J Plant Res (2006) 119:11–21
Digital Object Identifier (DOI) 10.1007/s10265-005-0253-z

JPR SYMPOSIUM

Javier Sampedro · Robert E. Carey · Daniel J. Cosgrove

Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs

Received: September 30, 2005 / Accepted: November 29, 2005 / Published online: January 13, 2006

Abstract Expansins comprise a superfamily of plant cell wall-loosening proteins that has been divided into four distinct families, EXPA, EXPB, EXLA and EXLB. In a recent analysis of *Arabidopsis thaliana* and *Oryza sativa* expansins, we proposed a further subdivision of the families into 17 clades, representing independent lineages in the last common ancestor of monocots and eudicots. This division was based on both traditional sequence-based phylogenetic trees and on position-based trees, in which genomic locations and dated segmental duplications were used to reconstruct gene phylogeny. In this article we review recent work concerning the patterns of expansin evolution in angiosperms and include additional insights gained from the genome of a second eudicot species, *Populus trichocarpa*, which includes at least 36 expansin genes. All of the previously proposed monocot-eudicot orthologous groups, but no additional ones, are represented in this species. The results also confirm that all of these clades are truly independent lineages. Furthermore, we have used position-based phylogeny to clarify the history of clades EXPA-II and EXPA-IV. Most of the growth of the expansin superfamily in the poplar lineage is likely due to a recent polyploidy event. Finally, some monocot-eudicot clades are shown to have diverged before the separation of the angiosperm and gymnosperm lineages.

Key words Expansin evolution · Arabidopsis · Rice · *Populus trichocarpa* (Poplar) · Genome duplication · Microsynteny

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s10265-005-0253-z and is accessible for authorized users.

J. Sampedro · R.E. Carey · D.J. Cosgrove (✉)
Department of Biology, Pennsylvania State University, 208 Mueller Lab, University Park, PA 16802, USA
Tel. +1-814-8633892; Fax +1-814-8659131
e-mail: dCosgrove@psu.edu

Introduction

One of the characteristics of land plants (Embryophyta) is the presence of a cellulosic cell wall that envelops the cell protoplasts and determines the mechanical properties of plant tissues. Plant cell walls contain not only cellulose but several other polysaccharides as well, which are modified through the action of numerous enzymes and other proteins secreted by the cell (Fry 1995). Expansins are one of these proteins, and they have been shown to loosen the cell walls in unique ways (Cosgrove 2000). Expansins are able to weaken the linkage between cellulose microfibrils, thereby allowing extension of the wall through polymer creep in response to the mechanical forces generated by cell turgor pressure. They are implicated in the control of plant cell growth, cell separation, wall dissolution, pollen-tube invasion of the stigma and in other processes involving cell-wall modifications (Rose and Bennett 1999; Lee et al. 2001; Cosgrove et al. 2002; Li et al. 2003). The mechanism of cell-wall enlargement must be compatible with the complex and subtle patterns of growth and morphogenesis in land plants, and expansin-mediated wall enlargement fulfills these requirements.

The first expansins were identified by McQueen-Mason et al. (1992) through the isolation of proteins responsible for rapid pH-dependent wall extension. Under current nomenclature rules these first expansins are considered to be part of the EXPA or α-expansin family (Kende et al. 2004). Members of a second family of expansins, referred to as β-expansin or EXPB, were first identified as the major allergens of grass pollen and only later were discovered to have cell wall-loosening activity characteristic of expansins (Cosgrove et al. 1997). Two smaller families, expansin-like A (EXLA) and expansin-like B (EXLB), have been identified from genomic analyses, but there is as yet no experimental evidence as to their activity on the cell wall. The expression patterns of some rice members of the EXLA family have been studied, but these do not suggest any particular physiological function (Lee and Kende 2002).

The four expansin families have a similar structure, with two distinct domains preceded by a signal peptide (Cosgrove 2000). Only proteins with this two-domain structure are considered to be members of the expansin superfamily. Apart from the four families known in plants, expansin-like proteins with the two characteristic domains have been found in *Dictyostelium* and some fungi and bacteria (Li et al. 2002; Darley et al. 2003). The evolutionary relationship of these diverse sequences to plant expansins is unclear. They are currently referred to as family EXLX or expansin-like X, a temporary "catch-all" group for non-plant expansins (Kende et al. 2004).

The expansin superfamily in Arabidopsis and rice

The completion of the genome sequencing of Arabidopsis and rice has allowed us to realize the full size and complexity of the expansin superfamily in these two highly divergent angiosperms. The genome of *Arabidopsis thaliana* (L.) Heynh ecotype Columbia (The Arabidopsis Genome Initiative 2000) includes 36 expansin genes, two of which are pseudogenes (*ΨAtEXPA19* and *ΨAtEXPB6*). These genes are divided among the four families as follows: 26 EXPA, six EXPB, three EXLA and one EXLB (Table 1). Even larger is the expansin superfamily found in the map-based sequence of *Oryza sativa* L. ssp. *japonica* cv. Nipponbare (International Rice Genome Sequencing Project 2005), which contains 58 expansin genes, including one pseudogene (*ΨAtEXPB13*) and two pairs of genes (*OsEXPA23a*/*OsEXPA23b* and *OsEXPB1a*/*OsEXPB1b*) that are 100% identical at the nucleotide level. The largest family is again EXPA, with 33 genes, and the smallest families are EXLA, with four genes, and EXLB, with a single sequence (Table 1). In contrast, the EXPB family, which has 18 genes, is threefold larger in rice than in Arabidopsis.

Traditional phylogenetic analyses of Arabidopsis and rice expansins showed a very complex picture, with many gene duplications occurring in the independent lineages of these plants (Lee et al. 2001; Li et al. 2002, 2003b). These studies encountered problems due to the low resolution of the gene phylogeny and could only resolve a few orthologous groups of genes below the family level, particularly in the case of the EXPA family, which is not only larger but also much more diverse than the others.

Position-based phylogeny

In order to clarify the evolutionary relationships of the Arabidopsis and rice expansins, we recently complemented traditional phylogenetic methods by making use of the history of the genome segments in which the expansin genes are located (Sampedro et al. 2005). This was possible because most expansin gene duplications in the Arabidopsis and rice lineages are due either to tandem duplications or to segmental duplications, and in both cases the genomic neighborhood of the expansin gene is preserved. Gene translocations appear to be unusual in this superfamily (Sampedro et al. 2005). Segmental duplications can usually be detected through the conservation of gene order and orientation – also referred to as microcollinearity or microsynteny. What makes the reconstruction of the history of the genomic segments easier is the fact that in both Arabidopsis and rice most of the detected segmental duplications appear to be part of whole genome duplication events.

Two detailed studies of the Arabidopsis genome have identified and dated numerous segmental duplications (Simillion et al. 2002; Bowers et al. 2003). In both studies the bulk of these duplications were attributed to three rounds of polyploidy based on the relative datings and overlapping age distributions. The intervening time between these three genome duplications is long enough that each pair of duplicated segments can usually be assigned with confidence to a particular polyploidy event. In the present work we follow the nomenclature of Bowers et al. (2003) in referring to these duplication events as α, β and γ, with α being the most recent and γ the oldest. Similarly, in the case of the rice genome, large duplicated segments covering most of the genome have been linked to a polyploidy event that predates the divergence of the grasses (Paterson et al. 2004).

Using this information as a starting point we were able to trace the history of most of the genomic segments in which expansins are found, in both rice and Arabidopsis, to the last ancestor of these two species. We were able to do this by finding microsynteny between groups of Arabidopsis and rice segments, which seems to be more extensive than previously thought. Since the history of the segments and the history of the genes they contain are expected to be similar, we could thereby identify groups of expansin genes that are orthologous between the two species – i.e. that they are all descendants of a single gene in the last ancestor of monocots and eudicots. By linking gene duplications to polyploidy events, we were also able to resolve phylogenetic uncertainties and detect nodes where sequence-based phylogenetic trees were misleading (Sampedro et al. 2005).

Position-based phylogeny, based on the history of the segments where the genes appear, provides information not only on topologies but also on branch lengths. This information can increase the confidence on nodes that are poorly supported in traditional phylogenetic trees. Where it contradicts sequence-based phylogeny, however, one has to decide between the alternative topologies on a case-by-case basis. In some instances position-based phylogeny may be misleading, but it can also alert us to the presence of long-branch attraction or other phylogenetic artifacts (Sampedro et al. 2005). Further study of problematic nodes – for example, through increased taxon sampling – should eventually resolve the contradictory topologies.

The results of the integration of sequence-based and position-based phylogeny suggested that Arabidopsis and rice expansins belong to 17 distinct lineages or clades (12 EXPA, two EXPB, one EXLA and two EXLB) that were already differentiated in the last common ancestor of the two species (Table 1). Some of these monocot-eudicot

Table 1. Expansin clades[a] in angiosperms

Families	Monocot-eudicot clades	Genes			
		Ath	*Osa*	*Ptr*	Other species[b]
EXPA	I	1, 5, 10, 15	5, 21	3, 4, 6, 9	*Car*: 1, 3, 4 / *Cpl*: 1 / *Csa*: 3, 7 / *Les*: 10 *Rac*: 4 *Fan*: 3, 4, 5 / *Ghi*: 3, 6 / *Les*: 5 / *Mdo*: 3, 4, 5 / *Nta*: 4 / *Pco*: 4, 5 / *Rac*: 1 / *Rpa*: 2, 3, 5, 13, 14, 15, 18 / *Sas*: 2 / *Tve*: 1, 3 / *Zel*: 3 / *Zma*: 2
	II	14	11	10, 11	
	III	2, 8	2, 4, 6	2, 12, 13, 14	*Car*: 2 / *Cpl*: 2, 3 / *Csa*: 1, 8 / *Fan*: 2, 7 / *Fpa*: 2 / *Ghi*: 1, 2 / *Les*: 2 / *Mac*: 1 / *Mdo*: 1, 2 / *Mja*: 1, 4, 3, 6, 7 / *Nta*: 5 / *Pco*: 1, 2, 3 / *Phy*: 3 / *Rac*: 2 / *Rpa*: 1, 4, 7, 8, 9, 16 / *Sni*: 1, 4 / *Zel*: 2 / *Zma*: 1, 5
	IV	3, 4, 6, 9, 16	7	1, 5, 7, 15, 16	*Ghi*: 4 / *Csa*: 4, 5, 6, 9 / *Fan*: 1, 6 / *Gma*: 2 / *Les*: 1, 4, 6, 9, 18 / *Mal*: 1 / *Mdo*: 6 / *Mja*: 2, 5 / *Nta*: 6 / *Pco*: 6, 7 / *Phy*: 1, 2 / *Psa*: 1/*Rac*: 3 / *Rpa*: 6, 10, 11, 12, 17, 19 / *Sas*: 1/*Sni*: 2 / *Spy*: 3 / *Tve*: 2 / *Zel*: 1 / *Zma*: 4
	V	11	1, 12, 13, 14, 15, 18, 19, 20, 22, 23, 24, 25, 27, 28, 29, 31	17, 18	*Csa*: 2 / *Fpa*: 1, 4, 5 / *Gma*: 1 / *Les*: 8 / *Nta*: 1, 2, 3 / *Sni*: 3 / *Spy*: 1 / *Zma*: 3
	VI	17	3, 8, 9, 33	19	*Fpa*: 3 / *Sas*: 3 / *Spy*: 4
	VII	12		20	*Ghi*: 5
	VIII	20	16	21, 22	
	IX	13	10	8, 23	*Les*: 3
	X	7, 18	17, 30	24	
	XI		26, 32	25, 26	
	XII	19, 21, 22, 23, 24, 25, 26		27	*Les*: 7
	?				*Spy*: 2
EXPB	I	2, 4, 5, 6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18	1	*Fpa*: 1, 2, 3 / *Gma*: 1 (*cim1*) / *Nta*: 1 (*PPAL*) / *Zma*: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
	II	1, 3	16, 17	2, 3	
EXLA	I	1, 2, 3	1, 2, 3, 4	1, 2	*Mja*: 1
EXLB	I	1		1, 2	
	II		1	3, 4	

[a]Monocot-eudicot clades (independent lineages existing in the last ancestor of monocots and eudicots) are used to classify angiosperm expansins. The assignment of Arabidopsis (*Ath*) and rice (*Osa*) expansin genes to the different monocot-eudicot clades is based on Sampedro et al. (2005). For *Populus trichocarpa* (*Ptr*) expansins see Fig. 1 and text. The classification of previously described expansins from other species is based on the results of neighbor-joining phylogenetic trees made with protein sequences of each species and a selection of Arabidopsis, poplar and rice expansins representing all monocot-eudicot clades. In the case of *LeEXPA10* and *RaEXPA4*, trees cannot resolve if they belong to clade EXPA-II or EXPA-I. *SpEXPA2* appears to belong to a lineage not present in the known angiosperm genomes

[b]Species and references: *Car*: *Cicer arietinum* (Sanchez et al. 2004), *Cpl*: *Craterostigma plantagineum* (Jones and McQueen-Mason 2004), *Csa*: *Cucumis sativus* (Shcherban et al. 1995; Link et al. 2001), *Fan*: *Fragaria ananassa* (Civello et al. 1999; Harrison et al. 2001; Rose et al. 1997), *Fpa*: *Festuca pratensis* (Reidy et al. 2001), *Ghi*: *Gossypium hirsutum* (Harmer et al. 2002), *Gma*: *Glycine max* (Crowell 1994; Lee et al. 2003), *Les*: *Lycopersicon esculentum* (Rose et al. 1997; Reinhardt et al. 1998; Brummell et al. 1999; Catala et al. 2000; Chen and Bradford 2000; Vogler et al. 2003), *Mac*: *Musa acuminata* (Trivedi and Nath 2004), *Mal*: *Melilotus alba* (Giordano and Hirsch 2004), *Mdo*: *Malus* × *domestica* (Wakasa et al. 2003), *Mja*: *Mirabilis jalapa* (Gookin et al. 2003), *Nta*: *Nicotiana tabacum* (Link and Cosgrove 1998; Pezzotti et al. 2002), *Pco*: *Pyrus communis* (Hiwasa et al. 2003), *Phy*: *Petunia* × *hybrida*(Zenoni et al. 2004), *Psa*: *Pisum sativum* (Michael 1996), *Rac*: *Rumex acetosa* (Vriezen et al. 2000), *Rpa*: *Rumex palustris* (Vriezen et al. 2000; Colmer et al. 2004), *Sas*: *Striga asiatica* (O'Malley and Lynn 2000), *Sni*: *Sambucus nigra* (Belfield et al. 2005), *Spy*: *Sagittaria pygmaea* (Ookawara et al. 2005), *Tve*: *Triphysaria versicolor* (Wrobel and Yoder 2001), *Zel*: *Zinnia elegans* (Im et al. 2000), *Zma*: *Zea mays* (Wu et al. 2001; Li et al. 2003a)

orthologous groups are not recognized in phylogenetic trees as monophyletic groups, and others are incorrectly resolved (Sampedro et al. 2005). Two of them appear to have been unilaterally lost in the Arabidopsis lineage, and three are not found in the rice genome (Table 1).

We proposed, in our previous work, that the independent lineages existing in the last ancestor of monocots and eudicots can also be used as the basis of a subdivision of the four expansin families. For this purpose we used roman numerals to denote the 17 orthologous expansin clades (Table 1). In our definition, each clade would include all of the genes that descend from a single gene present in the last common ancestor of Arabidopsis and rice and would thus constitute a monocot-eudicot orthologous group of genes. In principle, it might be preferable to base a classification system like this on a more significant divergence, like the basal node of all extant angiosperms or of all extant vascular plants. On the other hand, the fully assembled genomes of Arabidopsis and rice, through microsynteny analysis, offer us a unique chance to solve phylogenetic uncertainties and

identify independent lineages with a much greater degree of confidence. This subdivision is also directly applicable to at least 90% of all angiosperms species and, if necessary, could be modified in the future. A division of the EXPA family into four groups had already been proposed when only a few sequences were known (Link and Cosgrove 1998). This scheme is compatible with ours, since their groups A, B, C and D correspond to clades IV, III, I, and V, respectively.

Embedding the history of the expansin superfamily into the histories of the Arabidopsis and rice genomes not only helps to clarify the phylogeny of this gene family but also provides a broader perspective of expansin evolution. Polyploidy events affect the whole genome, but they are followed by massive gene loss. Following event α, the most recent polyploidy detected in Arabidopsis, one of the copies was lost in 85% of the duplicated gene pairs (Blanc and Wolfe 2004). In the case of the Arabidopsis expansin superfamily we have estimated that there were at least 67 gene duplications since divergence from the rice lineage, 56 of which were due to events α–γ. At the same time a minimum of 48 gene deaths occurred, most of those did not involve large deletions since we still can recognize the genomic segments where the missing expansin should have been (Sampedro et al. 2005). We could say that the Arabidopsis expansin superfamily, as we see it today, is the result of the repeated pruning of redundant genes.

After a gene duplication, the long-term survival of both copies is only expected when they evolve differentiated functions. This can happen if one of the copies acquires a novel function or, more frequently, when the function of the ancestral gene is partitioned between both copies (Prince and Pickett 2002). Some expansin genes, such as *AtEXPA10*, have complex patterns of expression that include separate expression domains in several organs (Cho and Cosgrove 2000). Other expansins, such as *AtEXPA7* and *AtEXPA18*, appear to be specific to particular cell types (Cho and Cosgrove 2002). The monocot-eudicot clades where Arabidopsis and rice still maintain one-to-one orthology, despite having gone through at least four duplications, could correspond to genes with simple expression patterns of this second type, where a specialized gene function does not leave much room for functional divergence. Similarly, the clades where most growth has occurred might represent generalized expansins that can function, as seems to be the case for *AtEXPA10*, in many different cell-wall modifying processes. In the case of rice, two of the 17 ancestral clades have grown to such a degree that they now include 58% of all the expansin genes in this species (Sampedro et al. 2005). It is possible that this asymmetrical growth is related to the unique changes in the composition of the grass cell wall (Carpita 1996).

Some of the conclusions from the integrated analysis of Arabidopsis and rice expansins, in particular with respect to the number of EXPA monocot-eudicot clades, were offered in our previous study as tentative working hypotheses. The availability of a fairly complete sequence of the genome of *Populus trichocarpa* represents an opportunity to test some of these ideas.

Expansin evolution and the poplar genome

The preliminary draft (version 1.0) of the *P. trichocarpa* genome has been made available by the DOE Joint Genome Institute and the Poplar Genome Consortium (http://www.genome.jgi-psf.org/Poptr1/Poptr1.home.html). The genome of *P. trichocarpa* contains approximately 480 Mbp and has been sequenced to 7.5× depth. Most of the sequence has been assigned to one of the 19 chromosomes of *P. trichocarpa*, but many small fragments still remain unmapped.

The genus *Populus* has become a model system for tree molecular biology, and the availability of the entire genome sequence of one of its species will reinforce this position (Brunner et al. 2004). As a resource for comparative genomics, *P. trichocarpa* represents a second eudicot species that is fairly divergent from Arabidopsis, although both belong to the rosid subclass. It is not surprising that a small-scale study has found a considerable degree of collinearity between fragments of the Arabidopsis and poplar genomes (Stirling et al. 2003).

Among the 58,036 predicted gene models in the current poplar genome release, we found 36 that correspond to members of the expansin superfamily. In one case (*PtEXPA24*) the gene model is missing part of an exon because of what appears to be a 1-bp deletion in the coding region. If the sequence is correct, it is very unlikely that this would be a functional gene. To investigate this point in further detail, we examined the relevant raw sequences deposited in the trace archive at NCBI. One of the sequences that cover this region, TI619020066, shows a very wide peak for the cytosine at position 724, and another sequence, TI620069350, actually shows a clear double peak (positions 300–301), which is indicative of a sequence error. The addition of an extra cytosine in this position allows the gene to be translated into a full-length expansin protein. We used this edited sequence for our phylogenetic analysis. All the protein sequences and their genomic locations are provided in the electronic supplementary material (ESM; S1). A previous study identified and named nine expansin genes from hybrid aspen (*P. tremula* × *P. tremuloides* Michx) represented in expressed sequence tag (EST) libraries (Gray-Mitsumune et al. 2004). We identified the closest genes in the genomic sequence of *P. trichocarpa* and gave them identical numbers. The remainder of the genes were numbered consecutively following the order of the monocot-eudicot clades to which they were assigned. Incomplete gene fragments were ignored.

Expansin clades in poplar

The *P. trichocarpa* genome contains at least 27 EXPA, three EXPB, two EXLA and four EXLB genes (Table 1). Figure 1 shows phylogenetic trees of all the poplar expansins and selected sequences from Arabidopsis and rice that represent all of the monocot-eudicot clades present in both species. EXPA sequences were analyzed separately due to the presence of family-specific insertions and

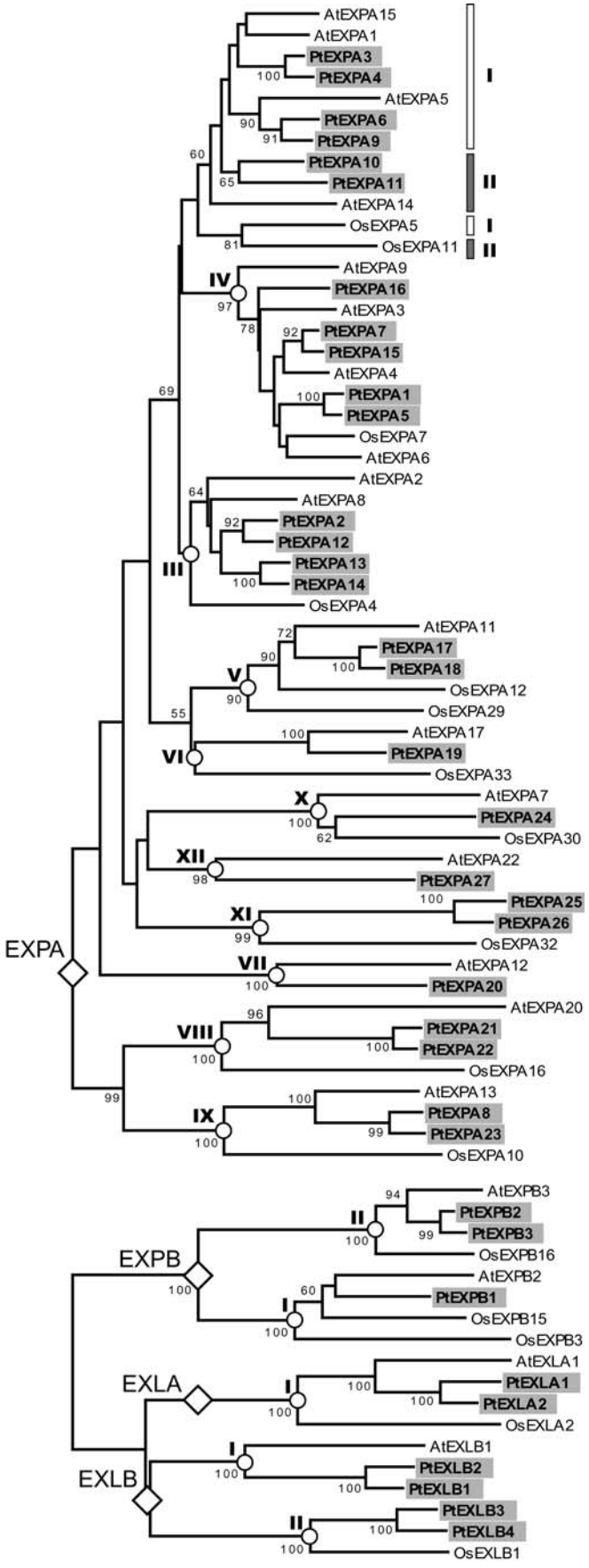

Fig. 1. Phylogenetic trees of *Populus trichocarpa* expansins and selected rice and Arabidopsis sequences. Protein alignments were made with CLUSTALW and a neighbor-joining tree was obtained with MEGA 3.1 (Kumar et al. 2004). Poisson correction with the complete deletion of gaps was used to calculate protein distances. Bootstrap values are based on 500 iterations. Bootstrap values below 55 are not shown. The four families are labeled at their respective roots (*white diamonds*). The roots of the different expansin monocot-eudicot clades (independent lineages in the last common ancestor of monocots and eudicots) are indicated by *white circles* and labeled as in Sampedro et al. (2005). In the case of clades EXPA-I and EXPA-II, which are not correctly resolved by this tree, clade assignments are based on microsynteny and indicated as alternating *white and grey bars* to the *right*

deletions. In most cases the assignment of the poplar expansins to the previously defined monocot-eudicot clades was straightforward. In the case of *PtEXPA10* and *PtEXPA11*, the phylogenetic tree neither supports nor clearly rejects the possibility that they belong to clade EXPA-II, which is represented in Arabidopsis by *AtEXPA14*. We decided to use position-based phylogeny to resolve this question.

Figure 2 shows the results of a small-scale microsynteny analysis. We can see that many of the genes surrounding *PtEXPA10* and *PtEXPA11* have their best Arabidopsis homologs in the segment that contains *AtEXPA14* or in other segments that are duplicates of this one. The four Arabidopsis segments had been identified as the products of the two most recent Arabidopsis polyploidy events (α and β) by whole genome studies (Simillion et al. 2002; Paterson et al. 2004). This result suggests that, even though *AtEXPA14* is not the Arabidopsis homolog with the highest sequence similarity to the poplar genes (the closest homologs are in clade EXPA-I), it is very likely that the three genes form an orthologous group and thus belong to clade EXPA-II. This possibility is further supported by the fact that some of the genes conserved between Arabidopsis and poplar in this region also appear in the same order and orientation in a putatively orthologous segment of the rice genome (which includes *OsEXPA11*), while the rice and Arabidopsis members of clade EXPA-I appear in segments that share a different set of genes (Sampedro et al. 2005). In our phylogenetic tree the Arabidopsis and poplar members of clade EXPA-II are not resolved as a monophyletic group, but this could be due to accelerated evolution on the part of *AtEXPA14*. As for the rice member of this clade, *OsEXPA11*, here it branches with a rice EXPA-I gene. As we explained in our previous work, this is likely due to a bias in amino acid composition between rice and Arabidopsis expansins (Sampedro et al. 2005).

Among the most noteworthy results from our analysis of expansins in the *P. trichocarpa* genome is the presence of members of the two monocot-eudicot clades that were lost in the Arabidopsis lineage. In the case of clade EXLB-II, we had already found members of this clade in other eudicots (Sampedro et al. 2005). Clade EXPA-XI, on the other hand, was only known from monocot species. In our previous work we had contemplated the possibility that this clade could actually be the monocot branch of clade EXPA-XII, which is only known from eudicots. Neighbor-joining and

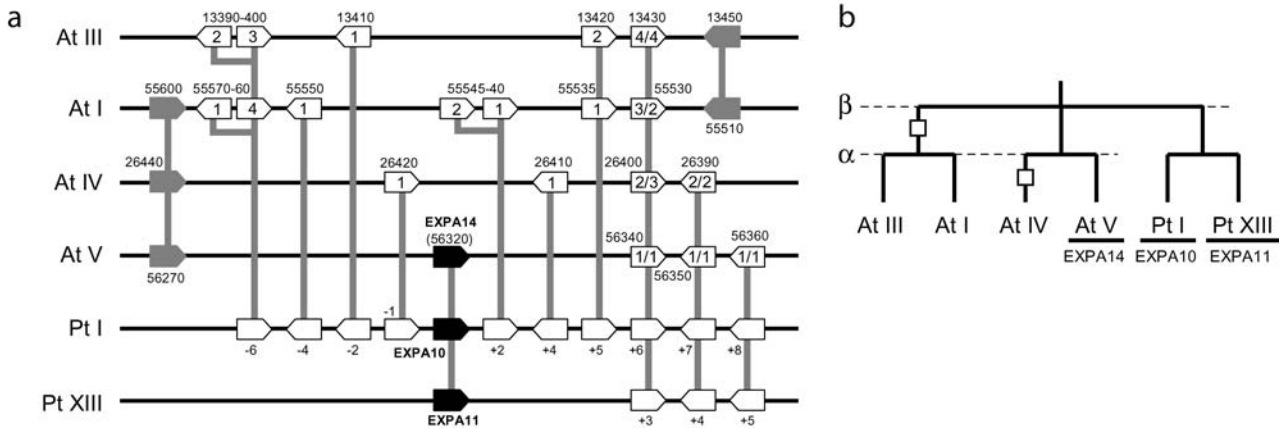

Fig. 2. Microsynteny analysis of clade EXPA-II in *Populus trichocarpa* and Arabidopsis. **a** Four Arabidopsis genomic segments and two *P. trichocarpa* segments that likely descended from a single segment in the last ancestor of these two species have been aligned to show conservation of gene order and orientation. Segments are identified to the *left* by the species abbreviation and chromosome number. Genes are indicated by *pentagons* oriented in the direction of transcription. *Vertical grey lines* connect homologous genes. The translated products of the poplar genes in these segments were used in BLASTP searches against the Arabidopsis proteome (http://www.ncbi.nlm.nih.gov/BLAST/Genome/PlantBlast.shtml?10). *Numbers inside* Arabidopsis genes indicate the order in which the product of that gene appeared in the list of BLASTP results; for example, number *1* indicates the first hit in the entire Arabidopsis proteome for the connected poplar gene. The *second number*, if present, indicates results obtained with proteins from segment PtXIII. Expansins are shown in *black* and *labeled in bold*. Arabidopsis genes are identified by their *annotation number*. *P. trichocarpa* genes are identified by their position relative to the expansin gene. *Grey* genes are previously described groups of duplicated Arabidopsis genes. **b** Cladogram indicating the relationship among the segments shown on the *left*. *Dashed lines* indicate polyploidy events in the Arabidopsis lineage. The attribution of Arabidopsis segmental duplications to these events is based on previously published whole genome studies (Simillion et al. 2002; Bowers et al. 2003). On the basis of expansin phylogenetic trees, the two poplar segments are shown as the product of a poplar-specific segmental duplication. It is unclear if the divergence between Arabidopsis and *P. trichocarpa* happened before or after event β. Segments with expansins are *underlined*, and the name of the expansin gene is shown *below* the underlining. *White boxes* indicate the loss of an expansin gene

parsimony trees suggested this possibility with weak support, while a Bayesian tree rejected it (Sampedro et al. 2005). Since no synteny was detected and protein distances between EXPA-XI and EXPA-XII expansins were very large, we opted to keep them as separate clades. The presence of clade EXPA-XI in poplar confirms this hypothesis. Furthermore, when we added the *P. trichocarpa* EXPA-XI and EXPA-XII expansins to the trees that had suggested a relationship, these two clades no longer appear as a monophyletic group (data not shown). Thus, it seems that the grouping of these two clades in our previous work was an artifact of long-branch attraction (Felsenstein 1978).

A similar question about another two clades, EXPA-VI (exclusive to Arabidopsis) and EXPA-VII, was also left open in our previous work. In this case the Bayesian and parsimony trees strongly supported the possibility that *AtEXPA12* from clade EXPA-VII and *AtEXPA17* from clade EXPA-VI were a monophyletic group, a possibility that was rejected, equally strongly, by the neighbor-joining tree (Sampedro et al. 2005). When the poplar sequences from these two clades are added, the eudicot members of both clades no longer appear as a monophyletic group in either Bayesian or parsimony trees (data not shown). These results increase our confidence in the independence of clades EXPA-VI and EXPA-VII.

Other questions are still left open. The branches represented in Fig. 1 by *OsEXPA29* and *OsEXPB3* were interpreted in our previous work as being highly divergent branches in clades EXPA-V and EXPB-I, on the basis of synteny evidence, even though phylogenetic trees suggested that they could represent independent monocot-eudicot clades lost in Arabidopsis (Sampedro et al. 2005). None of the poplar expansins fall into these branches and, consequently, we have no reason to change our previous interpretation. On the basis of position-based phylogeny, we also proposed in our previous study that clade EXPA-IV has a similar rooting problem in sequence-based trees due to divergent branches in the Arabidopsis lineage (Sampedro et al. 2005). When poplar expansins are added to the neighbor-joining tree, the topology of this clade still shows the same pattern, *OsEXPA7* appears nested deep inside the clade, although support for many branches is now very low (Fig. 1).

We performed a small-scale synteny analysis for clade EXPA-IV and found that the *P. trichocarpa* genomic segments that include expansins from this clade can be divided into three groups, each of them showing clear synteny with a different group of Arabidopsis segments (Fig. 3a). Synteny diagrams are provided in the ESM (S2). Combining these results with our previous analysis of synteny between Arabidopsis and rice, we obtained a position-based tree for the expansins in clade EXPA-IV present in the three sequenced genomes (Fig. 3b). The topology of the position-based tree is compatible with that of a sequence-based tree, if low support nodes are ignored, where the root has been manually placed between monocot and eudicot expansins (Fig. 3c). We still believe that the more parsimonious hypothesis is to assume that the root of the clade in family-wide trees is misplaced on account of accelerated rates of evolution in several branches, particularly *AtEXPA9*. This

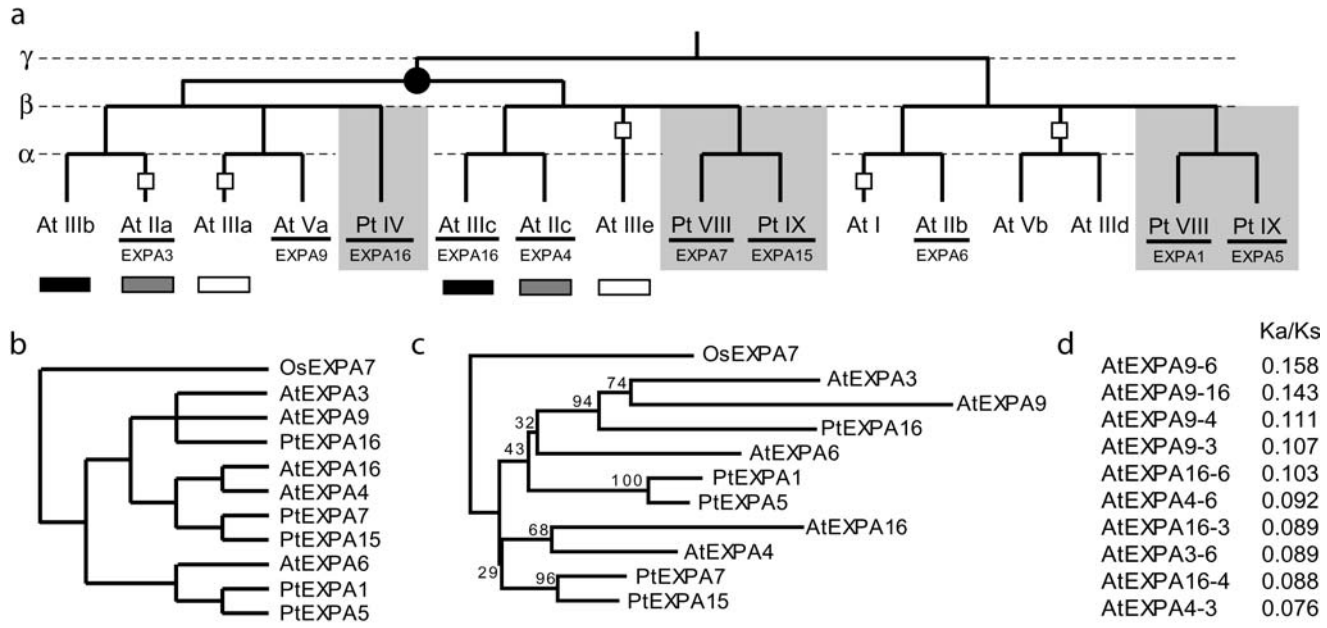

Fig. 3. Microsynteny analysis of clade EXPA-IV. **a** Cladogram indicating the relationship among genomic segments from Arabidopsis and *Populus trichocarpa* (*grey boxes*) that contain clade EXPA-IV expansins or contained them in the past. Segments are identified by the species *abbreviation*, chromosome *number* and a *letter* when necessary. *Dashed lines* indicate polyploidy events detected in the Arabidopsis genome. Event α is specific to the Arabidopsis lineage, event γ is common to both lineages and the dating of event β is uncertain. Segments with expansins are *underlined* and the name of the expansin gene shown *below* the underlining. A *black circle* indicates a segmental duplication in tandem. *Boxes of the same color below* pairs of genomic segments indicate that they are in close physical proximity. *White boxes* indicate the loss of an expansin gene. Synteny diagrams and details on the construction of this tree are provided as ESM S2. **b** Cladogram of clade EXPA-IV expansins based on the segmental cladogram. **c** Neighbor-joining phylogenetic tree of clade EXPA-IV expansins, manually rooted between monocot and eudicot species. Protein alignments were made with CLUSTALW, and a tree was obtained with MEGA 3.1 (Kumar et al. 2004). Poisson correction with complete deletion of gaps was used to calculate protein distances. Bootstrap values are based on 500 iterations. **d** Pairwise *Ka*/*Ks* ratios for Arabidopsis genes in clade EXPA-IV in order of decreasing magnitude. Rates of nonsynonymous (*Ka*) and synonymous (*Ks*) substitutions were calculated, according to the Pamilo-Bianchi-Li method as implemented by MEGA 3.1

idea is supported by the high ratio of nonsynonymous (*Ka*) to synonymous (*Ks*) substitution rates observed in pairwise comparisons that include *AtEXPA9* (Fig. 3d). The synteny between the segments that contain *PtEXPA16* and *AtEXPA9* is also particularly good (see ESM S2), and this would be difficult to explain if the family-wide trees were correct, since they suggest that the two genes had diverged long before the separation of the rice and Arabidopsis lineages (Fig. 1). Furthermore, the family-wide tree requires the branch represented by *AtEXPA9* to be lost independently in both rice and poplar.

The only remaining inconsistency is the fact that *PtEXPA24* branches with the rice member of clade EXPA-X instead of branching with the Arabidopsis member, as is seen in all the other monocot-eudicot clades represented in all three species (Fig. 1). The bootstrap for this node is low (62) and clade EXPA-X appears at the end of a very long branch that might cause rooting problems. For these reasons we prefer to keep the more parsimonious hypothesis of a single monocot-eudicot clade, as suggested by synteny (Sampedro et al. 2005).

Finally, we found no evidence in *P. trichocarpa* for the existence of previously unknown monocot-eudicot clades that had been lost both in Arabidopsis and rice (Fig. 1). This makes the discovery of many new monocot-eudicot clades very unlikely.

In conclusion, the analysis of the poplar expansin superfamily reinforces our confidence in the existence of 17 expansin clades in angiosperms that have been independent since the last ancestor of eudicots and monocots. All of these clades are represented in the genome of *P. trichocarpa* (Table 1).

Genome duplications

One of the remarkable features of the trees shown in Fig. 1 is the presence of 14 pairs of poplar expansins that appear to be the product of gene duplications specific to this lineage. Based on their short branch length, many of the duplications appear to be recent. Furthermore, in only one of these cases do the duplicated genes appear in close physical proximity (*AtEXPA25* and *AtEXPA26* are around 221 kbp apart); all the other gene pairs that have been mapped (nine of them) are located on different chromosomes (see ESM S1 for gene locations). Furthermore, the three poplar-specific expansin duplications in clades EXPA-II and EXPA-IV, the only ones in which we analyzed microsynteny, were clearly the product of segmental duplications (Fig. 2; ESM S2). On the basis of an analysis of EST collections Sterck et al. (2005) proposed that the *Populus* lineage went through a polyploidy event between 8 and 13 million

years ago. It is likely that many of the duplications of expansin genes in the *P. trichocarpa* genome are the result of this event.

Another interesting observation about the genomic locations of expansin genes in poplar is the fact that the two genes in clade EXLB-I appear to be approximately 500 kbp apart from genes in clade EXLB-II (*PtEXLB1* is close to *PtEXLB4*; *PtEXLB2* is close to *PtEXLB3*). This suggests that the separation between the two monocot-eudicot clades might have been the product of an ancient tandem duplication or a small-scale segmental duplication.

An important issue is the dating of the divergence of the Arabidopsis and poplar lineages with respect to the polyploidy events previously detected in the Arabidopsis genome. The EXPA phylogenetic tree shown in Fig. 1 strongly supports the possibility that event γ happened before the divergence of the two lineages, since poplar expansins are found on both sides of the node separating *AtEXPA15* and *AtEXPA5*, which has previously been attributed to this polyploidy event (Sampedro et al. 2005). The other node attributed to this event separates *AtEXPA6* from the remaining Arabidopsis members of clade EXPA-IV. Sequence-based phylogenetic trees for this clade are unreliable, but synteny analysis strongly supports the hypothesis that event γ happened in the common lineage giving rise to *P. trichocarpa* and Arabidopsis (Fig. 3).

In the case of genome duplication β, the expansin phylogenetic trees are not very informative. There are no cases of two branches in poplar whose divergence can be clearly attributed to this event. In contrast, three Arabidopsis expansin duplications have been attributed to this event: *AtEXPA15-AtEXPA1*, *AtEXPA2-AtEXPA8* and *AtEXPA3-AtEXPA9* (Sampedro et al. 2005). The position of *P. trichocarpa* expansins with respect to these nodes is resolved in contradictory ways, and none with very high support (Figs. 1, 3). Previous datings of the Arabidopsis genome duplications in relation with the divergence of different lineages did not include *P. trichocarpa* among the taxons analyzed, but the results would imply that event β predates the divergence of *P. trichocarpa* and Arabidopsis (Bowers et al. 2003; Chapman et al. 2004). We have discussed the potential pitfalls of this dating method in our previous study (Sampedro et al. 2005).

Our limited synteny analyses could be interpreted as suggesting that *P. trichocarpa* diverged from the Arabidopsis lineage before event β. In the segments shown in Fig. 2, for example, the putative Arabidopsis orthologs of the poplar genes in segment PtI are found in five cases exclusively on one side of the β node (segments AtI and AtIII) and in another five cases on the other side (segments AtIV and AtV). Only in one case do the putative orthologs appear on both sides. If event β occurred in a common ancestor of Arabidopsis and *P. trichocarpa*, PtI would be orthologous only to the Arabidopsis segments on one side of the node. Whichever side we choose for PtI, this would require five of the ancestral genes in this region to survive as duplicates between event β and the divergence of the two species, only to be lost later on the Arabidopsis segments orthologous to PtI, while surviving both in poplar and on the Arabidopsis segments paralogous to PtI. A similar case can be made for the segments analyzed on ESM S2. A genome-wide analysis of synteny between *P. trichocarpa* and Arabidopsis would probably clarify this important issue.

The dating of polyploidy event β determines the estimate of the minimal number of expansin genes present in the last ancestor of Arabidopsis and *P. trichocarpa*. If this genome duplication is specific to the Arabidopsis lineage, the number would be 19 (14 EXPA, two EXPB, one EXLA and two EXLB). Three more EXPA genes would be required in this ancestor if event β occurred in the common lineage of the two species.

Introns

The intron pattern of all the poplar expansins is provided in the ESM (S1). In our previous analysis of the Arabidopsis and rice expansins we established the minimal ancestral intron patterns for each of the monocot-eudicot clades (Sampedro et al. 2005). The analysis of the *P. trichocarpa* genome requires a change in the ancestral pattern of clade EXPA-XI, which should be [AB] instead of [B]. As for poplar-specific intron losses, the two members of clade EXPA-VIII have lost intron A; all the other genes have maintained the ancestral intron patterns of their clades. In the case of intron G, which is usually located in the 5′-untranslated region, its presence could be confirmed only in *PtEXPA3* (based on cDNA sequences) and in *PtEXPA4*, where it is located in the start codon. For the other genes where this intron might be expected (clades EXPA-I and EXPA-II), no cDNA sequences were found, and so we could not assess the presence of intron G in these genes.

Expansin clades in other angiosperm species

It would be useful, in our opinion, to extend the subdivision of the expansin families into monocot-eudicot clades to other species as well. We offer here a first attempt, the results of which can be seen in Table 1. We searched the literature for previously named expansin genes from angiosperm species and assigned them to one of the monocot-eudicot clades on the basis of phylogenetic trees similar to those in Fig. 1. It appears that most of the expansin genes that have been cloned to date belong to just a few of these clades (EXPA-I, EXPA-III, EXPA-IV, EXPA-V and EXPB-I). In some cases, particularly with gene fragments or species that branch close to or before the divergence of the eudicot and monocot lineages, the assignment is based on nodes with low support. It is particularly difficult to discriminate between clades EXPA-I and EXPA-II on the basis of sequence-based trees. On the other hand, only one expansin gene – *SpEXPA2* (*Spy*: 2 in Table 1) from the monocot *Sagittaria pygmaea* – appears to belong to a lineage distinct from those represented in the three known

angiosperm genomes (data not shown). We could not find any other sequences from this lineage in GenBank (including EST collections), so we can not exclude the possibility that it is a highly divergent branch of a known monocot-eudicot clade (EXPA-I, EXPA-II and EXPA-III are the best candidates).

Expansins in gymnosperms

To determine how far back the differentiation of the expansin monocot-eudicot clades took place, we also analyzed the expansin sequences available in the TIGR *Pinus* gene index, release 6.0 (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=pine), which includes full-length sequences or fragments of more than 45,000 genes (Quackenbush et al. 2001). We were able to confirm the presence, in the *Pinus taeda* L. transcriptome, of lineages that are sister to eight of the expansin monocot-eudicot clades defined in angiosperms (Fig. 4). Furthermore, a number of pine EXPA sequences appear to belong to a branch not present in the genomes of rice, Arabidopsis or poplar; this branch is not related to that represented by *SpEXPA2*. These results suggest that the last ancestor of angiosperms and gymnosperms had at least six EXPA genes and one EXPB, one EXLA and one EXLB gene. However, this is likely to be an underestimation.

Functional implications

The survival of most of the expansin monocot-eudicot clades in Arabidopsis, rice and poplar, together with the fact that many of them appear to have diverged before the separation of the angiosperm and gymnosperm lineages, could be interpreted as indicating an important degree of ancestral functional specialization. By comparing expression patterns in different species, it has recently been proposed that expansins belonging to group A, which corresponds to clade EXPA-IV in our classification, may play an important role in xylem development (Gray-Mitsumune et al. 2004). This idea is also supported by an earlier study (Im et al. 2000) on *Zinnia elegans* expansins. Similarly, both Arabidopsis members of clade EXPA-X have been shown to be specifically expressed in trichoblasts, while one of their rice orthologs (*OsEXPA17*) appears to be root-specific (Cho and Cosgrove 2002; Lee and Kende 2002). The clarification of expansin phylogeny is a necessary first step in order to extend this comparative approach and understand the evolution of gene function in the different lineages.

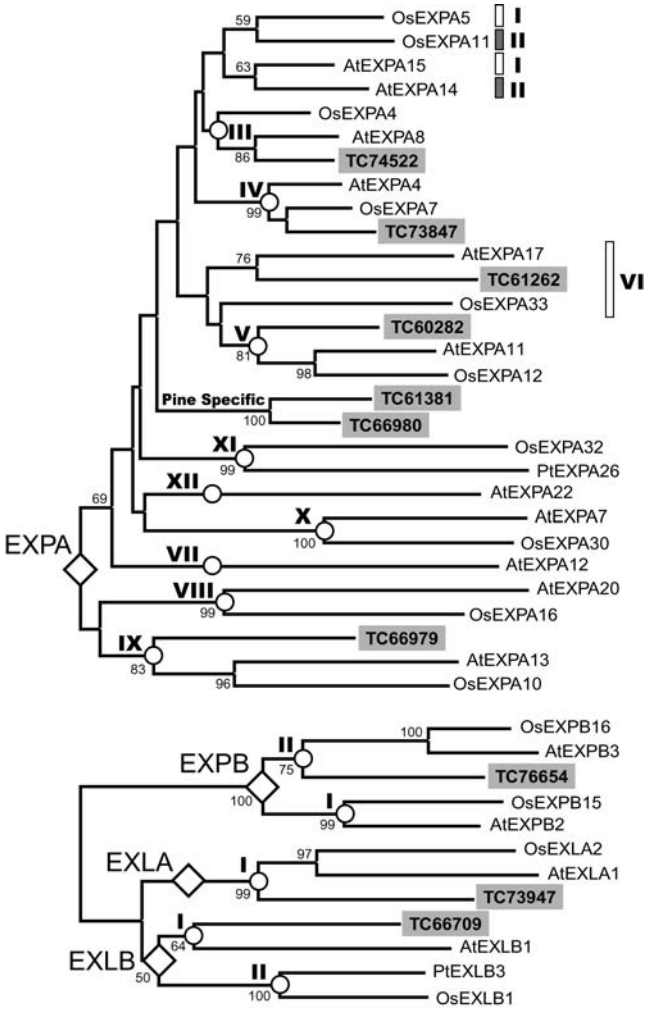

Fig. 4. Phylogenetic trees of *Pinus taeda* L. expansins and angiosperm sequences. Protein alignments were made with CLUSTALW and a neighbor-joining tree was obtained with MEGA 3.1 (Kumar et al. 2004). Poisson correction with complete deletion of gaps was used to calculate protein distances. Bootstrap values below 55 are not shown. Pine sequences are shown in *grey boxes* and identified by their TIGR Pine Index accession *numbers*. The four families are labeled at their respective roots (*white diamonds*). The roots of the different expansin monocot-eudicot clades (independent lineages in the last common ancestor of monocots and eudicots) are indicated by *white circles* and labeled as in Sampedro et al., (2005). In the case of clades EXPA-I, EXPA-II and EXPA-VI, which are not resolved by this tree, the clade identification is indicated as alternating *white and grey bars* to the *right*

References

Belfield EJ, Ruperti B, Roberts JA, Queen-Mason S (2005) Changes in expansin activity and gene expression during ethylene-promoted leaflet abscission in *Sambucus nigra*. J Exp Bot 56:817–823

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16:1679–1691

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Brummell DA, Harpster MH, Dunsmuir P (1999) Differential expression of expansin gene family members during growth and ripening of tomato fruit. Plant Mol Biol 39:161–169

Brunner AM, Busov VB, Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends Plant Sci 9:49–56

Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol 47:445–476

Catala C, Rose JK, Bennett AB (2000) Auxin-regulated genes encoding cell wall-modifying proteins are expressed during early tomato fruit growth. Plant Physiol 122:527–534

Chapman BA, Bowers JE, Schulze SR, Paterson AH (2004) A comparative phylogenetic approach for dating whole genome duplication events. Bioinformatics 20:180–185

Chen F, Bradford KJ (2000) Expression of an expansin Is associated with endosperm weakening during tomato seed germination. Plant Physiol 124:1265–1274

Cho HT, Cosgrove DJ (2002) Regulation of root hair initiation and expansin gene expression in Arabidopsis. Plant Cell 14:3237–3253

Cho HT, Cosgrove DJ (2000) Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 97:9783–9788

Civello PM, Powell ALT, Sabehat A, Bennett AB (1999) An expansin gene expressed in ripening strawberry fruit. Plant Physiol 121:1273–1279

Colmer TD, Peeters AJ, Wagemaker CA, Vriezen WH, Ammerlaan A, Voesenek LA (2004) Expression of alpha-expansin genes during root acclimations to O_2 deficiency in *Rumex palustris*. Plant Mol Biol 56:423–437

Cosgrove DJ (2000) Loosening of plant cell walls by expansins. Nature 407:321–326

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

Cosgrove DJ, Li LC, Cho HT, Hoffmann-Benning S, Moore RC, Blecker D (2002) The growing world of expansins. Plant Cell Physiol 43:1436–1444

Crowell DN (1994) Cytokinin regulation of a soybean pollen allergen gene. Plant Mol Biol 25:829–835

Darley CP, Li Y, Schaap P, McQueen-Mason SJ (2003) Expression of a family of expansin-like proteins during the development of *Dictyostelium discoideum*. FEBS Lett 546:416–418

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

Fry SC (1995) Polysaccharide-modifying enzymes in the plant cell wall. Annu Rev Plant Physiol Plant Mol Biol 46:497–520

Giordano W, Hirsch AM (2004) The expression of MaEXP1, a *Melilotus alba* expansin gene, is upregulated during the sweetclover-*Sinorhizobium meliloti* interaction. Mol Plant Microbe Interact 17:613–622

Gookin TE, Hunter DA, Reid MS (2003) Temporal analysis of alpha and beta-expansin expression during floral opening and senescence. Plant Sci 164:769–781

Gray-Mitsumune M, Mellerowicz EJ, Abe H, Schrader J, Winzell A, Sterky F, Blomqvist K, Queen-Mason S, Teeri TT, Sundberg B (2004) Expansins abundant in secondary xylem belong to subgroup A of the alpha-expansin gene family. Plant Physiol 135:1552–1564

Harmer SE, Orford SJ, Timmis JN (2002) Characterisation of six alpha-expansin genes in *Gossypium hirsutum* (upland cotton). Mol Genet Genomics 268:1–9

Harrison EP, McQueen-Mason SJ, Manning K (2001) Expression of six expansin genes in relation to extension activity in developing strawberry fruit. J Exp Bot 52:1437–1446

Hiwasa K, Rose JK, Nakano R, Inaba A, Kubo Y (2003) Differential expression of seven alpha-expansin genes during growth and ripening of pear fruit. Physiol Plant 117:564–572

Im KH, Cosgrove DJ, Jones AM (2000) Subcellular localization of expansin mRNA in xylem cells. Plant Physiol 123:463–470

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jones L, McQueen-Mason S (2004) A role for expansins in dehydration and rehydration of the resurrection plant *Craterostigma plantagineum*. FEBS Lett 559:61–65

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

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

Lee DK, Ahn JH, Song SK, Choi YD, Lee JS (2003) Expression of an expansin gene is correlated with root elongation in soybean. Plant Physiol 131:985–997

Lee Y, Kende H (2002) Expression of alpha-expansin and expansin-like genes in deepwater rice. Plant Physiol 130:1396–1405

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

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

Li Y, Darley CP, Ongaro V, Fleming A, Schipper O, Baldauf SL, McQueen-Mason SJ (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin. Plant Physiol 128:854–864

Li Y, Jones L, McQueen-Mason S (2003) Expansins and cell growth. Curr Opin Plant Biol 6:603–610

Link BM, Cosgrove DJ (1998) Acid-growth response and alpha-expansins in suspension cultures of bright yellow 2 tobacco. Plant Physiol 118:907–916

Link BM, Wagner ER, Cosgrove DJ (2001) The effect of a microgravity (space) environment on the expression of expansins from the peg and root tissues of *Cucumis sativus*. Physiol Plant 113:292–300

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

Michael AJ (1996) A cDNA from pea petals with sequence similarity to pollen allergen, cytokinin-induced and genetic tumour-specific genes: identification of a new family of related sequences. Plant Mol Biol 30:219–224

O'Malley RC, Lynn DG (2000) Expansin message regulation in parasitic angiosperms: marking time in development. Plant Cell 12:1455–1465

Ookawara R, Satoh S, Yoshioka T, Ishizawa K (2005) Expression of α-expansin and xyloglucan endotransglucosylase/hydrolase genes associated with shoot elongation enhanced by anoxia, ethylene and carbon dioxide in arrowhead (*Sagittaria pygmaea* Miq.) Tubers. Ann Bot 96:693–702

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Pezzotti M, Feron R, Mariani C (2002) Pollination modulates expression of the PPAL gene, a pistil-specific beta-expansin. Plant Mol Biol 49:187–197

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827–837

Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J (2001) The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res 29:159–164

Reidy B, McQueen-Mason S, Nosberger J, Fleming A (2001) Differential expression of alpha- and beta-expansin genes in the elongating leaf of *Festuca pratensis*. Plant Mol Biol 46:491–504

Reinhardt D, Wittwer F, Mandel T, Kuhlemeier C (1998) Localized upregulation of a new expansin gene predicts the site of leaf formation in the tomato meristem. Plant Cell 10:1427–1437

Rose JK, Bennett AB (1999) Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening. Trends Plant Sci 4:176–183

Rose JK, Lee HH, Bennett AB (1997) Expression of a divergent expansin gene is fruit-specific and ripening- regulated. Proc Natl Acad Sci USA 94:5955–5960

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

Sanchez MA, Mateos I, Labrador E, Dopico B (2004) Brassinolides and IAA induce the transcription of four alpha-expansin genes related to development in *Cicer arietinum*. Plant Physiol Biochem 42:709–716

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

Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99:13627–13632

Sterck L, Rombauts S, Jansson S, Sterky F, Rouze P, Van de Peer Y (2005) EST data suggest that poplar is an ancient polyploid. New Phytol 167:165–170

Stirling B, Yang ZK, Gunter LE, Tuskan GA, Bradshaw HD (2003) Comparative sequence analysis between orthologous regions of the *Arabidopsis* and *Populus* genomes reveals substantial synteny and microcollinearity. Can J For Res 33:2245–2251

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Trivedi PK, Nath P (2004) MaExp1, an ethylene-induced expansin from ripening banana fruit. Plant Sci 167:1351–1358

Vogler H, Caderas D, Mandel T, Kuhlemeier C (2003) Domains of expansin gene expression define growth regions in the shoot apex of tomato. Plant Mol Biol 53:267–272

Vriezen WH, De Graaf B, Mariani C, Voesenek LACJ (2000) Submergence induces expansin gene expression in flooding tolerant *Rumex palustris* and not in flooding intolerant *R. acetosa*. Planta 210:956–963

Wakasa Y, Hatsuyama Y, Takahashi A, Sato T, Niizeki M, Harada T (2003) Divergent expression of six expansin genes during apple fruit ontogeny. Eur J Hortic Sci 68:253–259

Wrobel RL, Yoder JI (2001) Differential RNA expression of alpha-expansin gene family members in the parasitic angiosperm *Triphysaria versicolor* (Scrophulariaceae). Gene 266:85–93

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

Zenoni S, Reale L, Tornielli GB, Lanfaloni L, Porceddu A, Ferrarini A, Moretti C, Zamboni A, Speghini A, Ferranti F, Pezzotti M (2004) Downregulation of the *Petunia hybrida* alpha-expansin gene PhEXP1 reduces the amount of crystalline cellulose in cell walls and leads to phenotypic changes in petal limbs. Plant Cell 16:295–308