conversion (P value = 1.4×10^{-4}). Even more revealing is the presence in inbred line W23 of what appears to be an unconverted version of *EXPB1c*, with 11 mismatches in the putative conversion tract (Supplemental Fig. S3B). No conversion events could be detected in the other complexes.

Genomic Localization of *EXPB10* Genes

Information regarding the localization of the different pollen β-expansin genes in the maize genome was obtained from three sources: database sequences, BAC library screening, and fluorescence in situ hybridization (FISH). The results are presented in Table I and Figure 3. Both the sequenced BACs and many of the BACs whose ends contain full or partial pollen β-expansin genes are included in the physical map that covers most of the maize genome. This information is detailed in Supplemental Tables S1 to S3. In the case of *EXPB10* genes, a BAC that contains *EXPB10a* has been localized to the long arm of chromosome 9 (9 L), whereas *ψEXPB10e* can be found in three overlapping BACs located in the same region, but at a considerable distance (Fig. 3E). Finally, *EXPB10c* has recently being sequenced from BAC c0129O19 in chromosome 3 (see below).

As part of the maize-mapping project, BAC libraries were hybridized with a large number of short probes derived from ESTs (Gardiner et al., 2004). Among the probes selected were three derived from the cDNA sequences of *EXPB1* (PCO124530), *EXPB10* (PCO110981), and *EXPB11* (PCO068046). We ordered a number of BAC clones that reportedly hybridized to these probes and screened them by PCR and Southern blots. A large number of them turned out to be false positives (Supplemental Table S4). On the other hand, four overlapping BACs were found to contain two genes from the EXPB10 complex, which we then proceeded to sequence (Fig. 2C). One of them, *EXPB10d*, is not represented in the databases by any other sequences. As for *EXPB10c*, we sequenced it in BAC c0154E21, and an identical sequence has recently been obtained from BAC c0129O19, confirming our hypothesis (Fig. 2C). These BACs have been mapped to 3S, the short arm of chromosome 3 (Fig. 3E). Finally, when FISH was performed using an EXPB10 probe, signals were observed on chromosomes 3 S, 5 L, and 9 L, although the signal on 9 L was very faint (Fig. 3C). The signals on 9 L and 3 S confirm the previous results, whereas the signal in chromosome 5 could indicate the location of *EXPB10b* or of other as yet unidentified genes belonging to the EXPB10 complex.

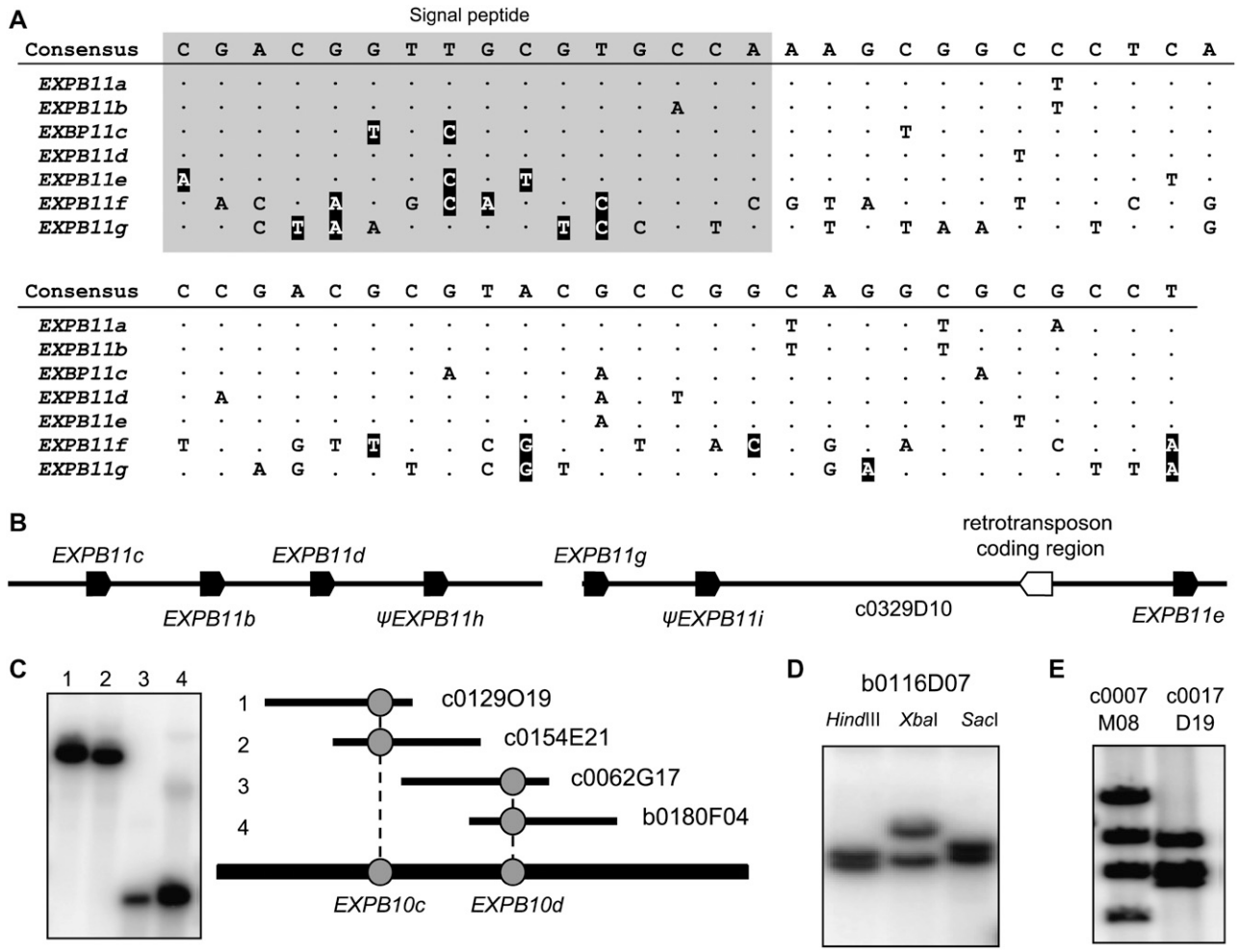

Figure 2. Sequence analysis and cluster arrangements. A, Mismatches between the different maize genes in the EXPB11 complex. Dots represent identical nucleotides. Nonsynonymous changes are underlined in black. The region that codes for the signal peptide region is boxed in gray. B, Annotation of two fragments from BAC c0329D10. Pollen β-expansin genes are shown in black, a retrotransposon in white. C, Southern blot of four BAC clones, where two *EXPB10* genes were located. Clones were digested with *Hin*dIII and *Pst*I. Blots were hybridized with class A-specific probes (Table II). The four BAC clones overlap as shown in the drawing to the right, where gray circles stand for *EXPB10* genes. The identity of the genes was confirmed through sequencing. D, Southern blot of BAC b0116D07, digested with *Hin*dIII, *Xba*I, and *Sac*I. Blots were hybridized with class B-specific probes. E, Southern blot of two BAC clones that overlap with each other and with BAC c0329D10, shown in B. Clones were digested with *Hin*dIII and *Pst*I. Blots were hybridized with class A-specific probes (Table II).

Automated analysis of maize-rice synteny, based on BAC ends and probe hybridization data (Soderlund et al., 2006), detected extensive conservation of gene order between maize chromosome 9 and rice chromosome 3, suggesting that the location of *ψEXPB10e* could be orthologous to that of the single cluster of class A genes in rice. The region surrounding *ψEXPB10e* has recently been sequenced, which allowed us to analyze in detail the conservation of microsynteny (Fig. 4A). We found that the order and orientation of the genes in this small region have been perfectly preserved since the separation of the maize and rice lineages. The main differences are the proliferation of class A genes in the rice lineage and the distribution of the maize genes over a much larger span than their rice counterparts. As for the other *EXPB10* genes, the genomic location of *EXPB10a* seems to be orthologous to a region in rice chromosome 3 hundreds of genes away from the class A genes, whereas the location of *EXPB10c* and *EXPB10d* on 3 S appears to be orthologous to rice chromosome 1, where no class A genes are found (see Supplemental Text for details). Both locations are therefore likely to be products of independent transposition events. This is indicated by the use of white triangles in Figure 3E.

Genomic Localization of *EXPB11* Genes

In the case of the EXPB11 gene complex, five active genes and two pseudogenes are located in close

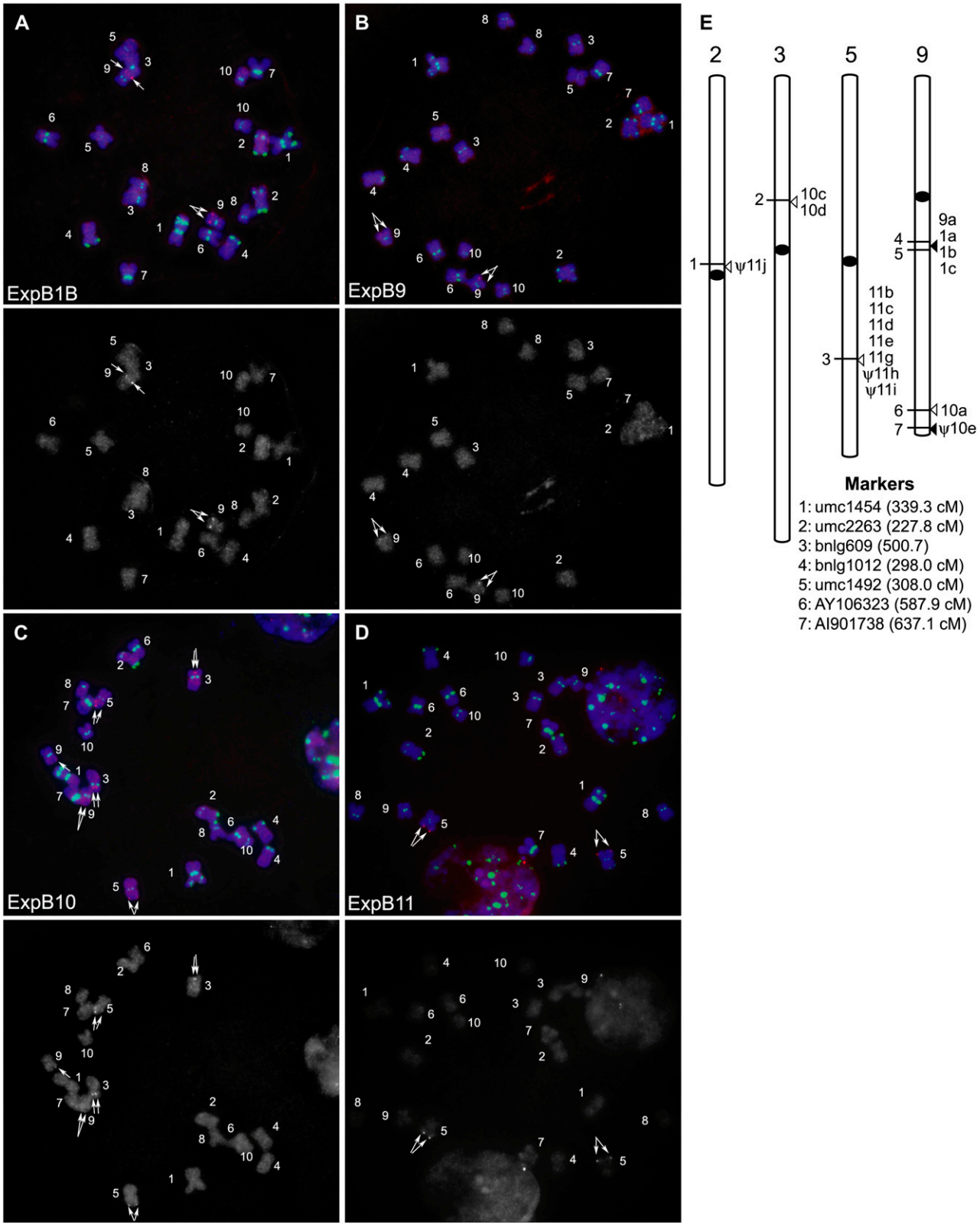

Figure 3. Localization of expansin gene clusters on maize chromosomes. Chromosome spreads from the B73 inbred line were hybridized with probes specific to *EXPB1b* (A), *EXPB9* (B), *EXPB10* (C), and *EXPB11* (D). The expansin probes (red) were hybridized simultaneously with a cocktail of probes that allows each chromosome to be identified in this inbred line and counterstained with DAPI (blue). The cocktail contains CentC (green), TAG microsatellite clone 1-26-2 (green), and the 180-bp

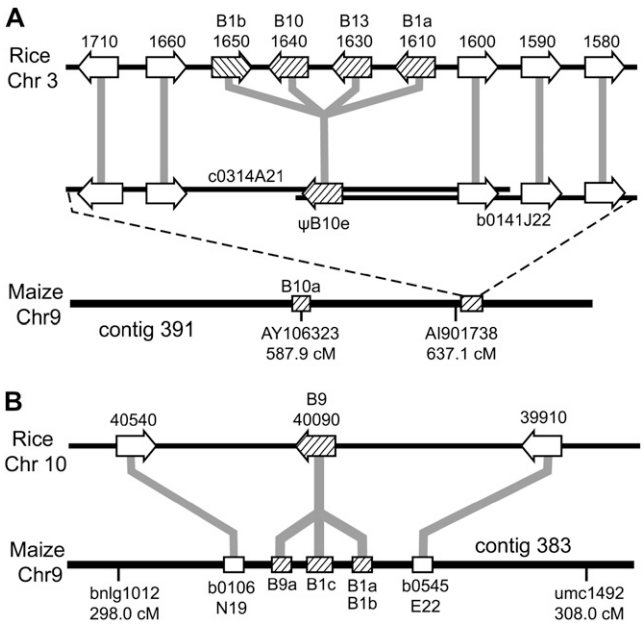

Figure 4. Synteny between maize group 1 allergen locations in rice and maize. A, Microsynteny of class A group 1 allergen genes (hatched arrows) and surrounding genes (white). Rice genes are identified by The Institute for Genomic Research annotation number. Putatively orthologous genes, as indicated by BLAST searches, are connected by gray lines. The third line shows the location of maize BACs b141J22 and c0314A21 in reference to selected markers (vertical lines), as well as that of a second BAC, where another maize *EXPB10* gene was found. B, Synteny of class B gene locations. Hatched boxes are BAC clones that contain complex *EXPB1* genes; white boxes are BACs whose end sequences include gene fragments putatively orthologous to connected rice genes.

proximity in BAC c0329D10, which has been partially sequenced (Fig. 2B). This BAC has been mapped to chromosome 5 (Fig. 3E). During the BAC library screening, we also detected this same group of genes through Southern hybridization in a series of BACs that overlap each other and also BAC c0329D10 (Fig. 2E). This result was further confirmed by FISH hybridization with an EXPB11 probe, which only produced a signal on the long arm of chromosome 5 (Fig. 3D). This location is orthologous to rice chromosome 2 and thus appears to represent another case of translocation.

An end sequence for BAC b0259K07, mapped to chromosome 2, is the only evidence for *ψEXPB11j*. On the other hand, this sequence is almost identical (576 of 581 bp) to *ψEXPB11i*, one of the pseudogenes in the aforementioned chromosome 5 BAC. The current limit of reliable detection using FISH is approximately 3 kb (Kato et al., 2006; Wang et al., 2006) and a single copy of an *EXPB11* pseudogene located on 2 S would not be detected. It is also possible that the position of this BAC was misassigned. At least two more *EXPB11* genes and two other pseudogenes have not yet been localized. It is possible that they are also part of the cluster of *EXPB11* genes in 5 L because the sequencing of c0329D10 is incomplete and that of surrounding BACs has not started.

Genomic Localization of *EXPB1* Genes

Two genes in this complex that have been mapped through genomic sequencing, *EXPB9a* through the end sequence of BAC b413H10 and *EXPB1c* in BAC c0454G13, are both located in close proximity on chromosome 9 (Figs. 3E and 4A). As part of the BAC library screening, we detected the presence of two other genes from this complex in BAC b0116D07 (Fig. 2D). Direct sequencing of this BAC, which is also located in the same region of chromosome 9, resulted in a hybrid sequence with double peaks where *EXPB1a* and *EXPB1b* have mismatches, indicating that all four known class A genes are located in the same genomic region (Fig. 4B).

In the FISH experiments, probes against both *EXPB1a* and *EXPB1c* genes produced similar results, hybridizing to the long arm of chromosome 9 (Fig. 3A). The same result was obtained with a probe designed against *EXPB9a* (Fig. 3B). This information confirms that all of the class B genes in maize are likely located in a single cluster in chromosome 9, although distributed over a considerable distance. In the case of class B genes, automated synteny analysis shows that the region of chromosome 9 where maize *EXPB1* genes are located is orthologous to the section of rice chromosome 10 that contains *OsEXPB9*, the single class B gene in this species. However, there is not yet enough information to determine how well microsynteny is conserved (Fig. 4B).

Analysis of Group 1 Allergen Expression

Using real-time PCR, we studied the expression of the three complexes of group 1 allergen genes both in immature tassel flowers (approximately 10 and 12 d before anthesis) and mature pollen (Fig. 5). Expression of the three gene complexes in immature flowers was very low, with only slight differences between the samples collected 10 and 12 d before anthesis, which correspond to the late uninucleate microspore stage (Mascarenhas, 1989). In contrast, in mature pollen, the levels of expression are approximately 10^6-fold higher,

Figure 3. (*Continued.*)
knob probe (blue). Gray values of the EXPB probe signals are displayed below the merged (RGB) image. Arrows indicate sites of hybridization to the EXPB probes. E, Maize chromosomes that contain group 1 allergen genes are represented to scale proportional to genetic distance. Black triangles indicate ancestral gene locations showing microsynteny with rice group 1 allergens (see Fig. 4); white triangles are for locations where group 1 allergen genes have been translocated; black circles indicate centromeres. The closest markers to each location are shown.

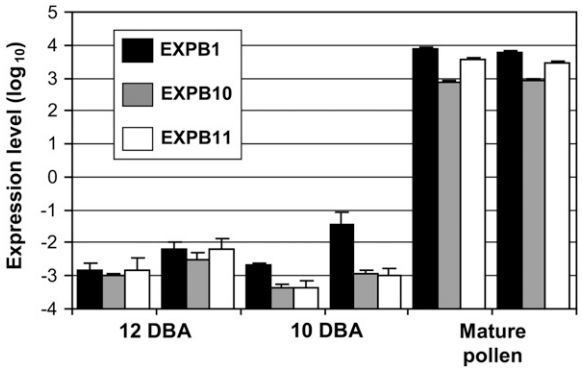

Figure 5. Expression levels of the three complexes of maize group 1 allergen genes relative to GADPH. RNA was extracted from two sibling plants at three developmental stages: immature tassel flowers at 12 and 10 d before anthesis (DBA) and mature pollen. For flower samples, two independent RNA extractions were obtained from each plant. Means and SDs of three amplifications (six in the case of flower samples) are shown.

indicating that these genes are only activated in full during the last phase of pollen development. A previous study had already found, using an *EXPB9* probe, weak expression of maize group 1 allergen genes before the first pollen mitosis, at the uninucleate microspore stage, followed by much stronger expression at later stages (Broadwater et al., 1993). We have shown that the same pattern seems to apply to both class A and class B and to all three maize gene complexes. These results are consistent with the existing hypothesis that the main function of group 1 allergens is to aid in the penetration and growth of the pollen tube through the maternal tissue (Cosgrove et al., 1997; Li et al., 2003). However, the expression of these genes in immature pollen suggests that they could also play a role in pollen development. As for the relative levels of expression of the three complexes, the number of ESTs deposited in databases suggests that they may be considerably lower in the case of EXPB10 (Table I). Although affected by amplification efficiency, the level of expression we detected in mature pollen for the EXPB10 complex was 4.3 and 8.2 times lower than those of EXPB11 and EXPB1, respectively (Fig. 5).

Conserved Elements in Group 1 Allergen Promoters

The availability of genomic sequences from two highly divergent grasses offers the opportunity of identifying conserved motifs in the upstream sequences of rice and maize pollen β-expansins. It has been shown that this region is sufficient to determine pollen-specific expression in one of the rice allergens (Xu et al., 1999). We were able to analyze the promoter sequence of all four maize class B genes, as well as seven class A genes (two *EXPB10*s and five *EXPB11*s), and compared them with their rice orthologs. All these genes are expressed in pollen, with the possible exception of *EXPB10c*, which has not yet been detected in any EST libraries. Because the genes share a common ancestor, one could also expect them to share some cis-regulatory sequences. We searched for potential regulatory elements of at least 8 bp, conserved in similar positions in rice and maize genes, using FootPrinter (Blanchette and Tompa, 2003). Promoter sequences

Table II. *Sequence of primers and probes used in this work*

Specificity	Primer 1	Primer 2	Probe	Template
BAC Screening Primers				
Class A	CGGGCATCATTGACATG	GGATGACATCGTCGTAGACG		
Class B	GGGCAGAAGATCGTGTTCC	GTGTAGACGGCATCGGGTC		
Complex EXPB10	GGTCACCTTCCACTTGG	GGCAGTGTTGGGCTTC		
Complex EXPB11	GCTGACAATACTTTAAGCCG	GACAACACAACATTCATGG		
BAC Screening Probes				
Class A	CGTGTCGGAGCCCTTCTC	GCGCAAGGCGGGC		
Class B	GACAAGCTTCGCCACTGC	CTCCAGTTCGCCGGG		
Quantitative RT-PCR				
Complex EXPB1	TCGGCTCCCTGGCAAAG	TCGCACCCTTCTGAACTCG	CGCCACTGCGGCATCATGGAC	
Complex EXPB10	GGTGTGGTCCTCCCAAGGTT	GGCATCTAGCCAATCACTACCAT	CCCCGGGTAAGAACATCACAGCCA	
Complex EXPB11	TGCGATAAGCCTGTGGAGTG	GGCTCATAGTTCATGTCCGTGA	CGGCAAGCCCGTGGTGGTG	
GADPH	TGACAGCAGGTCGAGCATCT	TTGACGAAGTGGTCGTTCAGAG	CGACGCCAAGGCCGGGATC	
FISH Probes				
EXPB1A	CTATCGCTCCCTACCACTTCGA	GCATACATGATTAGAATTGCCTCCTA		EXPB1a
EXPB1B	CTATCGCTCCCTACCACTTCGA	CACTCTTTGGAATTCGATCATGAA		EXPB1c
EXPB9	CTATCGCTCCCTACCACTTCGA	AACATTAGTCCCTAGAAACCAAACACC		EXPB9a
EXPB10	GCCCGTGGTGGTATACATTACAGAC	CCTCCTCTTCAACGCGTTCC		EXPB10a
EXPB11	CGGGCATCATCGACATG	GCGGTAGACCTCCTTCC		EXPB11a

evolve at a very fast rate so that the probability of a motif of this size being conserved by chance between these two species is less than 0.001 (Kaplinsky et al., 2002).

With the exception of *ZmEXPB9*, for which we only have a short upstream fragment, all the analyzed rice and maize upstream sequences include one or two RY repeats (CATGCATG), perfectly conserved in most cases (Supplemental Fig. S4). This element is bound by the B3 domain of several transcription factors involved in seed maturation (Reidt et al., 2000; Braybrook et al., 2006). It is possible that a related transcription factor is active in grass pollen and involved in the transcriptional control of pollen β-expansins. Our results also suggest that the conservation of regulatory sequences between species is more extensive for class B genes, with at least four 8-bp motifs conserved in the same order and orientation than for class A, where we could only detect conservation of the RY repeats (Supplemental Fig. S5). Due to their recent divergence, the motifs shared by maize *EXPB10* and *EXPB11* genes may have little functional significance.

DISCUSSION

It has long been known that in some grass species, group 1 pollen allergens consist of many protein variants with small differences in pI and molecular size (Petersen et al., 1993). Whether this isoform diversity is due to multiple genes or variations in post-translational modifications cannot be resolved without identifying the corresponding genes. The cloning of cDNA sequences encoding group 1 allergens from grass pollen began in the early 1990s (Perez et al., 1990; Broadwater et al., 1993; Xu et al., 1995), but gave few hints of the large genetic complexity we have discovered in maize. On the other hand, N-terminal sequencing of timothy grass group 1 allergens showed the existence of at least four slightly different protein sequences, all of which appear to belong to class A (Petersen et al., 1993). A previous study in maize identified only two class A and two class B genes (Li et al., 2003) and the recent sequencing of the rice genome revealed that this species had a single class B gene and also a tandem of three genes and one pseudogene belonging to class A (Sampedro et al., 2006). In contrast to these numbers, this work reveals the presence of at least 15 group 1 allergen genes in maize, at least 13 of which are simultaneously expressed in mature pollen (Table I). From the limited evidence we have collected for sorghum and the results of immunological studies, it would not be surprising to also find large numbers of group 1 allergen genes in other grass species. These results have implications for our understanding of the function and evolution of group 1 allergens. Additionally, they need to be taken into account, for example, in attempts to silence the expression of these genes, as was done for the group 5 allergens (Bhalla et al., 1999).

Genomic Localization

We have shown that class B genes are still found in regions of collinear gene order in rice and maize (Fig. 4). For class A genes, on the other hand, we have detected four translocation events from an ancestral location now occupied by *ψEXPB10e* (Fig. 3). However, the fact that this pseudogene can be confidently placed in the EXPB10 complex (Supplemental Fig. S1A) indicates that the oldest of these translocation events (barring long-distance gene conversion) did not happen before the divergence of the EXPB10 and EXPB11 complexes. Our phylogenetic analysis shows that this divergence is more recent than the split of maize from sorghum (Fig. 1). It is thus likely that class A genes in the last ancestor of these two species were found in their ancestral location, which has also been conserved in the rice lineage (Fig. 4).

The last common ancestor of rice and maize is also an ancestor to all but a few basal grasses (Kellogg, 2001). In particular, it is also the ancestor of turf grasses, such as *Lolium perenne*, *Poa pratensis*, *Phleum pratense*, and *Cynodon dactylon*, whose pollen is most frequently responsible for human allergic disease. The evidence that class A and class B gene locations have been preserved in the maize and rice lineages for most of their history suggests a strategy for isolating orthologous genes in turf grasses and other species by positional cloning because there are already detailed comparative genetic maps that cover many grass groups (Devos, 2005). It should also be noted that the translocation events we have detected in maize appear to be quite exceptional in the expansin superfamily at large. In the rice lineage, since its divergence from Arabidopsis (*Arabidopsis thaliana*), we only found six possible cases of translocation out of a total of 58 genes (Sampedro et al., 2005).

There is no evidence that the recent polyploidization of maize had any effect on the proliferation and diversification of pollen β-expansins. We found only one location for class B genes and none of the class A locations appear to be part of recently duplicated segments because they are orthologous to different regions of the rice genome. The ancestral class A location in chromosome 9 is part of a large region duplicated in chromosome 1. This section of chromosome 1 has been fully sequenced and does not contain any pollen β-expansins (Bruggmann et al., 2006). On the other hand, the timing of the divergence of the EXPB10 and EXPB11 complexes, shortly after the split of maize and sorghum, is probably compatible with the polyploidy event, but this possibility would require further translocations of *EXPB11* genes (Swigonova et al., 2004).

Selection for Intron Size and GC Content of Pollen β-Expansins

Pollen β-expansin genes in maize are expressed at high levels, representing 4% of the protein solubilized from mature pollen (Li et al., 2003). Like all genes expressed in the haploid phase, they are expected

to be under particularly intense purifying selection (Bernasconi et al., 2004; Seoighe et al., 2005), especially because expansins are likely to be directly involved in pollen competition (Valdivia et al., 2007). This selection could affect not only protein sequence, but also intron size and base composition via their effects on transcription and translation.

It has been shown that highly expressed genes in several species tend to have shorter introns, probably due to the time and energy required for their transcription (Castillo-Davis et al., 2002). In Arabidopsis, it has been found that genes expressed in pollen have shorter introns than those expressed in the sporophyte, most likely due to the effect of stronger selection (Seoighe et al., 2005). In this context, it is notable that class A genes in both rice and maize lack introns. Moreover, the 12 introns found in the four class B maize genes are all very short (66–83 bp; see Supplemental Fig. S3), both when compared to the 123-bp median size of the introns in rice expansins or the 166-bp median size of introns for maize genes in general (Haberer et al., 2005). These observations are consistent with the idea that selection has operated on the maize pollen β-expansin genes to reduce intron size.

Selection at the translational level has also been proposed as an explanation for the very high GC content at the third codon position of many grass genes (Wong et al., 2002). All pollen β-expansins have high GC content at the third codon positions, but class B genes appear to have particularly high levels (85.5%–87% in maize, 91% in rice) when compared with class A genes (70%–75% for maize EXPB10s, 81%–83% for EXPB11s, and 71%–84% for rice). It has previously been shown that grass genes with higher GC content have a lower rate of synonymous substitution (Alvarez-Valin and Jabbari, 1999). This correlation, which could be caused by selection or other mechanisms, seems to apply to the two classes of pollen β-expansins and is evident in phylogenetic trees. Those trees based on DNA distances show much shorter branches for the GC-rich class B than for class A (Supplemental Fig. S1), a difference that is much reduced when the trees are based on protein sequences (Fig. 1). This suggests that the differential GC enrichment of the two classes could be an ancestral characteristic maintained through evolutionary history, maybe caused by a different degree of selective pressure for efficient or rapid translation. The unequal rate of evolution is also likely to be the reason why trees based on DNA sequence can have difficulty recognizing class B as a clade (Supplemental Fig. S1B).

Multiplicity of Pollen β-Expansin Genes

One of the most remarkable findings of this work is the presence in maize of multiple gene copies that are expressed simultaneously, with minor or no differences at the protein level, and all of them apparently under purifying selection, particularly in the case of the genes in the EXPB11 complex (Fig. 2). The expansin superfamily has been studied in rice, Arabidopsis, and Populus, and in these species the level of identity found inside the maize gene complexes is only seen in groups of two genes (Sampedro et al., 2005, 2006). Recent duplications appear to be more common in the maize genome, but a recent survey found evidence of copies with higher than 98% identity for only 1% of maize genes (Emrich et al., 2007).

The evidence for purifying selection in the pollen β-expansin duplicates suggests that each of these genes provides a selective advantage for the pollen grains that carry them, even though their function is likely to be identical. A possible explanation is that having multiple copies of group 1 allergen genes allows higher levels of total group 1 allergen expression to be reached. The apparently recent proliferation of class A and, to a lesser extent, class B genes in maize could be linked to a particularly strong selection for high levels of pollen β-expansin in this lineage. Maize pollen grains have to grow through silks that can be up to 30 cm long, which could cause the levels of pollen β-expansin to be particularly critical for male-to-male competition. This idea is supported by reduced transmission through pollen of a β-expansin insertional mutant allele (Valdivia et al., 2007).

We should also consider the possibility that the phenomenon of gene proliferation in this group of genes is older than phylogenetic trees indicate. The existence of a cluster of class A genes in rice, together with the evidence from sorghum and the immunological results in other species, suggests that this tendency to duplicate may be a common characteristic of pollen β-expansins in grasses, particularly for class A genes. The appearance of recent proliferation in multiple lineages could thus be the result of concerted evolution. Both unequal crossing over through a process of repeated gene births and deaths, as well as gene conversions, can have homogenizing effects on clusters of closely related genes. We have seen clear evidence of a gene conversion event in the case of the EXPB1 complex and the numerous pseudogenes show how gene copies are being lost in all the lineages. If homogenization is a common occurrence in these genes, then long-range transposition (or the maize polyploidy) may have provided the opportunity for the divergence and possible specialization of the EXPB10 and EXPB11 complexes. A similar situation of specialization associated with physical distance has been described for Arabidopsis resistance genes, which also appear in clusters and are subject to frequent homogenization (Baumgarten et al., 2003).

The degree of functional differentiation in the grass group 1 allergens still needs to be addressed. We have found no evidence of differential expression in the three complexes (Fig. 5), but it seems likely that at least class A and class B genes have separate functions in view of their considerable sequence divergence. One possibility is that they act on different substrates within the wall. The question of the functional specialization of the EXPB10 and EXPB11 complexes is still open because their divergence is very recent and they still

show a considerable degree of sequence conservation at the promoter level.

The number of genes and pseudogenes arranged in clusters, together with the evidence of very recent duplications, show that maize pollen β-expansins have large potential for rapid genetic change. In the course of this work, we have found evidence in maize line B73 of gene conversion in *EXPB1c* and the insertion of a mobile element in the 5′-untranslated region of *EXPB9a* (see Supplemental Text), two events that are not seen in any of the other lines where sequence information is available. Not only the sequences, but also the number of genes and possibly even the locations, are likely to vary in different maize lines. For all these reasons, some caution is needed when applying to other lines the results we have obtained with B73.

Establishing the sequences and numbers of the genes that code for the two classes of group 1 allergens should facilitate further research into their biological functions. Moreover, the situation in maize is likely to apply in some degree to other species and it should serve as a caution about the potential genetic complexity underlying these proteins in turf grasses, where their allergenic properties are of significant medical concern.

MATERIALS AND METHODS

Sequence Analysis

EST sequences from B73 libraries, as well as maize (*Zea mays*) genomic sequences available in GenBank as of November 2, 2006 (Nonredundant, Genome Survey Sequence, and High-Throughput Genomic Sequence databases), were analyzed and grouped into genes using SeqMan (DNAstar). Sorghum EST assemblies were obtained from the PlantGDB database (www.plantgdb.org).

Phylogenetic Analysis

Sequences were aligned with the help of ClustalW, as implemented in MEGA3.1 (Kumar et al., 2004). Neighbor-joining phylogenetic trees were obtained with the same program and maximum-likelihood trees with PHYML 2.4.4 (Stéphane and Gascuel, 2003). For protein trees, the Jones, Taylor, Thornton model with γ-distributed rates (γ parameter of 1.16, as estimated by PHYML) was used for both methods. For DNA trees, substitution models were chosen using MODELTEST (David and Keith, 2001), as implemented in Findmodel (http://hcv.lanl.gov/content/hcv-db/findmodel/findmodel.html). For neighbor joining, K2P with γ-distributed rates (γ parameters of 0.822, 0.574, and 0.975 for complete, class A, and class B trees) was selected. All neighbor-joining trees were constructed using complete deletion of alignment gaps. The maximum-likelihood tree was obtained using the General Time Reversible model with a γ parameter of 0.708 and empirical base frequency estimates. Bootstrap values were based on 500 replicates for all trees.

BAC Clone Identification

Cell stocks for BAC clones were grown in 500 mL of Luria-Bertani broth with 20 μg/mL of chloramphenicol and BAC DNA was purified using NucleoBond BAC Maxi plasmid purification kit (CLONTECH Laboratories). BAC clones were initially screened by PCR using class-specific primers (Table II). PCR fragments were used for direct sequencing using an ABI Hitachi 3730XL DNA analyzer.

Southern Blot

Twenty micrograms of DNA from each BAC were digested with *Hin*dIII, *Xho*I, and *Eco*RV at 37°C overnight. DNA was precipitated and separated on 0.8% agarose gels in 1× Tris-acetate EDTA and then transferred using a Vacugene Xl vacuum-blotting system (Pharmacia Biotech, Amersham-Pharmacia) to a Hybond-N nylon membrane (Amersham Biosciences). The membrane was prehybridized using Ultrahyb solution (Ambion) overnight at 37°C. Hybridizations were performed at 68°C for 10 h in a mixture containing the probe added to the prehybridization buffer. Blots were washed twice at 62°C for 20 min in 5× SEN (5% SDS, 1 mM EDTA, 40 mM Na_2HPO_4) and later for 20 min in 1× SEN (1% SDS, 1 mM EDTA, 40 mM Na_2HPO_4), also twice. Blots were exposed to phosphor screens (Molecular Dynamics) for 4 h at room temperature and imaged using a Typhoon scanner and Image Quant TL software (GE Healthcare).

Two PCR-amplified fragments from the coding region of *EXPB11* and *EXPB1* genes were used as class-specific probes (Table II). These probes were labeled with ^{32}P-dCTP using the rediprime II random prime labeling system (Amersham-Pharmacia). Gene-specific probes and PCR fragments were cloned with a TOPO TA cloning kit (Invitrogen).

Real-Time PCR

Immature tassel flower and pollen tissue were collected from B73 inbred plants grown at the Rock Springs Agricultural Experiment Station at Rock Springs, PA, in 2004. Immature tassel flowers (anthers) were collected from these plants approximately 10 and 12 d before pollen shed (stage V15). Mature pollen was collected at the time of pollen shed (stage VT). Flowers and pollen were collected in the field and immediately frozen in liquid nitrogen and stored at −80°C. Total RNA was extracted using Concert Plant RNA reagent (Invitrogen) and cDNA was prepared, after DNase treatment, using the ABI high capacity RT kit (Applied Biosystems). Amplification and quantification were done on the ABI 7300 real-time PCR system (Applied Biosystems). Primers and probes are listed in Table II. Cycling conditions were as follows: 2 min at 50°C, 10 min at 95°C followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. Relative abundance of each gene, normalized to the GADPH transcript level, was determined using the $\Delta\Delta$CT method (Applied Biosystems). Probes were synthesized by Biosearch Technologies and were labeled with the reporter dye 6-carboxyfluorescein (6′-FAM) at the 5′ end and with the quencher dye Black Hole Quencher at the 3′ end. To eliminate the possibility of DNA contamination, control reactions were run with RNA samples where the cDNA synthesis step was not carried out.

FISH Protocol

FISH was performed according to published protocols (Kato et al., 2004, 2006). Expansin probes were made from PCR products amplified using the primers and templates listed in Table II. PCR-amplified DNA was labeled with Texas Red-dUTP using a nick translation protocol with a high concentration of polymerase I (Kato et al., 2006). To identify the 10 chromosomes in the B73 line, a cocktail of probes was used simultaneously with the expansin probes. The cocktail contained CentC (labeled with AlexaFluor 488-dUTP; Molecular Probes), the 1-26-2 TAG microsatellite probe (AlexaFluor 488-dUTP), and the 180-bp knob repeat (Coumarin-dUTP). Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) containing the antifade solution Vectashield (Vector Laboratories) diluted 1:20 in Vectashield without DAPI. Probe concentrations were 15 to 30 ng/μL each in a total volume of 5 μL per slide. Images were captured with an Olympus BX61 microscope using Applied Spectral Imaging software and a CCD Cool-1300QS camera. Images were then processed by using the sharpen feature of ASI software. To remove background haze, images were copied, severely blurred using the Gaussian blur feature of Adobe Photoshop CS, and then subtracted from the original image. When merged color images are presented, the digital contrast was further adjusted so that red signals can be more easily distinguished. Signals from the single loci probes are also displayed in gray values after only minimal digital manipulation to ensure that printed images accurately reflect the captured images.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers DQ525684 and DQ525685.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Nucleotide-based phylogenies.

Supplemental Figure S2. Class A gene sequence analysis.

Supplemental Figure S3. Class B gene sequence analysis.

Supplemental Figure S4. Pairwise ratios of nonsynonymous to synonymous substitutions for maize genes.

Supplemental Figure S5. Promoter analysis of maize and rice genes.

Supplemental Table S1. GenBank accession numbers of sequences diagnostic of genes and pseudogenes in the EXPB10 complex.

Supplemental Table S2. GenBank accession numbers of sequences diagnostic of genes and pseudogenes in the EXPB11 complex.

Supplemental Table S3. GenBank accession numbers of sequences diagnostic of genes and pseudogenes in the EXPB1 complex.

Supplemental Table S4. Screened BAC clones that do not contain pollen β-expansin genes.

Supplemental Text. Detailed description of how the available sequences were grouped in genes or excluded as artifacts.

ACKNOWLEDGMENTS

We thank Edward Wagner and Daniel M. Durachko for expert technical assistance; Dr. Deborah Grove for design of the primers and probes and running the Taqman assays at the Nucleic Acid Facility (Huck Institute, Pennsylvania State University); and R. Oberheim and the Department of Horticulture for use of the Agricultural Experimental Station at Rock Springs, PA.

Received November 3, 2006; accepted December 29, 2006; published January 12, 2007.

LITERATURE CITED

Alvarez-Valin F, Jabbari K (1999) Synonymous and nonsynonymous substitutions in genes from Gramineae: intragenic correlations. J Mol Evol **49:** 330–342

Baumgarten A, Cannon S, Spangler R, May G (2003) Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics **165:** 309–319

Bernasconi G, Ashman TL, Birkhead TR, Bishop JD, Grossniklaus U, Kubli E, Marshall DL, Schmid B, Skogsmyr I, Snook RR, et al (2004) Evolutionary ecology of the prezygotic stage. Science **303:** 971–975

Bhalla PL (2003) Genetic engineering of pollen allergens for hayfever immunotherapy. Expert Rev Vaccines **2:** 75–84

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

Blanchette M, Tompa M (2003) FootPrinter: a program designed for phylogenetic footprinting. Nucleic Acids Res **31:** 3840–3842

Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc Natl Acad Sci USA **103:** 3468–3473

Broadwater AH, Rubinstein AL, Chay CH, Klapper DG, Bedinger PA (1993) Zea mI, the maize homolog of the allergen-encoding Lol pI gene of rye grass. Gene **131:** 227–230

Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, Pontaroli AC, Wei F, Haberer G, Fuks G, Du C, et al (2006) Uneven chromosome contraction and expansion in the maize genome. Genome Res **16:** 1241–1251

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002) Selection for short introns in highly expressed genes. Nat Genet **31:** 415–418

Cosgrove DJ, Bedinger P, Durachko DM (1997) Group I allergens of grass pollen as cell wall-loosening agents. Proc Natl Acad Sci USA **94:** 6559–6564

David P, Keith AC (2001) Selecting the best-fit model of nucleotide substitution. Syst Biol **50:** 580–601

Devos KM (2005) Updating the ''crop circle''. Curr Opin Plant Biol **8:** 155–162

Dong Q, Lawrence CJ, Schlueter SD, Wilkerson MD, Kurtz S, Lushbough C, Brendel V (2005) Comparative plant genomics resources at PlantGDB. Plant Physiol **139:** 610–618

Emrich SJ, Li L, Wen TJ, Yandeau-Nelson MD, Fu Y, Guo L, Chou HH, Aluru S, Ashlock DA, Schnable PS (2007) Nearly identical paralogs (NIPs): implications for maize (Zea mays L.) genome evolution. Genetics **175:** 429–439

Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanafey M, et al (2004) Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol **134:** 1317–1326

Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S (2002) SSR variation in important U.S. maize inbred lines. Crop Sci **42:** 951–957

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, et al (2005) Structure and architecture of the maize genome. Plant Physiol **139:** 1612–1624

Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M (2002) Utility and distribution of conserved noncoding sequences in the grasses. Proc Natl Acad Sci USA **99:** 6147–6151

Kato A, Albert PS, Vega JM, Birchler JA (2006) Sensitive fluorescence *in situ* hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. Biotech Histochem **81:** 71–78

Kato A, Lamb JC, Birchler JA (2004) From the cover: chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA **101:** 13554–13559

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol **125:** 1198–1205

Kende H, Bradford K, Brummell D, Cho HT, Cosgrove D, Fleming A, Gehring C, Lee Y, Queen-Mason S, Rose J, et al (2004) Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol **55:** 311–314

Knox B, Suphioglu C (1996a) Environmental and molecular biology of pollen allergens. Trends Plant Sci **1:** 156–164

Knox RB, Suphioglu C (1996b) Pollen allergens: development and function. Sex Plant Reprod **9:** 318–323

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform **5:** 150–163

Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park YJ, Jeong OY, Bennetzen JL, et al (2004) Gene loss and movement in the maize genome. Genome Res **14:** 1924–1931

Lee Y, Choi D, Kende H (2001) Expansins: ever-expanding numbers and functions. Curr Opin Plant Biol **4:** 527–532

Li L-C, Bedinger PA, Volk C, Jones AD, Cosgrove DJ (2003) Purification and characterization of four beta-expansins (Zea m 1 isoforms) from maize pollen. Plant Physiol **132:** 2073–2085

Mascarenhas JP (1989) The male gametophyte of flowering plants. Plant Cell **1:** 657–664

McQueen-Mason S, Durachko DM, Cosgrove DJ (1992) Two endogenous proteins that induce cell wall expansion in plants. Plant Cell **4:** 1425–1433

Messing J, Dooner HK (2006) Organization and variability of the maize genome. Curr Opin Plant Biol **9:** 157–163

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology **265:** 218–225

Paterson AH, Freeling M, Sasaki T (2005) Grains of knowledge: genomics of model cereals. Genome Res **15:** 1643–1650

Perez M, Ishioka GY, Walker LE, Chestnut RW (1990) cDNA cloning and immunological characterization of the rye grass allergen Lol pI. J Biol Chem **265:** 16210–16215

Petersen A, Becker WM, Schlaak M (1993) Characterization of grass group I allergens in timothy grass pollen. J Allergy Clin Immunol **92:** 789–796

Reidt W, Wohlfarth T, Ellerstrom M, Czihal A, Tewes A, Ezcurra I, Rask L, Baumlein H (2000) Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. Plant J **21:** 401–408

Salse J, Piegu B, Cooke R, Delseny M (2004) New in silico insight into the synteny between rice (Oryza sativa L.) and maize (Zea mays L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J **38:** 396–409

Sampedro J, Carey RE, Cosgrove DJ (2006) Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs. J Plant Res **119:** 11–21

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol **6:** 242

Sampedro J, Lee Y, Carey RE, dePamphilis C, Cosgrove DJ (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. Plant J **44:** 409–419

Seoighe C, Gehring C, Hurst LD (2005) Gametophytic selection in Arabidopsis thaliana supports the selective model of intron length reduction. PLoS Genet **1:** e13

Shcherban TY, Shi J, Durachko DM, Guiltinan MJ, McQueen-Mason SJ, Shieh M, Cosgrove DJ (1995) Molecular cloning and sequence analysis of expansins—a highly conserved, multigene family of proteins that mediate cell wall extension in plants. Proc Natl Acad Sci USA **92:** 9245–9249

Soderlund C, Nelson W, Shoemaker A, Paterson A (2006) SyMAP: a system for discovering and viewing syntenic regions of FPC maps. Genome Res **16:** 1159–1168

Stéphane G, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:** 696–704

Suphioglu C (2000) What are the important allergens in grass pollen that are linked to human allergic disease? Clin Exp Allergy **30:** 1335–1341

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res **14:** 1916–1923

Troyer AF (1999) Background of U.S. hybrid corn. Crop Sci **39:** 601–626

Valdivia ER, Wu Y, Li L-C, Cosgrove DJ, Stephenson AG (2007) A group-1 grass pollen allergen influences the outcome of pollen competition in maize. PLoS ONE **2:** e154

Wang CJR, Harper L, Cande WZ (2006) High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. Plant Cell **18:** 529–544

Wong GK, Wang J, Tao L, Tan J, Zhang J, Passey DA, Yu J (2002) Compositional gradients in Gramineae genes. Genome Res **12:** 851–856

Wu Y, Meeley RB, Cosgrove DJ (2001) Analysis and expression of the alpha-expansin and beta-expansin gene families in maize. Plant Physiol **126:** 222–232

Xu H, Goulding N, Zhang Y, Swoboda I, Singh MB, Bhalla PL (1999) Promoter region of Ory s 1, the major rice pollen allergen gene. Sex Plant Reprod **12:** 125–126

Xu H, Theerakulpisut P, Goulding N, Suphioglu C, Singh MB, Bhalla PL (1995) Cloning, expression and immunological characterization of Ory s 1, the major allergen of rice pollen. Gene **164:** 255–259

Yennawar NH, Li LC, Dudzinski DM, Tabuchi A, Cosgrove DJ (2006) Crystal structure and activities of EXPB1 (Zea m 1), a beta-expansin and group-1 pollen allergen from maize. Proc Natl Acad Sci USA **103:** 14664–14671