

# Evolutionary divergence of $\beta$ -expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits

Javier Sampedro<sup>\*†</sup>, Mara Guttman<sup>‡</sup>, Lian-Chao Li<sup>§</sup> and Daniel J. Cosgrove<sup>\*</sup>

<sup>1</sup>Department of Biology, Pennsylvania State University, University Park, PA 16802 USA

Received 4 September 2014; revised 19 October 2014; accepted 21 October 2014; published online 29 October 2014.

<sup>\*</sup>For correspondence (e-mails javier.sampedro@psu.edu; dcosgrove@psu.edu).

<sup>†</sup>Present address: Departamento de Fisiología Vegetal, Universidad de Santiago, Santiago de Compostela 15782, Spain.

<sup>‡</sup>Present address: USDA-ARS 800 Buchanan St, Albany, CA 94710, USA.

<sup>§</sup>Present address: Pharmaceutical Development & Manufacturing Sciences Janssen R&D, LLC, Malvern, PA 19355, USA.

## SUMMARY

Expansins are wall-loosening proteins that promote the extension of primary cell walls without the hydrolysis of major structural components. Previously, proteins from the EXPA ( $\alpha$ -expansin) family were found to loosen eudicot cell walls but to be less effective on grass cell walls, whereas the reverse pattern was found for EXPB ( $\beta$ -expansin) proteins obtained from grass pollen. To understand the evolutionary and structural bases for the selectivity of EXPB action, we assessed the extension (creep) response of cell walls from diverse monocot families to EXPA and EXPB treatments. Cell walls from Cyperaceae and Juncaceae (families closely related to grasses) displayed a typical grass response (' $\beta$ -response'). Walls from more distant monocots, including some species that share with grasses high levels of arabinoxylan, responded preferentially to  $\alpha$ -expansins (' $\alpha$ -response'), behaving in this regard like eudicots. An expansin with selective activity for grass cell walls was detected in Cyperaceae pollen, coinciding with the expression of genes from the divergent EXPB-I branch that includes grass pollen  $\beta$ -expansins. The evolutionary origin of this branch was located within Poales on the basis of phylogenetic analyses and its association with the 'sigma' whole-genome duplication. Accelerated evolution in this branch has remodeled the protein surface in contact with the substrate, potentially for binding highly substituted arabinoxylan. We propose that the evolution of the divergent EXPB-I group made a fundamental change in the target and mechanism of wall loosening in the grass lineage possible, involving a new structural role for xylans and the expansins that target them.

**Keywords:**  $\beta$ -expansins, group-I grass pollen allergens, cell walls, wall loosening, arabinoxylan, monocots, commelinids.

## INTRODUCTION

Plant cell growth begins with wall stress relaxation, a process mediated at least in part by expansins (Cosgrove, 2005). Expansins are currently divided into four families: EXPA, EXPB, EXLA and EXLB (Sampedro and Cosgrove, 2005). Evidence of cell wall loosening activity is available for both  $\alpha$ -expansins (EXPA) and  $\beta$ -expansins (EXPB), and in both cases appears to involve destabilization of non-covalent, load-bearing contacts between cellulose microfibrils (McQueen-Mason and Cosgrove, 1995; Li and Cosgrove, 2001; Tabuchi *et al.*, 2011; Park and Cosgrove, 2012b).

Evolutionary analysis indicates that the last common ancestor of monocots and eudicots had at least 12 EXPA genes and two EXPB genes (Sampedro *et al.*, 2005). The

descendants of each of these genes form clades that are identified with roman numerals. In eudicots like Arabidopsis and poplar the EXPB family has remained small, with six and three genes, respectively, compared with 26 and 27 EXPA genes (Sampedro *et al.*, 2006). In contrast, in grasses the  $\beta$ -expansin family has expanded considerably, with rice having 18 EXPB versus 32 EXPA genes. This selective radiation of EXPB genes was exclusively the result of diversification of the EXPB-I clade, which has 16 genes in *Oryza sativa* (rice), compared with four in Arabidopsis and one in poplar.

The best-studied members of the EXPB family are proteins that accumulate to high levels in grass pollen and that belong to a highly divergent branch within the EXPB-I

clade. In *Zea mays* (maize) they are encoded by numerous genes and constitute 4% of the extractable protein in mature pollen grains (Li *et al.*, 2003; Valdivia *et al.*, 2007a). Reductions in  $\beta$ -expansin levels caused by a transposon insertion or by silencing did not impair pollen tube growth *in vitro*, but strongly impaired competition with normal pollen for reproductive success (Valdivia *et al.*, 2007b, 2009). These results, as well as the weakening of the middle lamella by EXPB1, the dominant  $\beta$ -expansin in maize pollen, indicate that these  $\beta$ -expansins facilitate intercellular penetration of the stylar tissues by the pollen tube (Tabuchi *et al.*, 2011). In addition to their physiological role these proteins are powerful elicitors of hay fever and seasonal asthma in humans; immunologists have named them group-1 grass pollen allergens (Petersen *et al.*, 2006).

Pollen  $\beta$ -expansins from maize loosen the cell walls of several grass species, but not of eudicot species (Li and Cosgrove, 2001; Li *et al.*, 2003). In contrast,  $\alpha$ -expansins display only weak activity on grass cell walls, but strong effects on eudicots, and also on some monocots with eudicot-like walls, such as *Allium cepa* (onion) or *Zephyranthes* (zephyr lily) (McQueen-Mason *et al.*, 1992; Cho and Kende, 1997). The different effects of grass pollen  $\beta$ -expansin on eudicot and grass cell walls are hypothesized to be the result of the distinctive structure of grass walls, which are enriched in arabinoxylan at the expense of pectins and xyloglucan (Carpita, 1996; Vogel, 2008). In maize walls arabinoxylan is hypothesized to link cellulose microfibrils into a load-bearing network, and to form the interstitial matrix between microfibrils (Carpita *et al.*, 2001). Maize pollen  $\beta$ -expansin preferentially binds to xylans *in vitro*, most likely on an open, conserved surface of the protein (Yennawar *et al.*, 2006). It also solubilizes highly substituted, feruloylated arabinoxylan from grass cell walls, accompanied by lesser quantities of homogalacturonan (Tabuchi *et al.*, 2011). Thus the arabinoxylan-cellulose interaction seems the likeliest target for the selective wall loosening of grass cell walls by grass pollen  $\beta$ -expansins.

In this work we adopted a phylogenetic framework to understand the molecular basis for the selectivity of grass pollen  $\beta$ -expansins for grass walls and the evolutionary origin of these divergent  $\beta$ -expansins. Our goals were: (i) to pinpoint the phylogenetic transition from the eudicot-type wall sensitivity to  $\alpha$ -expansin ( $\alpha$ -response) to the grass-type wall sensitivity to grass pollen  $\beta$ -expansins ( $\beta$ -response), (ii) to establish the phylogenetic occurrence of pollen with high  $\beta$ -expansin activity, (iii) to identify changes in the expansin protein sequence associated with the evolutionary emergence of the divergent subgroup of the EXPB-I clade that includes the grass pollen proteins, and (iv) to infer possible connections between these protein changes and the distinctive activity of the grass pollen

$\beta$ -expansins by mapping the changes onto the structure of ZmEXPB1.

Our results show that the predominance of arabinoxylan is not sufficient to render monocot walls susceptible to loosening by divergent EXPB-I proteins. The change in wall susceptibility (i.e. walls with  $\beta$ -response) occurred in a common ancestor of Poaceae, Juncaceae and Cyperaceae, in close association with the emergence and radiation of the divergent EXPB-I branch. Furthermore, the rapid evolution of this branch is likely to have involved a change in substrate preference that could be linked to a new structural role for highly substituted arabinoxylan.

## RESULTS

### Cell wall responses to expansins in an evolutionary framework

Previous studies noted a diversity of cell wall compositions within the monocots, spanning from eudicot-type walls in the lily family to the highly derived walls of grasses (Jarvis *et al.*, 1988; Carpita, 1996; Harris *et al.*, 1997; Hsieh and Harris, 2009; Harris and Trethewey, 2010). A major transition was the appearance of feruloylated arabinoxylans in primary walls, which are characteristic of the commelinid clade. To assess the relationship between the evolutionary changes in cell wall composition and the cell wall response to expansins, 10 monocot species were selected to cover the range of wall composition within this group. These species belong to six commelinid families (Poaceae, Cyperaceae, Juncaceae, Typhaceae, Commelinaceae and Musaceae) and two non-commelinid families (Asparagaceae and Araceae). Their phylogeny, based on the most recent analyses, is shown on the left side of Figure 1 (Givnish *et al.*, 2010; 'Hont *et al.*, 2012; Singh *et al.*, 2013). One eudicot species, *Cucumis sativus* (cucumber), was also included for comparison.

Heat-inactivated wall specimens from each species were clamped in a constant-force extensometer, and protein extracts from maize pollen (containing  $\beta$ -expansin ZmEXPB1) or cucumber hypocotyls (containing  $\alpha$ -expansin CsEXPA1) were added to the sample holder; the resulting increase in wall creep rate was compared for EXPA and EXPB treatments. The average increase in creep rate is shown in the middle columns of Figure 1. At the outset of these experiments, the extracts containing EXPA and EXPB were adjusted in concentration to result in a subsaturating creep response of  $\sim 6 \mu\text{m min}^{-1}$  in *Triticum aestivum* (wheat) coleoptile and cucumber segments, representing grass and eudicot wall responses, respectively (compare Figure 2a,b). Under these conditions, EXPA treatment caused an increase in creep rate seven times larger in cucumber walls than in wheat walls, whereas the reverse was true for EXPB treatment (compare the numbers in the

|             | Species                            | Response to expansins ( $\mu\text{m min}^{-1}$ ) |            | Expansin activity in pollen |   | Western blot |
|-------------|------------------------------------|--------------------------------------------------|------------|-----------------------------|---|--------------|
|             |                                    | $\alpha$                                         | $\beta$    | C                           | W |              |
| Poales      | <i>Zea mays</i>                    |                                                  |            | ○                           | ● | ●            |
|             | <i>Poa</i>                         |                                                  |            |                             |   |              |
|             | <i>Triticum aestivum</i>           | 0.6                                              | <b>4.4</b> | ○                           | ● | ●            |
|             | <i>Carex</i> sp.                   | 0.5                                              | <b>1.9</b> | ○                           | ● | ●            |
|             | <i>Cyp</i>                         |                                                  |            |                             |   |              |
|             | <i>Eleocharis</i> sp.              |                                                  |            | ○                           | ○ | ●            |
|             | <i>Scirpus</i> sp.                 | N.S.                                             |            | ○                           | ○ | ○            |
|             | <i>Jun</i>                         |                                                  |            |                             |   |              |
|             | <i>Juncus effusus</i>              | 0.2                                              | <b>1.6</b> | ○                           | ○ | ○            |
|             | <i>Typ</i>                         |                                                  |            |                             |   |              |
| Commelinids | <i>Typha</i> sp.                   | <b>1.6</b>                                       | 0.6        | ○                           | ○ | ○            |
|             | <i>Sparganium</i> sp.              |                                                  |            | ○                           | ○ | ○            |
|             | <i>Bro</i>                         |                                                  |            |                             |   |              |
|             | <i>Tillandsia</i> sp.              |                                                  |            | ○                           | ○ | ○            |
|             | <i>Com</i>                         |                                                  |            |                             |   |              |
|             | <i>Setcreasea purpurea</i>         |                                                  |            | ○                           | ○ | ○            |
|             | <i>Tradescantia fluminensis</i>    | <b>2.6</b>                                       | 0.2        | ○                           | ○ | ○            |
|             | <i>Mus</i>                         |                                                  |            |                             |   |              |
|             | <i>Musa nana</i>                   | N.S.                                             |            | ○                           | ○ | ○            |
|             | <i>Str</i>                         |                                                  |            |                             |   |              |
| Asterids    | <i>Strelitzia reginae</i>          |                                                  |            | ○                           | ○ | ○            |
|             | <i>Can</i>                         |                                                  |            |                             |   |              |
|             | <i>Canna</i> sp.                   |                                                  |            | ○                           | ○ | ○            |
|             | <i>Are</i>                         |                                                  |            |                             |   |              |
|             | <i>Coccothrinax argentata</i>      |                                                  |            | ○                           | ○ | ○            |
|             | <i>Syagrus romanzoffiana</i>       |                                                  |            | ○                           | ○ | ○            |
|             | <i>Phoenix</i> sp.                 |                                                  |            | ○                           | ○ | ○            |
|             | <i>Asp</i>                         |                                                  |            |                             |   |              |
|             | <i>Asparagus officinalis</i>       | <b>4.5</b>                                       | 0.1        |                             |   |              |
|             | <i>Agave americana</i>             | <b>4.4</b>                                       | 0.3        |                             |   |              |
| Eudicot     | <i>Ama</i>                         |                                                  |            |                             |   |              |
|             | <i>Zephyranthes grandiflora</i>    |                                                  |            | ○                           | ○ | ○            |
|             | <i>Xan</i>                         |                                                  |            |                             |   |              |
|             | <i>Heimerocallis</i> sp.           |                                                  |            | ○                           | ○ | ○            |
|             | <i>Orc</i>                         |                                                  |            |                             |   |              |
|             | <i>Phalaenopsis</i> sp.            |                                                  |            | ○                           | ○ | ○            |
| Eudicot     | <i>Ara</i>                         |                                                  |            |                             |   |              |
|             | <i>Pinellia ternata</i>            | <b>1.5</b>                                       | 0          | ○                           | ○ | ○            |
|             | <i>Spathiphyllum x clevelandii</i> |                                                  |            | ○                           | ○ | ○            |
| Eudicot     | <i>Cuc</i>                         |                                                  |            |                             |   |              |
|             | <i>Cucumis sativus</i>             | <b>5.8</b>                                       | 0.8        |                             |   |              |

**Figure 1.** Cell wall responses to expansins and presence of expansins in pollen of different monocots. Horizontal lines separate species from different families, which are indicated to the left by three-letter abbreviations: Poaceae (Poa), Cyperaceae (Cyp), Juncaceae (Jun), Typhaceae (Typ), Bromeliaceae (Bro), Commelinaceae (Com), Musaceae (Mus), Strelitziaceae (Str), Cannaceae (Can), Arecaceae (Are), Asparagaceae (Asp), Amaryllidaceae (Ama), Xanthorrhoeaceae (Xan), Orchidaceae (Orc), Araceae (Ara) and Cucurbitaceae (Cuc), as an eudicot reference. The response to expansins is shown as the average increase in creep rate after the addition of expansins from cucumber hypocotyls ( $\alpha$ ) and maize pollen ( $\beta$ ). When the difference is significant according to a Student's *t*-test ( $P < 0.05$ ), the strongest response is shown in bold with a grey background; N.S., responses to both types of expansin did not significantly differ from zero. Expansin activity in pollen extracts was tested on wheat (W) and cucumber (C) segments. Western blot experiments were performed using antibodies against *Lolium* group-1 pollen allergen Lol pl. Positive results are indicated with filled circles; negative results are indicated with empty circles.

middle panel of Figure 1 and see the creep responses in Figure 2a,b).

Among the non-grass monocots tested, only wall segments from *Carex* and *Juncus* species gave creep responses similar to the cell walls of wheat coleoptiles, showing significantly higher extension in response to

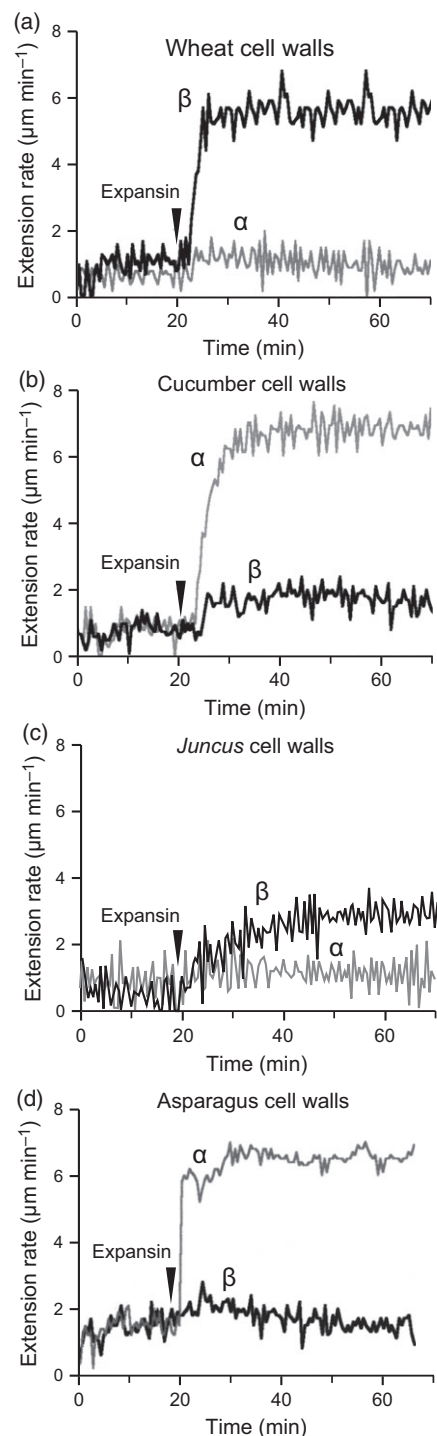


**Figure 2.** Differential cell wall responses of selected species to *Zea mays* (maize) pollen  $\beta$ -expansins (black) and *Cucumis sativus* (cucumber)  $\alpha$ -expansins (grey). Extension traces show two representative results for each species. The addition of expansins is indicated with black arrowheads.

maize pollen  $\beta$ -expansins (i.e.  $\beta$ -response; see Figures 1 and 2c). These two species belong to Cyperaceae and Juncaceae, the two families most closely related to Poa-

ceae (the grass family) of those tested (Figure 1). In contrast commelinid species from both Commelinaceae and Typhaceae, as well as all non-commelinid species tested, displayed an  $\alpha$ -response, with significantly stronger extension in response to cucumber EXPA proteins, as illustrated in Figure 2d for *Asparagus officinalis*. Results for two of the species tested (*Scirpus* sp. and *Musa nana*) were inconclusive because of very weak creep responses to both protein treatments. Weak responses are seen in wall segments with low growth rates. It is also possible that the extensometer conditions were not optimized for these materials. From these results we conclude that the 'grass-type' wall response to expansins, characteristic of Poaceae, is also present in two closely related families, Cyperaceae and Juncaceae, but is not a general property of the commelinid group nor of other groups more distant to Poaceae.

#### Phylogenetic survey of pollen $\beta$ -expansin activity in monocots

High  $\beta$ -expansin activity, assayed as induction of creep in wheat coleoptile cell walls, has previously been detected in pollen extracts from maize and two other grass species, *Phleum pratense* and *Lolium perenne*, in all three cases linked to the presence of EXPB-I proteins (Li and Cosgrove, 2001). To test whether expansin activity is common to pollen of other grasses, protein extracts were obtained from 19 additional species belonging to three subfamilies and seven tribes of Poaceae (Table 1). All the pollen extracts were capable of inducing an increase in the creep rate of inactivated wheat coleoptiles. Thus, this activity is a common feature of pollen from diverse grass species. We currently cannot determine if this character is ancestral for the family, as the earliest diverging Poaceae subfamilies, Anomochlooideae and Pharoideae, have not been tested.

To assess the evolutionary origin of this trait, pollen extracts from 20 monocot species from 13 different families were surveyed for creep activity (Figure 1). Whole-pollen extracts were applied to heat-inactivated cucumber and wheat walls clamped in a constant-force extensometer, as described above. Of the 20 species representing a diverse set of monocot groups, expansin activity in pollen was detected in only one Cyperaceae species in the genus *Carex*. The *Carex* expansin activity could induce creep on wheat cell walls but not on cucumber walls, showing the same specificity as pollen extracts from grasses (Figure 3a). Preliminary results also showed  $\beta$ -expansin activity in pollen extracts of *Cyperus alternifolius*, another Cyperaceae, but pollen was not available at the later date for these specific assays.

The same pollen extracts were analyzed by western blots with polyclonal antibodies against group-I allergens from *Lolium perenne* (Tamborini *et al.*, 1997). These

**Table 1** Grass species with extractable pollen expansin activity

| Subfamily                 | Tribe                | Species                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Panicoideae               | Paniceae             | <i>Paspalum notatum</i>                                                                                                                                                                                                                                                                                                                                                                                      |
|                           | Andropogoneae        | <i>Sorghum halepense</i><br><i>Zea mays</i> <sup>a</sup>                                                                                                                                                                                                                                                                                                                                                     |
| Chloridoideae<br>Pooideae | Cynodonteae<br>Poeae | <i>Cynodon dactylon</i><br><i>Lolium perenne</i> <sup>a</sup><br><i>Poa pratensis</i><br><i>Poa compressa</i><br><i>Festuca arundinacea</i><br><i>Dactylis glomerata</i><br><i>Lolium multiflorum</i><br><i>Bromus inermis</i><br><i>Phalaris arundinacea</i><br><i>Avena sativa</i><br><i>Holcus lanatus</i><br><i>Phleum pratense</i> <sup>a</sup><br><i>Anthoxanthum odoratum</i><br><i>Agrostis alba</i> |
|                           |                      | <i>Elymus repens</i>                                                                                                                                                                                                                                                                                                                                                                                         |
|                           |                      | <i>Pascopyrum smithii</i>                                                                                                                                                                                                                                                                                                                                                                                    |
|                           |                      | <i>Leymus condensatus</i>                                                                                                                                                                                                                                                                                                                                                                                    |
|                           |                      | <i>Triticum aestivum</i>                                                                                                                                                                                                                                                                                                                                                                                     |
|                           |                      | <i>Secale cereale</i>                                                                                                                                                                                                                                                                                                                                                                                        |
|                           | Bromeae<br>Aveneae   |                                                                                                                                                                                                                                                                                                                                                                                                              |
|                           |                      |                                                                                                                                                                                                                                                                                                                                                                                                              |
|                           | Triticeae            |                                                                                                                                                                                                                                                                                                                                                                                                              |
|                           |                      |                                                                                                                                                                                                                                                                                                                                                                                                              |
|                           |                      |                                                                                                                                                                                                                                                                                                                                                                                                              |
|                           |                      |                                                                                                                                                                                                                                                                                                                                                                                                              |

<sup>a</sup>Species where this activity has been previously detected (Li and Cosgrove, 2001).

antibodies had previously been shown to recognize pollen  $\beta$ -expansins from maize, which is quite distantly related to *Lolium* within Poaceae (Li *et al.*, 2003). Of the 20 non-grass monocots tested, similar proteins were only detected in *Carex* and in a second Cyperaceae species from the genus *Eleocharis* in which no expansin (wall creep) activity had been detected (Figure 3b). Antibodies may be more sensitive than activity measurements, especially as proteins may lose activity. In both *Carex* and *Eleocharis*, as well as maize and wheat, the antibodies recognized one or more protein bands in the range 30–35 kDa (Figure 3b), which is characteristic of pollen  $\beta$ -expansins. The other 18 species, representing a diversity of monocot groups, did not show evidence of high  $\beta$ -expansin content in pollen extracts by this immunoassay. This included *Juncus effusus*, the walls of which extended in response to maize pollen expansins. The results of these activity and antibody assays indicate that pollen with high  $\beta$ -expansin content is restricted to Poaceae and Cyperaceae among the tested species, a distribution that is similar to (though slightly narrower than) that of cells walls displaying a  $\beta$ -response to applied expansins.

#### Sequence and expression of *Carex* EXPB genes

In order to identify the protein responsible for the expansin activity detected in *Carex*, RT-PCR was performed on *Carex* pollen RNA. Degenerate primers, designed using an alignment of grass pollen  $\beta$ -expansins, were used to amplify part of a  $\beta$ -expansin gene. This sequence was then used to

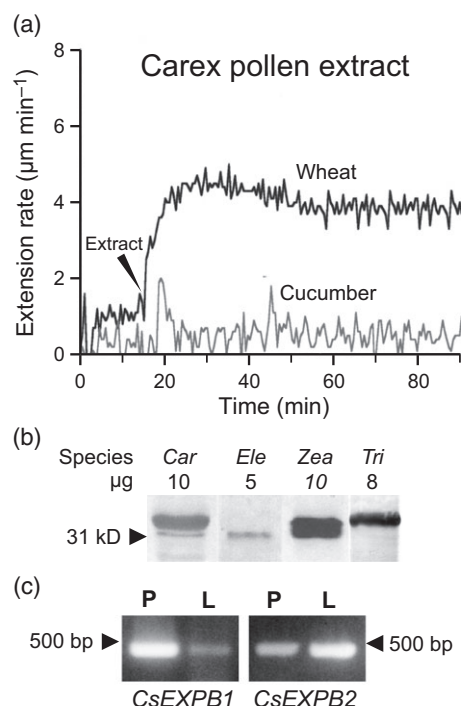


**Figure 3.** Divergent EXPB-I expansins in Cyperaceae.

(a) Effect of *Carex* pollen extracts on the creep rate of *Triticum aestivum* (wheat) coleoptiles and *Cucumis sativus* (cucumber) hypocotyl segments.

(b) Western blot of pollen extracts from *Carex* (Car), *Eleocharis* (Ele), maize (Zea) and wheat (Tri). The quantity of extract used in each lane is indicated below the species. The membrane was probed with antibodies against group-1 allergens from *Lolium perenne*. The location of a 31-kDa protein marker is indicated to the left.

(c) Amplification of *CsEXPB1* and *CsEXPB2* from pollen (P) and leaf (L) cDNA. The location of a 500-bp marker is indicated.

clone a full-length cDNA that was named *CsEXPB1* (GenBank accession no. KM396910). Using a similar strategy, a second expansin cDNA, named *CsEXPB2* (GenBank accession no. KM396909), was cloned from growing leaves of the same *Carex* species. The primers in this case were designed on the basis of a wider alignment of EXPB-I proteins. RT-PCR indicated that *CsEXPB1* was highly expressed in pollen, as well as at a lower level in leaves (Figure 3c). The cDNA sequence amplified from leaves showed only 1 bp difference in 428 bp, compared with *CsEXPB1*, and thus may represent an allelic variant. With RNA samples collected from several individuals we amplified a number of closely related sequences for both *CsEXPB1* in pollen and *CsEXPB2* in leaves. Sequences highly similar to *CsEXPB2* were also amplified from pollen RNA, although this gene seems to be more highly expressed in leaves (Figure 3c).

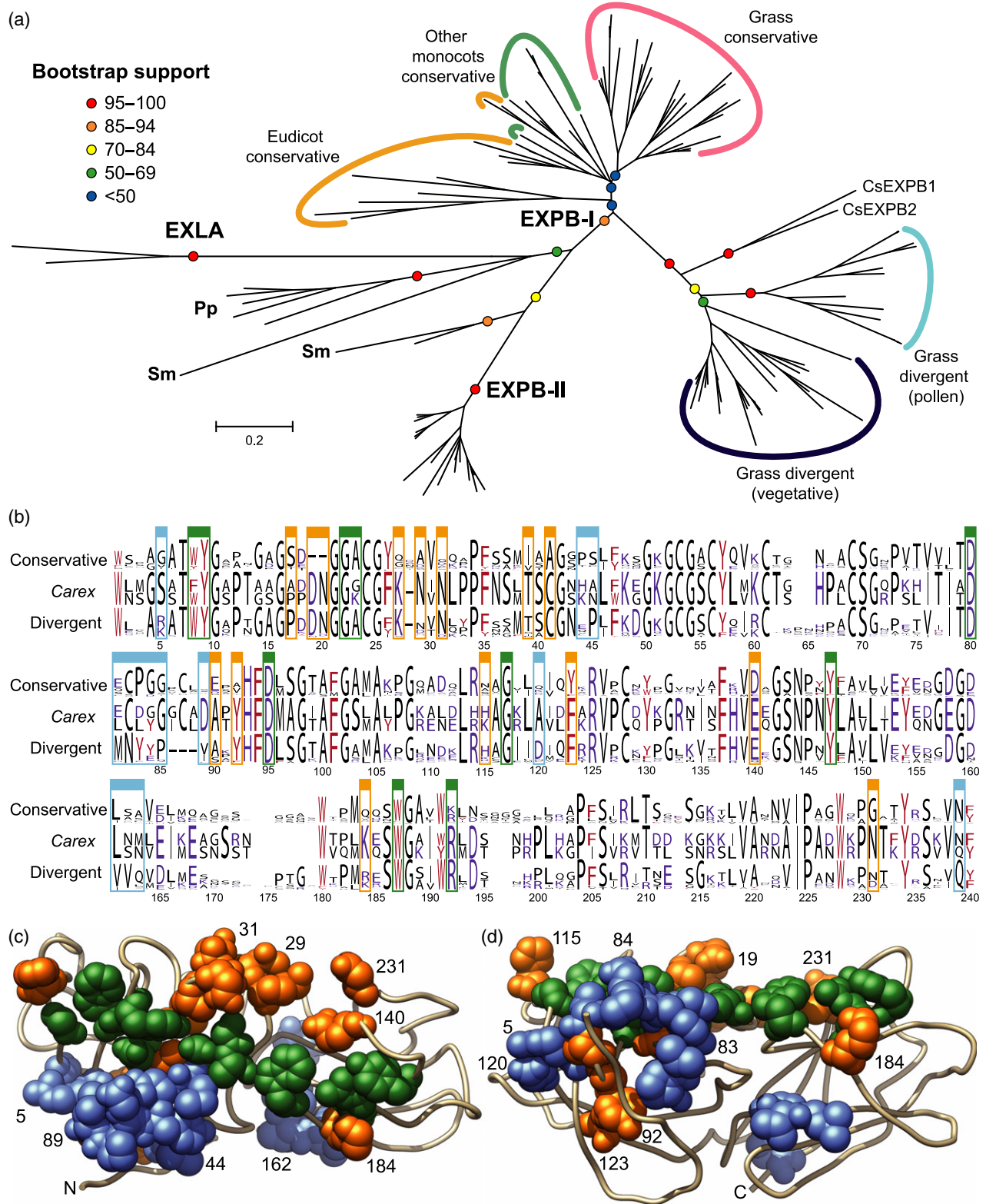
#### Evolutionary divergence of the EXPB-I clade in monocots

A phylogenetic analysis of the EXPB family was carried out with an alignment of protein sequences translated from 11

plant genomes, as well as the *Carex* expansins (Figure S1). In addition to three grasses (rice, sorghum, and *Brachypodium*), the analysis included the  $\beta$ -expansins in the non-grass monocot genomes of *Musa acuminata*, *Elaeis guineensis* and *Phoenix dactylifera*. Three divergent eudicot species (*Arabidopsis*, grapevine and tomato), as well as two non-flowering plants (*Selaginella moellendorffii* and *Physcomitrella patens*), were also included. The root of the tree could be located between the EXLA and EXPB families or, more likely, in the branch leading to the *Physcomitrella* expansins, as neither this species nor *Selaginella* has EXLA genes (Carey *et al.*, 2013). A maximum-likelihood phylogenetic reconstruction, shown in Figure 4a, gives strong support for monophyletic EXPB-I and EXPB-II angiosperm clades (86 and 99% bootstrap values). Both *Carex* expansins appear in a divergent branch that includes grass pollen  $\beta$ -expansins and a sister group of proteins that are expressed in vegetative tissues, judging from expressed sequence tag (EST) data and expression studies (Lee and Kende, 2001). We will refer to the sequences in this branch as divergent EXPB-I genes, and to the remaining EXPB-I sequences as conservative. A similar tree constructed with neighbor-joining, a method that is more prone to long-branch attraction, results in a similar topology, but with less support for the monophyly of EXPB-I (Figure S2).

This analysis shows that the separation of the divergent and conservative EXPB-I branches happened before Cyperaceae and Poaceae diverged, but cannot resolve the timing further. All the relevant nodes have bootstrap values below 50% in both maximum-likelihood and neighbor-joining analyses (Figure 4a; Figure S2); however, a previous study of rice expansins showed that divergent and conservative EXPB-I genes are found in paralogous genomic regions that were created by segmental duplication (Sampedro *et al.*, 2005). A later analysis of the rice and sorghum genomes found that these paralogous regions were the result of a whole-genome duplication, denoted as 'sigma' (Tang *et al.*, 2010). Divergent and conservative EXPB-I genes in rice are found in duplicated regions that form the  $\sigma 6$  block (Figure 5). The recent sequencing of the banana genome has revealed that the  $\sigma$  polyploidy event took place in Poales after they split from a sister group that includes Zingiberales, Commelinales and Arecales ('Hont *et al.*, 2012).

We found no evidence of divergent EXPB-I genes in the banana, *Phoenix dactylifera* (date palm) or *Elaeis guineensis* (oil palm) genomes (Figure 4a). In addition to the full-length sequences included in the phylogenetic tree, all available fragments of  $\beta$ -expansin genes from these three genomes belong to the conservative EXPB-I branch or the EXPB-II clade (Table S2). Finally, no divergent EXPB-I genes were found in 27 eudicot genomes available in Phytome, or among non-grass protein sequences deposited in GenBank. These results, together with the duplication history of the genomic regions where divergent and con-



servative grass EXPB-I genes are located, support a recent separation of the two branches, in stem Poales at the earliest (Figure 1). Rice divergent EXPB-I genes expressed

in pollen are located a short chromosomal distance from those expressed in vegetative tissues, probably as the result of a short-range translocation event that took place

**Figure 4.** Origin and differentiation of divergent EXPB-I proteins.

(a) Maximum-likelihood phylogenetic tree of EXPB proteins from *Carex* as well as 11 plant genomes. Bootstrap support values based on 200 replicates are indicated on selected branches by colored circles (see key in panel). The two *Carex* proteins cloned in this work are named CsEXPB1 and CsEXPB2. Sm and Pp indicate *Selaginella* and *Physcomitrella* sequences, respectively (these are not angiosperms). All sequences are identified in Figure S1.

(b) Alignment of protein sequence logos for conservative EXPB-I proteins, *Carex* expansins and divergent EXPB-I proteins from grasses. Green boxes mark conserved amino acids that form part of the conserved central binding groove. Sites where divergent and conservative EXPB-I proteins show different conserved residues were identified with SPEER-SERVER. Sites where *Carex* and grass sequences have similar conserved amino acids are boxed in orange; sites where they show differences are boxed in blue. Aromatic amino acids are brown and charged amino acids are purple.

(c, d) Top and side views of ZmEXPB1 with conserved and divergent residues shown as spheres, colored as in (b). Selected residues are identified with numbers corresponding to the alignment in (b).

before the later  $\rho$  whole-genome duplication of grasses (Figure 5).

In addition to the radiation of the divergent EXPB-I branch, the conservative EXPB-I group also appears to have diversified in grasses. The three species we analyzed have between nine and 18 divergent EXPB-I genes, not counting duplicates that code for identical proteins, and between six and 11 conservative EXPB-I genes. In contrast the genomes of eudicots and other monocots only have between one and six conservative EXPB-I genes. Thus the EXPB-I clade shows a remarkable radiation and diversification specific to the grass lineage, which might have involved positive selection.

#### Evidence of functional differentiation in divergent EXPB-I proteins

To identify the specific sequence changes responsible for the phylogenetic differentiation of the divergent and conservative EXPB-I branches, we used SPEER-SERVER to identify protein residues that show different conservation patterns in the two groups (Chakraborty *et al.*, 2012). Twenty-seven sites with significance  $P < 0.1$  were mapped onto an alignment of three sequence logos, one for grass divergent EXPB-I proteins, another for the two *Carex* expansins, and a third for conservative EXPB-I proteins from both monocots and eudicots (Figure 4b). A two-amino acid insertion, highly conserved in the divergent branch (positions 19–20), was also included in the analysis, since the software ignores insertions and deletions. In 15 of the 29 changes in amino acid preferences, the *Carex* sequences share the conservation pattern observed in the grass divergent EXPB-I group (orange boxes in Figure 4b). In the remaining 14 positions the *Carex* sequences are either closer to conservative EXPB-I sequences or show their own conservation pattern (blue boxes in Figure 4b). The *Carex* expansins thus appear to occupy a position intermediate between divergent grass EXPB-I and conservative sequences. The amino acids that form the proposed binding surface of the protein (boxed in green) are highly conserved and similar in the three groups, with the exception of position 23, where *Carex* expansins show a different preference.

To understand the possible functional significance of these protein modifications, the changes in amino acid

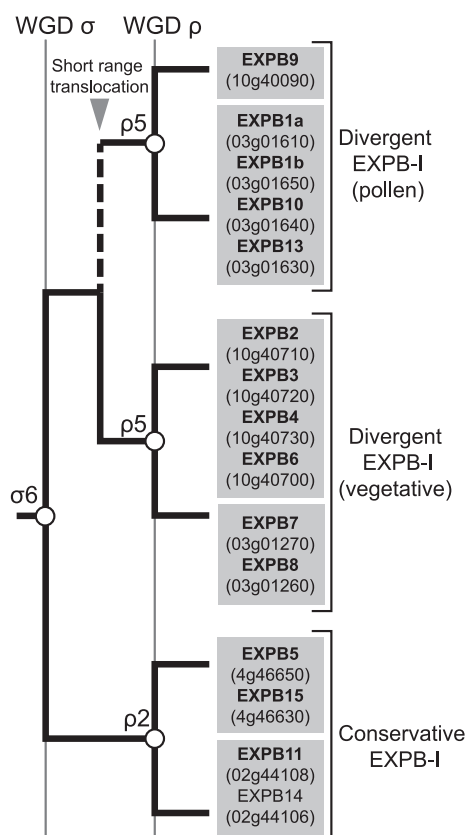


**Figure 5.** Duplication history of paralogous genomic regions in rice that contain EXPB-I genes. Nodes that correspond to whole-genome duplications (WGDs) are identified by the codes used for duplicated blocks in Tang *et al.* (2010). A dashed line indicates a short-range translocation event. Expansin genes included in each genomic region (grey blocks) are listed with their annotation.

conservation pattern were mapped onto the 3D structure of ZmEXPB1, a divergent EXPB-I protein from maize pollen (Yennawar *et al.*, 2006). Most of the changes are located on the protein face proposed to interact with bound polysaccharide (Figure 4c,d). The changes specific to grass divergent EXPB-I proteins (blue spheres) are mainly located in a large cluster situated to one side of the open binding groove (green spheres). This region of blue spheres also encompasses the location of a three-amino-acid deletion specific to divergent grass EXPB-I expansins (positions 86–88 in Figure 4b). In this region conservative

and *Carex* EXPB-I proteins have a conserved pair of cysteines that could form a disulfide bond, but the cysteines are missing in grass divergent EXPB-I proteins (positions 82 and 87). These cysteines are also present in EXPB-II expansins and in most  $\alpha$ -expansins (Sampedro and Cosgrove, 2005). The changes shared by both *Carex* and grass divergent EXPB-I proteins (orange spheres) are located mainly on the opposite side of the binding groove, spanning both domains. This analysis suggests that the evolution of the divergent EXPB-I expansins involved a focused remodeling of the edges of the binding surface, not affecting the central binding groove, and that the changes progressed further in the grass lineage after the divergence of Cyperaceae.

## DISCUSSION

### Evolution of wall composition and expansin responsiveness in monocots

We found that the  $\beta$ -response pattern, previously documented for grass cell walls, is also found in species from Cyperaceae and Juncaceae (Figure 1), whereas other commelinids and monocot species more distant to grasses display an  $\alpha$ -response pattern, behaving similarly to eudicots in this assay. The most parsimonious explanation is that the change in wall responsiveness took place in a common ancestor of Poaceae, Cyperaceae and Juncaceae, after the divergence of Typhaceae (black node in Figure 1). This dating could be narrowed further when future work determines the expansin responsiveness of the Rapateaceae family, which diverged during this interval. The wall structure evolved from one that is sensitive to the action of  $\alpha$ -expansin to one that is preferentially loosened by  $\beta$ -expansin from grass pollen. In view of the large compositional differences between eudicot and grass walls, this may have been a complex set of changes.

Several lines of evidence point to a xyloglucan-cellulose junction as the most likely target for  $\alpha$ -expansin action. In addition to binding preferentially to this structure (McQueen-Mason and Cosgrove, 1995),  $\alpha$ -expansins are less effective on cell walls of an *Arabidopsis* mutant lacking xyloglucan (Park and Cosgrove, 2012a), and  $\alpha$ -expansin elicits a larger extension in artificial cellulose composites containing xyloglucan compared with those made of cellulose alone (Whitney *et al.*, 2000). Consistent with this, NMR analysis showed that a bacterial expansin selectively targeted a site in the *Arabidopsis* cell wall containing a modified form of cellulose in close proximity to xyloglucan (Wang *et al.*, 2013). This NMR-detected structure is very similar to the 'biomechanical hot spots', limited cellulose-xyloglucan-cellulose junctions that control wall extensibility, proposed by Park and Cosgrove (2012b). Thus the preponderance of evidence indicates  $\alpha$ -expansin loosens a mechanically important xyloglucan-cellulose junction.

The fraction of xyloglucan in eudicot walls with this mechanical function is apparently small because much of the wall xyloglucan could be digested with a xyloglucan-specific endoglucanase, without causing wall creep or changing cell wall compliance (Park and Cosgrove, 2012b). In contrast, digestion of <1% of the wall xyloglucan by endoglucanases able to digest both xyloglucan and cellulose resulted in a high creep rate and large increases in wall compliance. From these and related results emerges the concept that biomechanical properties of primary walls such as creep and extensibility are controlled by specific interactions among wall components (at biomechanical hot spots) rather than by bulk polymer viscoelasticity. This concept may also account for some of the perplexing results found in the current work.

Xyloglucan content in grass cell walls (~5%) is much lower than in eudicots (~20%), and it has been suggested that arabinoxylan, which is highly abundant in grasses (20–40%), plays an equivalent role in these walls (Carpita and Gibeaut, 1993; Carpita, 1996; Vogel, 2008). Arabinoxylan-rich walls have been termed 'type II', in contrast with xyloglucan-dominated 'type-I' walls. Our initial expectation, therefore, was that the  $\beta$ -response pattern would be associated with 'type-II' walls, but this was not borne out by our results. Although high arabinoxylan content is a common wall property of commelinid monocots (Harris *et al.*, 1997; Harris and Trethewey, 2010), we found that *Typha* sp. and *Tradescantia fluminensis* showed an  $\alpha$ -response pattern resembling that of eudicots (Figure 1). The monosaccharide composition of primary walls in these two species suggests an arabinoxylan content similar to that of grasses (Harris *et al.*, 1997). Thus, our results show that the presence of arabinoxylan, even in large quantities, is not sufficient to confer responsiveness to maize pollen  $\beta$ -expansins or to prevent wall loosening by  $\alpha$ -expansin, as in 'type-I' walls. Evidently a subset of commelinids (grasses and closely related families) evolved structural modifications that altered their response to expansins, without large changes in arabinoxylan content.

The change from  $\alpha$ -response to  $\beta$ -response does not appear to be linked with changes in xyloglucan structure. Although grass xyloglucan is based on XXG<sub>n</sub> subunits, lacks fucose and has reduced galactosylation, Cyperaceae and Juncaceae have an XXXG type of xyloglucan, like eudicots, with numerous galactose and fucose side chains (Hsieh and Harris, 2009). A similar xyloglucan is also found in *Typha*, which shows an  $\alpha$ -response (Figure 1). A reduction in xyloglucan content, on the other hand, might have been part of the change in expansin response within Poales. Xyloglucan content has been estimated at 16% in Bromeliaceae, a level closer to eudicots than to grasses (Smith and Harris, 1995). Bromeliaceae is likely to be the sister group of all other Poales. In contrast, a low proportion of xyloglucan was found in a few families more

closely related to grasses than Juncaceae and Cyperaceae (Smith and Harris, 1999). Future work to quantify xyloglucan contents in Typhaceae, Juncaceae and Cyperaceae would help to establish if its reduction coincided with the change in expansin response we have observed.

We also note that the change in cell wall susceptibility to expansins coincides with the appearance of (1→3),(1→4)- $\beta$ -glucans, which are well-known components of the primary wall in Poaceae, and have been reported in Juncaceae and Cyperaceae, but have not been found in *Typha* and other early diverging Poales (Trethewey *et al.*, 2005). The (1→3),(1→4)- $\beta$ -glucans in Juncaceae and Cyperaceae have only been detected with antibodies, however, and are evidently present in trace quantities. In addition, the antibody signal was strongest in external epidermal walls and secondary walls, with only occasional labeling in other walls, a pattern that does not suggest a role in cell expansion. Finally, maize pollen  $\beta$ -expansins showed negligible binding to this polymer, making it an unlikely target for these proteins (Yennawar *et al.*, 2006).

In view of the foregoing discussion and prior results implicating xylans in  $\beta$ -expansin action (Yennawar *et al.*, 2006; Tabuchi *et al.*, 2011), we propose that the selective  $\beta$ -response of walls from Poaceae and closely related families involved a subset of arabinoxylans with a unique biomechanical role that evolved after the grass lineage diverged from the common ancestor with Typhaceae. The reduction of pectin and xyloglucan contents, and their replacement by arabinoxylan in the branch leading to comelinids, may have set the stage for the evolution of a new biomechanical role for arabinoxylan, possibly also involving (1→3),(1→4)- $\beta$ -glucans. A recent study of polysaccharide epitopes in maize root walls (Kozlova *et al.*, 2014) proposed a possible arrangement of cellulose, arabinoxylan and (1→3),(1→4)- $\beta$ -glucan that might resemble the type of structure proposed here. It should also be noted that the size of protein families that act on xyloglucan, such as EXPA or XTH (xyloglucan endotransglycosylase/hydrolase) is similar in grasses and eudicots (Yokoyama *et al.*, 2004; Sampedro *et al.*, 2005), suggesting that xyloglucan still has an important function in grasses.

### Origin and evolution of divergent EXPB-I proteins

Our phylogenetic analysis shows that divergent EXPB-I genes are not present in eudicots or in the recently sequenced banana, date palm and oil palm genomes. The two *Carex* genes reported in this work are the only divergent EXPB-I genes identified from a non-grass species (Figure 4). Although we do not have direct proof of their activity, it is likely that these or related genes are responsible for the expansin activity on grass cell walls detected in pollen from the same *Carex* species (Figure 3a). Therefore, the selective ability to loosen grass cell

walls appears to be a common characteristic of divergent EXPB-I expansins. This selectivity may be unique to the divergent group of  $\beta$ -expansins, but this point has not been tested at this time because other  $\beta$ -expansins have not been isolated in active form (e.g. Lee and Choi, 2005).

Whereas phylogenetic trees cannot date the separation of the divergent EXPB-I group from the conservative EXPB-I group, it appears to have been a product of the sigma whole-genome duplication (Figures 4a and 5), an event that can be detected in grasses, but not in banana (Hont *et al.*, 2012). The lack of divergent EXPB-I genes in all non-grass sequenced genomes also supports a recent split, sometime between the first Poales and the last common ancestor of Juncaceae, Cyperaceae and Poaceae. As was previously discussed, the change from  $\alpha$ -response to  $\beta$ -response in cell walls seems to have evolved after the divergence of Typhaceae. These results suggest that the appearance of cell walls with a  $\beta$ -response happened coincidentally or shortly after the origin of the divergent EXPB-I proteins capable of loosening these new walls. Both the change in cell wall responsiveness and the evolution of divergent EXPB-I genes may have been made possible by the sigma polyploidy event. Whole-genome doublings, by maintaining gene balance, allow the retention of duplicated networks of interacting genes, providing raw material for phenotypic innovations (Freeling, 2009).

The great length of the divergent EXPB-I branch suggests accelerated evolution through positive selection (Figures 4a and S2), and the comparison of conservative and divergent EXPB-I proteins shows that many of the changes happened in residues flanking the conserved polysaccharide binding groove (Figure 4b–d). Some of the differences in amino acid preference identified in the divergent EXPB-I group are absent in both *Carex* sequences (blue in Figure 4b–d). These changes, which are grouped in two clusters, do not seem to be necessary for the ability to loosen grass cell walls (Figure 3a). Of the fifteen changes common in both *Carex* and grass divergent EXPB-I proteins (orange in Figure 4b–d), a cluster of six is located at the surface and grouped around a two-amino acid insertion that is diagnostic for the divergent EXPB-I branch (positions 19 and 20). In the ZmEXPB1 structure an aspartic acid residue (D19) in this insertion is ideally located to interact with side chains of a polysaccharide bound to the conserved groove. Another two changes (positions 39 and 41) are internal residues located just below the binding groove in domain 1 and a tight cluster of three amino acid replacements (positions 90, 92 and 123) subtend a shoulder that flanks the shallow binding groove, and that was extensively remodeled in Poaceae. Finally, four changes shared by *Carex* and grasses (115, 140, 184 and 231) are located at the surface, and are occupied in divergent EXPB-I proteins by charged or polar residues suitably positioned for substrate recognition.

It seems likely that these changes in divergent EXPB-I proteins resulted in an alteration in preferred substrate, perhaps even mode of action, and simultaneously generated the specific epitopes recognized by immunoglobulins from individuals allergic to group-1 grass pollen allergens (Flicker *et al.*, 2006; Ball *et al.*, 2009). Unfortunately there is currently no biochemical evidence concerning the activity, substrate or allergenic epitopes of conservative EXPB-I proteins. One hypothesis that could explain the differences in preferred amino acids is that conservative and divergent EXPB-I proteins recognize xylans with different patterns of side chains, e.g. pattern or degree of arabinose substitution.

Although there are few detailed analyses, the general structure of arabinoxylan appears to be similar within commelinids, with arabinosyl substitutions mainly at O-3, some of them ester-linked to ferulic acid (Smith and Harris, 1995, 1999; Carnachan and Harris, 2000; Harris and Trethewey, 2010, 2010). The ratio of arabinose to xylose in primary walls is similar in most commelinid families (Harris *et al.*, 1997). Lacking more detailed analyses of arabinoxylans in different commelinids, the evolution of divergent EXPB-I proteins cannot be explained at this moment by a change in arabinoxylan linkage structure or degree of substitution.

Highly substituted arabinoxylan is the major component of the interstitial material of primary walls in grasses, whereas secondary walls contain mainly arabinoxylan with a low degree of substitution, more similar to xylan in eudicot secondary walls (Smith and Harris, 1999; Suzuki *et al.*, 2000; Scheller and Ulvskov, 2010). A possible evolutionary scenario is that secondary wall xylan was the ancestral substrate of conservative EXPB-I expansins (present in both eudicot and monocot genomes), and that divergent EXPB-I proteins evolved to act on highly substituted xylans, enabling enhanced roles in the mechanics of primary cell walls for both the proteins and their polysaccharide targets.

In this work we have tried to elucidate the capacity of grass pollen expansins to loosen grass cell walls in an evolutionary framework. We discovered important differences in the response to expansins within commelinid monocots, indicating an important but poorly defined change in wall structure within Poales. The abundance of arabinoxylan in commelinid primary walls is clearly not sufficient to make them susceptible to grass pollen  $\beta$ -expansins. A tentative hypothesis that would tie together all strands of the evidence is that when highly substituted arabinoxylan first evolved in commelinid primary walls,  $\beta$ -expansins had limited action on this polymer. After a whole-genome duplication in Poales the evolution of one of the copies of an EXPB-I gene resulted in a  $\beta$ -expansin with changes in surface amino acids that allowed it to act on highly

substituted arabinoxylan with greater effectiveness. The expression of divergent EXPB-I genes in growing tissues would have enabled arabinoxylan to supplant xyloglucan as a major structural (load-bearing) element of the wall and as a dominant target of expansin-mediated cell wall loosening in an ancestor of Poaceae, Juncaceae and Cyperaceae. Its modern descendants comprise more than 16 000 species, or 30% of all monocots, and include major food crops such as rice, wheat, maize, sorghum and sugarcane, as well as potential biofuel crops such as switchgrass. Future studies of cell wall evolution within Poales, as well as the characterization of the activity of conservative EXPB-I proteins, could test this hypothesis and clarify the evolution and functional differences of the highly modified primary walls of grasses.

## EXPERIMENTAL PROCEDURES

### Plant material

For wall extension assays, 2-cm apical segments were excised from wheat (*Triticum aestivum*) coleoptiles, cucumber (*Cucumis sativus*) hypocotyls and asparagus (*Asparagus officinalis*) stems, from seedlings germinated in darkness on wet vermiculite. Basal portions of leaves and pollen were collected from wild populations at Waddle Creek (State College, PA, USA) for Cyperaceae and Juncaceae species, and at Stone Valley (Huntington, PA, USA) for *Typha* sp. Pollen from Arecaceae species was obtained from Dr Alan Meerow at USDA, Miami, FL, USA. Maize pollen was collected from plants grown in the field at State College, PA, USA. Pollen from other grasses was purchased from Greer Laboratories, Lenoir, NC, USA. Other pollen and plant segments were obtained from the collections maintained at Penn State University (State College, PA, USA). Segments were obtained from floral bracts (*Musa nana* and *Agave americana*) or pedicels (*Tradescantia fluminensis* and *Pinellia ternata*).

### Protein extractions

Cucumber expansins were extracted as described previously (McQueen-Mason *et al.*, 1992). In brief, 3-cm apical segments of cucumber hypocotyls were homogenized with 20 mM sodium acetate (pH 4.5) and 2 mM EDTA in a Waring blender. The homogenate was then filtrated through a nylon screen (70- $\mu$ m mesh) and the retained fraction washed twice before being extracted in 1 M NaCl, 20 mM HEPES (pH 6.8), 2 mM EDTA and 3 mM sodium metabisulfite, for 1 h at 4°C. After removing cell wall fragments by filtration, ammonium sulfate fractionation was carried out to separate the material precipitated between 20% and 60% saturation. This precipitate was resuspended in 50 mM sodium acetate (pH 4.5), and the resulting solution was used in wall extension assays, adjusted in concentration to induce an increase in creep rate of  $\sim 6 \mu\text{m min}^{-1}$  for heat-inactivated walls of cucumber hypocotyls.

Expansins from pollen, stored at  $-80^\circ\text{C}$ , were extracted by suspending thawed pollen in 50 mM sodium acetate (pH 4.5), at a concentration of 250 mg of pollen per ml of buffer, for 1 h at 4°C under mild agitation. After centrifugation, the supernatant was used in wall extension assays or western blotting experiments. For the experiments shown in Figures 1 and 2, the extract concentration was adjusted to yield an increase in creep rate of  $\sim 6 \mu\text{m min}^{-1}$  for heat-inactivated walls of wheat coleoptiles.

### Wall extension assays

For most species plant segments (1–2 cm) kept frozen at  $-20^{\circ}\text{C}$  were thawed and then abraded to disrupt the cuticle by rubbing them with a carborundum slurry (320 grit; Fisher Scientific, now Thermo Scientific, <http://www.thermoscientific.com>). Segments were then submerged in boiling distilled water for 15 s to inactivate endogenous expansin activity. To retard oxidation, the basal portions of leaves from Cyperaceae, Juncaceae and Typhaceae species were first submerged in 5 mM ascorbic acid, immediately boiled, and then segments were frozen, thawed and abraded.

Treated segments were clamped in a constant load extensometer (Cosgrove, 1989), 50 mM HEPES (pH 6.8) was added to the cuvettes and a 20-g weight was applied to provide the tension. The bathing solution was changed, after approximately 15 min, to 50 mM sodium acetate (pH 4.5), to test the inactivation of endogenous expansins. After 20 min, if there was no change in extension rate, protein extracts in the same buffer were added and the extension rate monitored for at least 1 h.

### Western blot

The protein concentration of pollen extracts was quantified with Coomassie Protein Assay Reagent (Pierce, <http://www.pierce-net.com>). Proteins were separated by SDS-PAGE in 15% polyacrylamide gels using a Protean II mini-gel apparatus (Bio-Rad, <http://www.bio-rad.com>), and then transferred electrophoretically to Protran BA nitrocellulose membranes (Whatman, now GE Life Sciences, <http://www.gelifesciences.com>) with the Mini Trans-Blot Module (Bio-Rad). Transfer was carried out overnight at  $4^{\circ}\text{C}$  in a solution of 192 mM glycine, 25 mM Tris and 20% (v/v) methanol, at 30 V. Membranes were washed and blocked with 10% (v/v) horse serum in phosphate-buffered saline containing 0.05% (v/v) Tween 20 and 5 mM sodium azide. Rabbit polyclonal antibodies against Lol p1 (Tamborini *et al.*, 1997) were diluted 1:150 000 in the blocking solution and incubated with the membrane for 1 h. After repeated washes, membranes were incubated for 1 h with alkaline phosphatase conjugated anti-rabbit IgG (Rockland, <http://www.rockland-inc.com>) diluted 1:8000 in Tris-buffered saline. Membranes were washed again and developed with a solution of 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium in 100 mM Tris (pH 9.5), 100 mM NaCl and 5 mM  $\text{MgCl}_2$ .

### Cloning of cDNAs and RT-PCR

Extractions of RNA from pollen and growing leaves of *Carex* were carried out with the RNeasy Plant Mini Kit (Qiagen, <http://www.qiagen.com>) and quantified by UV absorbance. Numerous combinations of degenerate primers were tested until an expansin gene fragment of 255 bp was amplified from pollen RNA, and another fragment of 194 bp was amplified from leaf RNA. Full-length cDNAs were cloned with GeneRacer Kit with Superscript III RT module (Invitrogen, now Life Technologies, <http://www.lifetechnologies.com>). In brief, 800 ng of pollen RNA and 4  $\mu\text{g}$  of leaf RNA were dephosphorylated, decapped, ligated to an RNA oligo and reverse transcribed. The 5' and 3' regions of *CsEXPB1* were amplified, respectively, with 5POL and 3POL primers (Table S1). For *CsEXPB2*, 5VEG was used to obtain the 5' region, whereas two rounds of nested amplification with 3VEG and 3VEGN were required for the 3' region. Gene fragments were cloned using the TOPO TA cloning kit (Invitrogen), and between four and eight clones were sequenced in each case. Full-length cDNAs were obtained using primer pairs POLFL/POLFR (*CsEXPB1*) and VEGFL/VEGFR (*CsEXPB2*), designed on the basis of the longest and most common variants among the sequenced fragments.

To analyze the expression of the *Carex* genes, 50 ng of pollen or leaf RNA were reverse transcribed using random nonamers with Enhanced Avian HS RT-PCR Kit (Sigma-Aldrich, <http://www.sigmaaldrich.com>), and 1  $\mu\text{l}$  of the reactions was used as the template for amplification. *CsEXPB1* was amplified with POLCL and POLCR primers for 40 cycles, with annealing at  $64^{\circ}\text{C}$ . VEGCL and VEGCR were used for *CsEXPB2* for 35 cycles and annealing at  $62^{\circ}\text{C}$ . POLCR and VEGCR are designed over an intron boundary. The primer pairs did not produce amplification products when the template was genomic DNA or RNA samples that had not been reverse transcribed.

### Sequence analyses

Sequences of EXPB proteins from *Physcomitrella patens*, *Selaginella moellendorffii*, *Arabidopsis thaliana*, *Vitis vinifera*, *Solanum lycopersicum*, *Brachypodium distachyon*, *Sorghum bicolor* and *Oryza sativa* were obtained through BLAST searches from Phytozome (<http://www.phytozome.net>). A similar procedure was followed to obtain protein sequences from *Musa acuminata* (<http://banana-genome.cirad.fr>), *Phoenix dactylifera* (<http://qatar-weill.cornell.edu/research/datepalmGenome/download.html>) and *Elaeis guineensis* (<http://genomsawit.mpob.gov.my/genomsawit/index.php>). Sequences were aligned with CLUSTALW, as implemented in MEGA 5.2 (Tamura *et al.*, 2011).

Neighbor-joining and maximum-likelihood phylogenetic trees were constructed with MEGA 5.2. After complete deletion of alignment gaps, distances were estimated with the Jones, Taylor and Thornton (JTT) model using a gamma parameter of 1.24. According to a maximum-likelihood model test the JTT+G model had the highest score among those available for both phylogenetic reconstruction methods. Maximum-likelihood analysis was carried out using the subtree-pruning-regrafting level-3 method and a strong branch-swap filter. Bootstrap values were based on 200 replicates for both trees. Specificity-determining sites were identified using SPEER-SERVER (<http://www.hpppi.icb.res.in/ss/>). WEBLOGO 2.8.2 (<http://weblogo.berkeley.edu/>) was used to obtain sequence logos (Crooks *et al.*, 2004). The images of ZmEXPB1 structure were created with UCSF CHIMERA (Pettersen *et al.*, 2004).

### ACKNOWLEDGEMENTS

This work was supported by US Department of Energy Grant DE-FG02-84ER13179 from the Office of Basic Energy Sciences. We thank Tony Omeis and Dr Karl Keener (Penn State University) for assistance in the collection of material, Dr Alan Meerow (USDA, Miami) for the palm pollen and Dr Alessandro Sidoli (San Raffaele Scientific Institute, Milano, Italy) for the antibodies. Daniel Durachko and Ed Wagner provided technical assistance.

### GENBANK ACCESSION NUMBERS

KM396909 and KM396910.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Alignment of protein sequences used in the phylogenetic tree.

**Figure S2.** Phylogenetic tree of EXPB proteins obtained with neighbor joining (same symbols as in Figure 4a).

**Table S1.** Primers used in this work.

**Table S2.** Analysis of partial EXPB sequences from banana, date palm and oil palm genomes.

## REFERENCES

- Ball, T., Linhart, B., Sonneck, K. *et al.* (2009) Reducing allergenicity by altering allergen fold: a mosaic protein of Phl p 1 for allergy vaccination. *Allergy*, **64**, 569–580.
- Carey, R., Hepler, N. and Cosgrove, D.J. (2013) *Selaginella moellendorffii* has a reduced and highly conserved expansin superfamily with genes more closely related to angiosperms than to bryophytes. *BMC Plant Biol.* **13**, 4.
- Carnachan, S.M. and Harris, P.J. (2000) Polysaccharide compositions of primary cell walls of the palms *Phoenix canariensis* and *Rhopalostylis sapida*. *Plant Physiol. Biochem.* **38**, 699–708.
- Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 445–476.
- Carpita, N.C. and Gibeau, D.M. (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* **3**, 1–30.
- Carpita, N.C., Defernez, M., Findlay, K., Wells, B., Shoue, D.A., Catchpole, G., Wilson, R.H. and McCann, M.C. (2001) Cell wall architecture of the elongating maize coleoptile. *Plant Physiol.* **127**, 551–565.
- Chakraborty, A., Mandloi, S., Lanczycki, C.J., Panchenko, A.R. and Chakrabarti, S. (2012) SPEER-SERVER: a web server for prediction of protein specificity determining sites. *Nucleic Acids Res.* **40**, W242–W248.
- Cho, H.T. and Kende, H. (1997) Expansins in deepwater rice internodes. *Plant Physiol.* **113**, 1137–1143.
- Cosgrove, D.J. (1989) Characterization of long-term extension of isolated cell walls from growing cucumber hypocotyls. *Planta*, **177**, 121–130.
- Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* **6**, 850–861.
- Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
- Flicker, S., Steinberger, P., Ball, T., Krauth, M.T., Verdino, P., Valent, P., Almo, S. and Valenta, R. (2006) Spatial clustering of the IgE epitopes on the major timothy grass pollen allergen Phl p 1: Importance for allergenic activity. *J. Allergy Clin. Immunol.* **117**, 1336–1343.
- Freeling, M. (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**, 433–453.
- Givnish, T.J., Ames, M., McNeal, J.R. *et al.* (2010) Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales. *Ann. Mo. Bot. Gard.* **97**, 584–616.
- Harris, P. and Trethewey, J. (2010) The distribution of ester-linked ferulic acid in the cell walls of angiosperms. *Phytochem. Rev.* **9**, 19–33.
- Harris, P.J., Kelderman, M.R., Kendon, M.F. and McKenzie, R.J. (1997) Monosaccharide compositions of unlignified cell walls of monocotyledons in relation to the occurrence of wall-bound ferulic acid. *Biochem. Syst. Ecol.* **25**, 167–179.
- Hont, A., Denoeud, F., Aury, J.M. *et al.* (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, **488**, 213–217.
- Hsieh, Y.S.Y. and Harris, P.J. (2009) Xyloglucans of monocotyledons have diverse structures. *Mol. Plant*, **2**, 943–965.
- Jarvis, M.C., Forsyth, W. and Duncan, H.J. (1988) A survey of the pectic content of nonlignified monocot cell walls. *Plant Physiol.* **88**, 309–314.
- Kozlova, L.V., Ageeva, M.V., Ibragimova, N.N. and Gorshkova, T.A. (2014) Arrangement of mixed-linkage glucan and glucuronarabinoxylan in the cell walls of growing maize roots. *Ann. Bot.* **114**, 1135–1145.
- Lee, Y. and Kende, H. (2001) Expression of beta-expansins is correlated with internodal elongation in deepwater rice. *Plant Physiol.* **127**, 645–654.
- Lee, Y. and Choi, D. (2005) Biochemical properties and localization of the beta-expansin OsEXPB3 in rice (*Oryza sativa* L.). *Mol. Cells*, **20**, 119–126.
- Li, L.C. and Cosgrove, D.J. (2001) Grass group I pollen allergens (beta-expansins) lack proteinase activity and do not cause wall loosening via proteolysis. *Eur. J. Biochem.* **268**, 4217–4226.
- Li, L.-C., Bedinger, P.A., Volk, C., Jones, A.D. and Cosgrove, D.J. (2003) Purification and characterization of four beta-expansins (*Zea m* 1 isoforms) from maize pollen. *Plant Physiol.* **132**, 2073–2085.
- McQueen-Mason, S.J. and Cosgrove, D.J. (1995) Expansin mode of action on cell walls. Analysis of wall hydrolysis, stress relaxation, and binding. *Plant Physiol.* **107**, 87–100.
- McQueen-Mason, S., Durachko, D.M. and Cosgrove, D.J. (1992) Two endogenous proteins that induce cell wall expansion in plants. *Plant Cell*, **4**, 1425–1433.
- Park, Y.B. and Cosgrove, D.J. (2012a) Changes in cell wall biomechanical properties in the xyloglucan-deficient xxt1/xtt2 mutant of Arabidopsis. *Plant Physiol.* **158**, 465–475.
- Park, Y.B. and Cosgrove, D.J. (2012b) A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. *Plant Physiol.* **158**, 1933–1943.
- Petersen, A., Dresselhaus, T., Grobe, K. and Becker, W.M. (2006) Proteome analysis of maize pollen for allergy-relevant components. *Proteomics*, **6**, 6317–6325.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
- Sampedro, J. and Cosgrove, D.J. (2005) The expansin superfamily. *Genome Biol.* **6**, 242.
- Sampedro, J., Lee, Y., Carey, R.E., dePamphilis, C. and Cosgrove, D.J. (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. *Plant J.* **44**, 409–419.
- Sampedro, J., Carey, R.E. and Cosgrove, D.J. (2006) Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs. *J. Plant. Res.* **119**, 11–21.
- Scheller, H.V. and Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289.
- Singh, R., Ong-Abdullah, M., Low, E.T.L. *et al.* (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, **500**, 335–339.
- Smith, B.G. and Harris, P.J. (1995) Polysaccharide composition of unlignified cell walls of pineapple [*Ananas comosus* (L.) Merr.] fruit. *Plant Physiol.* **107**, 1399–1409.
- Smith, B.G. and Harris, P.J. (1999) The polysaccharide composition of Poales cell walls: Poaceae cell walls are not unique. *Biochem. Syst. Ecol.* **27**, 33–53.
- Suzuki, K., Kitamura, S., Kato, Y. and Itoh, T. (2000) Highly substituted glucuronarabinoxylans (hGAXs) and low-branched xylans show a distinct localization pattern in the tissues of *Zea mays* L. *Plant Cell Physiol.* **41**, 948–959.
- Tabuchi, A., Li, L.C. and Cosgrove, D.J. (2011) Matrix solubilization and cell wall weakening by  $\beta$ -expansin (group-1 allergen) from maize pollen. *Plant J.* **68**, 546–559.
- Tamborini, E., Faccini, S., Lidholm, J., Svensson, M., Brandazza, A., Longhi, R., Groenlund, H., Sidoli, A. and Arosio, P. (1997) Biochemical and immunological characterization of recombinant allergen Lol p1. *Eur. J. Biochem.* **249**, 886–894.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tang, H., Bowers, J.E., Wang, X. and Paterson, A.H. (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl Acad. Sci. USA*, **107**, 472–477.
- Trethewey, J.A.K., Campbell, L.M. and Harris, P.J. (2005) (1 $\rightarrow$ 3), (1 $\rightarrow$ 4)- $\beta$ -D-Glucans in the cell walls of the Poales (*sensu lato*): an immunogold labeling study using a monoclonal antibody. *Am. J. Bot.* **92**, 1660–1674.
- Valdivia, E.R., Sampedro, J., Lamb, J.C., Chopra, S. and Cosgrove, D.J. (2007a) Recent proliferation and translocation of pollen group 1 allergen genes in the maize genome. *Plant Physiol.* **143**, 1269–1281.
- Valdivia, E.R., Wu, Y., Li, L.-C., Cosgrove, D.J. and Stephenson, A.G. (2007b) A group-1 grass pollen allergen influences the outcome of pollen competition in maize. *PLoS ONE*, **2**, e154.
- Valdivia, E.R., Stephenson, A., Durachko, D. and Cosgrove, D. (2009) Class B  $\beta$ -expansins are needed for pollen separation and stigma penetration. *Sex. Plant Reprod.* **22**, 141–152.
- Vogel, J. (2008) Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **11**, 301–307.

- Wang, T., Park, Y.B., Caporini, M.A., Rosay, M., Zhong, L., Cosgrove, D.J. and Hong, M. (2013) Sensitivity-enhanced solid-state NMR detection of expansin's target in plant cell walls. *Proc. Natl Acad. Sci. USA*, **110**, 16444–16449.
- Whitney, S.E., Gidley, M.J. and McQueen-Mason, S.J. (2000) Probing expansin action using cellulose/hemicellulose composites. *Plant J.* **22**, 327–334.
- Yennawar, N.H., Li, L.C., Dudzinski, D.M., Tabuchi, A. and Cosgrove, D.J. (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.
- Yokoyama, R., Rose, J.K.C. and Nishitani, K. (2004) A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. *Plant Physiol.* **134**, 1088–1099.

**Figure S1.** Alignment of protein sequences used in the phylogenetic tree. Two letter abbreviations are used to indicate the following species: Pp (*Physcomitrella patens*), Sm (*Selaginella moellendorffii*), At (*Arabidopsis thaliana*), Pt (*Populus trichocarpa*), Vv (*Vitis vinifera*), Sl (*Solanum lycopersicum*), Bd (*Brachypodium distachyon*), Sb (*Sorghum bicolor*), Cs (*Carex sp.*), Os (*Oryza sativa*), Ma (*Musa acuminata*), Pd (*Phoenix dactylifera*) and Eg (*Elaeis guineensis*).

```

AtEXPL1      LHRSKAAYFSS----ASALS-----SGACAYG-SMATSFAGHIAAAIPSIYKDGAGCGACFQVRCKN---PKLCSTKGT-IVMITDLNKS-----NQTDLVLS
OsEXPL2      VRRSKAGFRDS----SIALN-----AGSCGYG-SLAASFNGGHIAAASPALFRGGVGCACFQVRCKD---GKLCSTAGA-KVVVTDEARST-----NRTDLVL
EXPB-II
Pp_022G002600 WLGGHATWYGD--PYGEGSS-----GGACGYT-QLTGTPIGSKIAAGNAPIFQEGKGCQCCEYVKNYPS-----CSPQGT-RIVITDLCPGGQ-----YCSTDQPAFDFSG
Pp_022G064000 WLPGHATWYGD--PYGEGSS-----GGACGYT-ELAGTPYGLSVGAGSAVIYQNGQCGCEYVKNCTYPS-----CKPTPS-RIVITDFCPGGT-----FCSTGEPAFDFSG
Pp_017G043100 WVPGHATWYGD--PYGEGSS-----GGACGYL-QLTGTPIGNKIAAGNEAIFQKGGKGCQCCEYVKNYPS-----CRPEGT-RIVITDLCPGGQ-----FCSGGNPAFDFSG
Pp_017G043200 WMAGHATWYGE--PLGEGSS-----GGACGYT-KLADTPYGPKIAAGNDPIFQGGSGCGACFEVKNYPS-----CKSEPT-RIIITDQCPGGT-----YCSTSQPAFDFSG
Pp_001G152100 WVRARGTWYGD--PYGEGSS-----GGNCGYT-KLWGTPIGPKIVAGSRSIYANGQCGQCQYQIRCVDPNGGPRLCNPQGT-NVVVTDFCPGGT-----YCSTGENAFDMSG
Sm_36179      WQTGRGTWYGP--PLGTGTN-----TGACGYG-ELEGTPTYDSNIVAIGSETFNNGLGCGACFEVRCVN---DTICREEPTTVVVVTDECPECP-----AD----QLDFSG
Sm_171427     WEPATATWYGS--PNGAGTD-----GGACGYGSLPNT-PYGSVDGAGSPILFMNGIGCGTCFEVKCVD---GQLCSPQPV-NVVITDECPCGG-----YCSGGRNTHFDLSG
Sm_234188     WQLATATWYGS--PSGAGTD-----GGACGYGELPNT-PYGLDVGAGSPVLFKNRGGCGACYKVRCLQ---QQLCSGKAV-TVVITDECPCGG-----YCAFGRTHFDLSG
AtEXPB1      WLPATATWYGS--AEGDGSS-----GGACGYGSLVDVKPFKARVGAVSPILFKGGEGCGACYKVRCLD---KTICSKRAV-TIIATDQSPSGP-----SAKAKHTHFDLSG
AtEXPB3      WLPATATWYGS--AEGDGSS-----GGACGYGSLVDVKPLHARVGAVNPILFKNGEGCGACYKVRCLD---KSICSRRAV-TVIIITDECPCG-----CSKTSTHFDLSG
PtEXPB2      WKPATATWYGS--PDGDGSD-----GGACGYGSLVDVKPFRARVGAVSPVLFKNKEGCGACYKVRCLD---KSICSRRAV-TIIITDECPCGG-----YCSNGNTHFDLSG
PtEXPB3      WKPATATWYGS--PDGDGSD-----GGACGYGSLVDVKPFRARVGAVSPVLFKNKEGCGACYKVRCLD---KSICSRRAV-TIIITDECPCGG-----YCSNGNTHFDLSG
Sl_07g054170 WYPATATWYGS--PEGDGST-----GGACGYGSMVDVKPFRARVGAVSPVLFKNKEGCGACYKVRCLD---RSVCSRRAA-TVIIITDECPCGG-----YCSGGRVHFDLSG
Sl_10g008440 WHSASATWYGS--PNGDGSD-----GGACGYGTMVDVKPFRARVGAVSSALFKKGGEGCGACYKVRCLD---KTICSKRAV-TIIITDESP-----ALTKGLVHFDLSG
Vv_01017505001 WASAGATWYGS--PEGAGSD-----GGSCGYGNAVSQPPFSSMITGIGPSLYKSGKECGACYQVKCTKRM---HPSCSGRPV-RVVITDFCPGGP-----CASQ---SAHFDLSG
Ma_randomP06310 WHPATATWYGS--AEGDGSD-----GGACGYGTLVDVKPLRARVGAVSPVLFKGGEGCGACYKVRCLD---PAICSRRAV-TVIITDECPCGG-----YCSGGRTHFDLSG
Ma_5P12080    WHPATATWYGS--PDGDGSD-----GGACGYGSLVDVRPLRARVGAVSPVLFKGGQCGACYKVRCLD---PAICSRRAV-TVIITDECPCGG-----YCSGGRTHFDLSG
Ma_1P00120    WHPATATWYGS--PNGDGSD-----GGACGYGSLVDVRPLRARVGAVSPVLFKGGEGCGACYKVRCLD---PVVCARRPV-TVIITDECPCGG-----YCAFGRTHFDLSG
Eg_4P60075    WQPATATWYGS--PDGDGSDVGGCVVPGGACGYGSLVDVRPMKARVGAVSPVLFKGGEGCGACYKVRCLD---GNICSRRAV-TVIITDECPCGG-----YCAFGRTHFDLSG
Bd3_g49850    WHQATATWYGS--ADGDGSD-----GGACGYGTLVDVAPMKARVGAVSPVLFKGGEGCGACYKVRCLD---AGICSRRAV-TVIITDECPCGG-----YCSGGRTHFDLSG
Bd_5g16497    WRPATATWYGD--AEGDGST-----GGACGYGTLVDVVPKMARVGVSVPVLFKGGEGCGACYKVRCLD---HGICSRRAV-TVIITDECPCGG-----LCGGGHTHFDLSG
Sb_004G227900 WHPATATWYGS--AEGDGSD-----GGACGYGTLVDVVPKMARVGAVSPVLFKSGEGCGACYKVRCLD---HGICSRRAV-TVIITDECPCGG-----VCSGGRTHFDLSG
Sb_006G156900 WRPATATWYGE--AEGDGSD-----GGACGYGTLVDVVPKMARVGVSVPVLFKDGECCGACYKVRCLD---RGICSRRAV-TVIITDECPCGG-----LCAFGRTHFDLSG
OsEXPB16      WHPATATWYGS--ADGDGSD-----GGACGYGTLVDVVPKMKTRVGAVSPVLFKGGEGCGACYKVRCLD---ASICSRRAV-TVIITDECPCGG-----YCAFGRTHFDLSG
OsEXPB17      WRPATATWYGD--AEGDGST-----GGACGYGSLVDVVPKMARVGVSVPVLFKDGECCGACYKVRCLD---HGICSRRAV-TVIITDECPCGG-----LCAFGRTHFDLSG
Conservative EXPB-I
AtEXPB2      WSIAGSTWYGN--PTGYGSD-----GGACGYGNAVAQPPFSKMSVAGGPSLFKSGKGCACYQVKCTS---KSACSKNPV-TVVITDECPCG-----CVKE---SVHFDLSG
AtEXPB4      SGTAGVTWYGE--PFGAGST-----GGACGYGSAVANPPLYAMVSAGGPSLFNNGKGCCTCYQVVCIG---HPACSGSPI-TVTITDECPCGP-----CASEP--VHIDLSG
AtEXPB5      WNTAGITWYGD--REGPGTT-----GGACGYGDAVAKHPYRCMVSAGGPSLFKDGKGCACYRLKCD---HPLCTKKPI-KVMISDECPCG-----CTKES--VHFDLSG
PtEXPB1      WSPAGATWYGS--PTGAGSD-----GGACGYGNAVEQAPFSSFI SAGGDSIYKSGQCGACYQVKCTS---NAACSGNPV-TVVITDQCPGSP-----CAQE---SFHFDLSG
Sl_10g018500 FSPAVATWYGD--SAGAGSD-----GGACGYGNDVKNPFSAMVSAGNIFKGGKGCACYQVMCK---E---KSACSEIPI-TVTITDECPCGS-----CGNYGAPFFHFDLSG
Sl_03g093390 WGTAGATWYGS--PHGAGSD-----GGSCGYGSASVEAPLSSFVTGIGPSLYKSGKECGACYQVKCTKKM---HRSCSGKGV-RVVITDFCPGGP-----CVAQ---SAHFDLSG
Sl_07g049540 FSLAVATFY----TDAGS-----GGACGLENDVANSPPYNSMITAGNQALFKQGVGCGACYQVFCNEAQ---NPHCSGNPI-TVTLTDECPCGS-----CNDD---PVHFDLSG
Sl_05g052330 FSLAVATWYGD--PTGAGS-----GGACGLEDDVGKIPYNAMITAGNQVLFKHGLGCGACYQVLCN---Q---NEECSQNPI-TVTLTDECPCGT-----CNDD---PIHFDLSG
Sl_01g090810 WSPAGATWYGS--PSGAGSD-----GGACGYTNAVDQPPFSSSLIAAGPSLFKSGKGCACYQVKCTG---NKACSGKPV-RVVITDSCPGGP-----CLSE---SAHFDLSG

```

Sl\_01g090820 FQPAVATFYGP--PTGAGSE-----GGACGYTDAVSKPPFNSFVSAGNGLFKKGLGCGACYQVLCK----LAPCTGRPV-TVTITDECPG-----CSG---SIHFDLSG  
 Vv\_01018163001 WSPAGATWYGP--ANGAGSD-----GGACGYEDAVEKAPFSKAVSAGGPSLFKSGKGCACGYQVKCTG---NEACSGSPV-TVTITDECPG-----CVSE---STHFDLSG  
 Vv\_01018168001 WSPAGATWYGP--AAGAGSD-----GGACGYGDAVEKAPFSKAVSAGGPSLFKSGKGCACGYQVKCTG---NEACSGSPV-TVTITDECPG-----CVSE---STHFDLSG  
 Vv\_01030375001 WHSATATWYGS--PDGDGSD-----GGACGYGSLVDVKKPFRARVGVAVSPVLFGGGVCGCACYQVKCLD---HNICSRRAV-TIIVTDECPGG-----YCSGGRTHFDLSG  
 Ma\_7P07040 WSPAGATWYGS--AHGAGSD-----GGACGYGNGVDRAFPSSMIAAGSPSIFKSGGEGCACYQIRCTK---NAACSGNPA-TIVLTDECPGG-----ICLAE--PVHFDLSG  
 Ma\_randomP23770 NSSAVATWYGS--PGGAGSD-----GGACGYGDAVSKPPFSSFIAAGGPSLFKSGMGCACGYQVACTS---NGACSGNPV-TVVITDECPGG-----PCASD--PVHFDLSG  
 Ma\_2P13740 MSPAVATWYGE--AEGPGST-----GGACGYGDGVAKAPFGSKIAAGGTSLYKSGKGCACGYQVACTA---NPACSGRPV-TVVITDQCPGG-----PCASD--AVHFDLSG  
 Ma\_2P13750 MSPAVATWYGA--AEGPGST-----GGACGYGDGVAKAPFESKISAGNSALFKSGKGCACGYQVACTA---NPACSGRPV-TVVITDQCPG-----ACDND--PLHFDLSG  
 Ma\_2P13730 MSPAVATWYGA--AEGPGST-----GGACGYGDGVAKAPFDSKISAGGTSLYKSGKGCACGYQVACTA---NPACSGSPV-TVVITDQCPGG-----PCASD--SVHFDLSG  
 Ma\_4P26910 WLPAGATWYGS--ANGAGSD-----GGACGYGGALEKPPFSSMISAGGPSLFKSGKGCACGYQVKCTE---NAACSGDPV-TVVITDECPGG-----PCLEK--SAHFDMG  
 Eg\_4P156078 WLPAGATWYGS--PRGAGSE-----GGACGYGNAVDQAPYSAMISAGSTIFNKGKGCACGYQVMCTT---HKACSGKPV-TVVIADSCPGTV-----CQRE--SAHFDGSG  
 Eg\_4P79642 LHLTAITLYQNDKVHAVKLT-----GGACGYRNAVDQAPFSAMIAAGSPSLFKSGKGCACGYQVMCTT---HEACSRKPV-TVVITDECPGGV-----CLKE--SAHFDLGSG  
 Pd\_30s825291g003 WSPAGATWYGS--AHGAGSD-----GGACGYGDAVGQAPLSAMIAAGSPSLFESGKGCACGYQVRCTT---HKACSGKPV-TVVITDSCPGGT-----CLRE--SAHFDLGSG  
 OsEXPB5 WSSGGATWYGG--PQGDGSE-----GGACGYQSAVGRQPFSSMIAAGGPSLFKNGKGCACGYQIKCTG---NRACSGRPV-TVVITDSCPGG-----VCLNE--AAHFDMGSG  
 OsEXPB11 WAAAGATYYGA--PNGDGSD-----GGACGYQTAVGQRPFFSSMIAAGSPSLYKGGKGCACGYEVKCTT---NAACSGQPA-TVVITDECPGG-----ICLAG--AAHFDMGSG  
 OsEXPB12 WSNGIATWYGD--PNGAGSE-----GGACGYQYAVDQPFSSMIAAGSPSIYQDGKGCACGYQVVCAG---NEACSGNPV-TVVITDQCPGGP-----CLLEE--LVHFDLSG  
 OsEXPB14 WSIGGATWYGP--ANGSGTD-----GGACGYQGDVGGQPPFSSMIAAGSPSIYESGKGCACGYQVKCSG---NPSCSGKPV-TVVLTDLCPGGA-----CLEE--PVHFDLSG  
 OsEXPB15 WSNGGATWYGA--ANGAGSD-----GGACGYQGAVFQAPFSSMIAAGSPSIYKSGLGCGACGYQVKCTG---NSACSGNPV-TVVLTDECPGGP-----CLSE--PVHFDLSG  
 OsEXPB18 WSNGGATWYGA--ANGAGSD-----GGACGYQAAVDQAPFSSMIAAGSPSIYKSGLGCGACGYQVKCSG---NSACSGNPV-TVVLTDECPGGP-----CLSE--PVHFDLSG  
 Bd\_3g50730 WSNGGATWYGA--SSGAGSD-----GGACGYQDAVGQPPFSSMIAAGGPSLFKNGKGCACGYQVRCKE---DPACSGKHV-TVVITDSCPDG-----CLEE--PVHFDLSG  
 Bd\_4g41110 WADARATWYGP--PGGAGTD-----GGACGYQGDVEQPPFSSAMITAGGPSIFQSGQGCACGYQVKCTG---NAACSGRPV-TVVVTDECPDGT-----CLSE--RFHFDLGG  
 Bd\_4g41117 WSDAGATWYGP--ANGAGTD-----GGACGYQRDVELPPFSSAMYTAGGPSIYQDGKGCACGYQVKCMD---NPACSGNPV-TVVVTDQCPGGP-----CAAD--RVHFDLSG  
 Bd\_5g17760 WSSAGATWYGA--ANGAGSD-----GGACGYKGAVDQAPFSSMIAAGGPSIYSGSLGCGACGYQVKCSG---NDACSGSVV-TVVITDECAGGP-----CLNE--P-HFDLSG  
 Bd\_5g17780 WADGGATWYGG--PYGDGSE-----GGACGYKSDVGQPPFSSMIAAGGPSLFKNGKGCACGYQVRCKE---DPACSGKHV-TVVITDSCPDG-----TCQKE--KAHFDMGSG  
 Bd\_5g17770 WSSGGATWYGS--PYGAGSD-----GGACGYQGAVGQRPFFSSMIAAGGPSFFKNGKGCACGYQVKCTG---NRACSGRPV-KVVITDSCPGG-----VCARE--A-HFDMGSG  
 Bd\_3g50740 WLAAGATYYGT--PNGDGSD-----GGACGYQSAVGHRPYSSMIAAGSTPLFLGGKGCACGYQVKCTA---NKACSGNPA-TVVITDLSPGG-----LYPGE--PAHFDMGSG  
 Bd\_3g50750 WLYGGATYYGA--PNGDGSD-----GGACGYQSAVGNRPFFSSMIAAGNPFLFKGGKGCACGYEVKCTG---NLACSGQPA-TVVITDECPGGA-----ACLGE--AAHFDMGSG  
 Sb\_004G294300 WLSAGATYYGA--PNGDGSD-----GGACGYQSAVGNRPFFSSMIAAGSPSLFKGGKGCACGYEVKCTG---NAACSGQPA-TVVITDECPGG-----VCLAE--AAHFDMGSG  
 Sb\_004G294400 WLSAGATYYGA--PDGDGSD-----GGACGYQTAVGQRPFFSSMIAAGSPSLFKGGKGCACGYEVKCDG---NAACSGQPA-TVVITDECPGG-----VCLAE--AAHFDMGSG  
 Sb\_002G124400 LSGAGATYYGA--PNGDGGE-----GGACGYQTAVGKQPFDSMIAAGSAPLYRGEGCAGCYEVKCTS---NAACSGQPV-TIVITDLSPGG-----LFPGE--VAHFDMGSG  
 Sb\_002G124500 LSGAGATYYGA--TNGDGSD-----GGACGYQTAVGKKPFDMSIAAGSTPLYKGGEGCAGCYEVKCKT---NAACSGQPV-TIVITDQSPGD-----LFPGE--VEHFDMGSG  
 Sb\_006G171100 WYSGGATWYGS--PYGAGSD-----GGACGYQGTVSQRPFFSSMIAAGGPSLFKNGNGCAGCYQIKCAG---NKACSVRPV-TVTITDSCPGG-----VCLAR--TAHFDMGSG  
 Sb\_006G171200 WSSGGATWYGA--ATGAGSD-----GGACGYRGAVDQAPFSSMIAAGSPSIYNSGKGCACGYQVKCTG---NGACSGNPV-TVVITDECPGGA-----CLNE--PGHFDMGSG  
 Sb\_006G171300 WSSGGATWYGN--ANGAGSD-----GGACGYKGDVSRPFRSMIAAGGPSLFKNGKGCACGYQIKCTG---NRACSGRPV-TVTITDSCPGG-----ACLAE--SAHFDMGSG  
 Sb\_001G155600 WSNGIATWYGA--ADGAGTD-----GGACGYQNAVDQPPFSSMIAAGCSSIYDSGKGCACGYQVVCTG---NEACSGNPV-TVVITDECPDCP-----CPDD--QVHLDMSG  
 Sb\_001G155700 WTTGIATWYGD--ANGAGSE-----GGACGYQYAVDQPPFSSMIAAGSPFIYDSGNGCAGCYQVVCSS---NQACSGYPV-TVVITDQGGGGPCLSQASDGMCLNE--GAHFDMGSG  
 Sb\_004G294500 SGDAGATWYGA--PTGAGSD-----GGACGYQNAVDQPPFSSMIAAGSPSIFQNGEGCAGCYQVKCTG---HACSGSPV-TVVLTDECPGGA-----CLAE--LVHFDLSG  
 Sb\_010G120100 WSDAGATWYGP--ANGAGSD-----GGACGYQGAHVQPPFSSMITAGSNSIYQDGKGCCTCYQVKCTG---HQSCSGSPV-TVVLTQCPG-A-----CQSE--PVHFDLSG  
 Divergent EXPB-I  
 CsEXPB1 WLSGSSTFYGS--PTASGPPDN----GGKCGFK-NINLPPFNLSLTCGSKALWKEGKGCACGYLMKCTS---HPLCSGQPK-LITITDECLYGGC-----ADAT---YHFDMAG  
 CsEXPB2 WNMGSATWYGA--PTGAGADDN----GGGCGFK-NVNLPFPNSMTSCGNHNLFKGGKGCACGYLVKCTG---HPACSGRPS-HITIADLCDGGIC-----LDAP---YHFDMAG  
 OsEXPB1 WLEAKATWYGA--PKGAGPKDN----GGACGYK-DVDKAPFLGMNSCGNDPIFKDGKGCACGYQVKCTG---E-ACSDKPA-LIHVTDMNDEP-----IAA---YHFDLGSG  
 OsEXPB2 WLPAKATWYGO--PNGAGPDDN----GGACGFK-NTNQYPFMSMTSCGNELFLQDGKGCACGYQIRCT--N--NPSCSGQPR-TVIITDMNYYP-----VAR---YHFDLGSG  
 OsEXPB3 WEVARATWYGA--PTGAGPDDD----GGACGFK-NTNQYPFMSMTSCGNELFKDGKGCACGYQIRCVN---HPACSGNPE-TVIITDMNYYP-----VSK---YHFDLGSG  
 OsEXPB4 WTKARATWYGO--PNGAGPDDN----GGACGFK-HTNQYPFMSMTSCGNPLFKDGKGCACGYQIRCT--K--DQSCSGRSE-TVIITDMNYYP-----VAP---YHFDLGSG  
 OsEXPB6 WLPARATWYGA--PTGAGPDDN----GGACGFK-NVNQYPFMSMTSCGNELFKDGKGCACGYQIRCNK---DPSCSGNIE-TVIITDMNYYP-----VAR---YHFDLGSG  
 OsEXPB7 WLNARATWYGA--PNGAGPDDN----GGACGFK-NVNLPFPFSAMTSCGNELFLQDGKGCACGYQIRCVG---HPACSGLPE-TVIITDMNYYP-----VSL---YHFDLGSG  
 OsEXPB8 WGNAKATWYGO--PNGAGAADN----GGACGFK-KVNQYPFMGMTSCGNPLFKGGKGCACGYVRVRCN--R--NPACSGNAQ-TVAITDMNYYF-----LSQ---YHFDLGSG  
 OsEXPB9 WLPARATWYGO--PYGSGSTDN----GGACGIK-NVNLPFGNMTSCGNVPIFKDGRGCGSCGYEVKCEQA---ACSKQPV-TVFITDMNYYE-----ISA---YHFDGSG  
 OsEXPB10 WLEARGTWYGK--PKGAGPDDN----GGACGYK-DIDKAPFLGMNSCGNDPIFKDGKGCACGYEVKCSKP---E-ACSDKPV-IIHITDMNTEP-----IAA---YHFDLGSG

Bd\_2g35007 WLHAKATWYGS--PTGAGPPDN-----GGSCGIK-ETNLPPYNSLTS CGNLPIFKDGGKGCSCFKIKCGNKV-----CSKKPV-VVFITDMNYYQ-----FAP---YHFDLSG  
Bd\_4g00360 WLDAKSTWYGR--PEGAGPKDN-----GGACGYK-DVDKEPFSGMTSCGNTPIFRDGRGCGSCFEIKCDKPA--E-FCSGQPV-LVHITDDNEEP-----IAA---YHFDLSG  
Bd\_1g78120 WLDAKSTWYGK--PTGAGPKDN-----GGACGYK-DVDKEPFSGMTSCGNTPIFRDGRGCGSCFEIKCTKP--D-ACSGDPV-LVHITDDNEEP-----IAA---YHFDLSG  
Bd\_1g78340 WMDARATWYGA--PNGAGPDDN-----GGACGFK-DVNLPFFSAMTSCGNEPLFKDGGKGCSCYQIRCLSRM--HPACSGVPE-TVIITDMNYYP-----VSR---FHFDLSG  
Bd\_1g78350 SGVARATWYGA--PNGAGPYDN-----GGACGFK-NVNKYPFMAMTSCGNQPLFKDGGKGCSCYQIRCLKCT--K--HKACSGRTE-TVVITDMNYYP-----VAP---YHFDLSG  
Bd\_3g33110 WLPAKATWYGA--PTGAGPDDN-----GGACGYK-FTNLYPFSSMTSCGNEPLFMDGAGGCTCYQIKCTSVN--NPACSGQPK-IVMITDMNYYP-----VAR---YHFDLSG  
Bd\_3g33150 WENARATWYGA--PTGAGPDDN-----GGACGFK-NVNQYPFSAMTSCGNEPLFKDGGKGCSCYQIRCN---DQSCSGSPE-TVVITDMNYYP-----VSK---YHFDLSG  
Bd\_3g33130 WLPAKATWYGA--PTGAGPANG-----GGACAYS-NNNLYPILSMNACGNQPLFMDGAGGCTCYEIKCDYKN--NRACSGQPK-RVTITDMNYYN-----IAQ---YHFDLSG  
Bd\_3g33140 WQAAKATWYGA--PTGAGPDDN-----GGACGFK-NVNQYPFSSMISCGNEPLFKDGGSGCACYEIRCTKVN--NPACSGKAK-KIMITDMNYYP-----VAK---YHFDLSG  
Bd\_3g32810 WLNAMATWYGK--PTGAGPKDN-----GGACGIK-DVNLPYPYSAMTACGNVPIFKDGRGCGSCYEVKCNVPA--K-LCSNKP-I-TVFITDMNYYE-----IAP---YHFDLSG  
Bd\_3g33160 WTPARATWYGO--PNGAGPDDN-----GGACGFK-HTNQYPFSAMTSCGNQPLFKDGGKGCSCYQIRCR--K--DQSCSGRTE-TVIITDMNYYP-----VAP---FHFDLSG  
Bd\_3g33120 WLPAKATWYGA--PTGAGPDDN-----GGACGFK-DTNQYPYSSMTSCGNEPLFMGGAGCCTCYQIRCNAN--NPACSGQPR-LVTITDMNYYP-----VAK---YHFDLSG  
Sb\_004G191600 WVAARATWYGA--PTGAGPYDN-----GGACGFK-NVNLPFFSAMTSCGNQPLFKDGGKGCSCYQIRCVN---HPACSGNPE-TVMITDMNYYP-----VAK---YHFDLSG  
Sb\_001G539700 WLEAKATWYGK--PTGAGPDDN-----GGACGYK-DVNKAPFNSMGACGNLPIFKDGLGCGSCFEIKCDKP--A-ECSGEAV-VVHITDMNYYEQ-----IAA---YHFDLAG  
Sb\_001G301000 WQMAKATWYGA--PYGAGPDDN-----GGACGFK-NTNQYPFMSMTSCGNQPLFQDGGCGCACYQIRCTAKN--NPACSGEVK-TVMITDMNYYP-----VAR---YHFDLSG  
Sb\_001G306200 WLVARATWYGO--PNGAGPDDN-----GGACGIK-NVNLPYPYNGFTACGNVPIFKDGGKGCSCYEVRCCKEMP---ECSGNPI-TVFITDMNYYE-----IAP---YHFDLSG  
Sb\_001G300500 WEAARATWYGA--PTGAGPDDN-----GGACGFK-NVNLPFFSAMTSCGNQPLFKDGGKGCSCYQIRCCQ--K--HPACSGNPE-TVIITDMNYYP-----VSK---YHFDLSG  
Sb\_001G301300 WLAAKATWYGA--PNGAGPDDN-----GGACGFK-NTNQYPYMSMTSCGNEPLFKDGGKGCSCYQIRCLNST--HDACSGRAK-RVITITDMNYYP-----VSK---YHFDLSG  
Sb\_001G301500 WLPAKATWYGA--PTGAGPDDN-----GGACGFK-GTNQYPFSSMTSCGNEPIFKDGGKGCSCYQIRCLKSN--HPACSGVPQ-TVIITDMNYYP-----VAK---YHFDLSG  
Sb\_001G306400 WLVARATWYGO--PNGAGPDDN-----GGACGIK-NVNLPYPYNGFTACGNVPIFKDGGKGCSCYEVRCCKEMP---ECSGNPI-TVFITDMNYYE-----IAP---YHFDLSG  
Sb\_001G300400 WQKARATWYGO--PNGAGPDDN-----GGACGFK-HTNQYPFMSMGSCGNQPLFKDGGKGCSCYQIRCT--K--DRSCSGRAE-TVIITDMNYYP-----VSK---YHFDLSG  
Sb\_001G306500 WLSARATWYGO--PNGAGPDDN-----GGACGIK-NVNLPYPYNGFTACGNVPIFKDGGKGCSCYEVRCCKEMP---ECSGNPI-TVFITDMNYYE-----IAP---YHFDLSG  
Sb\_001G539600 WLEAKATWYGK--PTGAGPDDN-----GGACGYK-DVNKAPFNSMGACGNLPIFKDGLGCGSCFEIKCDKP--A-ECSGEAV-VVHITDMNYYEQ-----IAA---YHFDLAG  
Sb\_001G301400 WLPKATWYGR--PNGAGPDNN-----GGCGYS-GTNLYPFNSMTSCGDRNVFRDGGKGCSCYQIKCVSSN--NPACSGVPQ-TIITDVNYDTS-----LGP---NRFDLSG  
Sb\_001G542200 WGNARATWYGO--PNGAGPYDN-----GGACGFK-KVNQYPFAMTSCGNQPLYRDGGKGCSCYQIRCSSK--HAACSGRTE-TVVITDMNYTPG-----VAP---YHFDLSG  
Sb\_001G542100 WLNARATWYGA--PNGAGPDDN-----GGACGFK-GVNLPFFSAMTSCGNEPLFKDGGKGCSCYQIRCKA---HPACSGVPE-TVIITDMNYYP-----VAP---YHFDLSG  
Sb\_001G540000 WLEAKATWYGK--PTGAGPDDN-----GGACGYK-DVNKAPFNSMGACGNLPIFKDGLGCGSCFEIKCDKP--A-ECSGEAV-VVHITDMNYYEQ-----IAA---YHFDLAG  
Sb\_001G300700 WEAARATWYGA--PTGAGPDDN-----GGACGFK-NVNLPFFSAMTSCGNEPLFKDGGKGCSCYQIRCN---HAACSGNPE-TVIITDMNYYP-----VAK---YHFDLSG  
Sb\_001G300800 WEAARATWYGA--PTGAGPDDN-----GGACGFK-NVNLPFFSAMTSCGNEPLFKDGGKGCSCYQIRCCQ--K--NDACSGNPE-TVIITDMNYYP-----VSK---YHFDLSG  
Sb\_001G300900 WEAARATWYGA--PTGAGPDDN-----GGACGFK-NVNLPFFSAMTSCGNEPLFKDGGKGCSCYQIRCCQ--K--HPACSGNPE-TVIITDMNYYP-----VAK---YHFDLSG

AtEXPL1 RAFRAMAKPIV-GADKDLLKQGIVDIEYQVRPCDYGN-KNMN---VRVEEASKKPNYLEIKLLYQGGQTEVVSIDIAQVGSPP-----NWGYMTRSHGAVVTDKVV--PTGAIQF  
OsEXPL2 AAYAAMARP---GMAAQLRTRRAVDVEYKRVPC EYAAGRNLNLS---IRVEEKSRPPREL SIRFLYQGGQTDIVAVDVATVGSS-----NWKFMTDRDYGPAWSTAQA--PAGPLQF  
EXPB-II  
Pp\_022G002600 AAITAMALPG---RDGELRNIGLYDIQYKRVPC EYP-NQNIA---FKVDAGSS-KYWFSFTVKYLGPGDINTVEVKCGKNY-----FQYAQHSWGANWMLINYS GPVPFQFPPL  
Pp\_022G064000 MAMTNMALPG---RDQELRNIGLYEIQYRRVPC YYP-NQNVA---FKVDPGST-PFWLSFTIEYQGGPGDIESVAIRQAGSSE-----FQQMQHNWGANWMLICYSGKPFKGPY  
Pp\_017G043100 AAISAMAKDG---QD GALRNIGLYDIQYKRVPC EYP-GQNIV---FKVDAGSS-PFWLSFTVKYMGPGDIESVSIQ-RDGS-----FIPAQHSWGANWMLINYS GAPFQGPY  
Pp\_017G043200 AAISDMAVSG---KD GELRNIGLYDILYKRVPC EYP-NQNIA---FQVDAGSS-AFWLSLVKYMGPGDIESVEIRTTGSSS-----FQPAKHNVGASWMLINTSGQPFKGPY  
Pp\_001G152100 AAINAMALRG---REGQLRNRGLYNLLYKRVPC RYR-GTNIE---FRVDNGSS-PFWLSILIKYVGGPGDIGQVYIRMANWYK-----FQPMRHAWGANWMI PN YDGKPFGRGPM  
Sm\_36179 TAFESLAIEG---QGDALRRARGIISIEYRRKSCDFA--ANIT---FEVVPGSN-EFWIAFVVKYVPGYALQSV EYVQSSGNT-----WENAAHLWGANWQIDG---PLTPT  
Sm\_171427 TAFGKMASGSAN---IQHLLAAGVLNVLYRRAPCIYK-SQG VV---FQVADGST-PFWFETVIRYLDGPGDLATVELQQFGSSA-----WQPMQSVWGANWCLNAGGGTPLRAPF  
Sm\_234188 TAFGRMAMAGR---TNQLLGS GVTQVLYKRVDCNYG-SRPM E---FQVNEGST-PFWLSLVRYAAGPGDLGHV ELMQAGSRV-----WQPMQVWGANWCFNGG---PLRGPF  
AtEXPB1 AAFGHMAIPGH---NGVIRNRGLLNILYRR TACKYR-GKNIA---FHVNAGST-DYWLSLLIEYEDGEGDIGSMHIRQAGSKE-----WISMKHIWGANWCIVEG---PLKGPF  
AtEXPB3 AVFGRLAIAIEG---SGPLRNRGLIPVIYRR TACKYR-GKNIA---FHVNEGST-EFWLSLVVEFEDGEGDIGSMHIRQAGRE-----WLEMKHVWGANWCIIGG---PLKGP  
PtEXPB2 AAFGRMAISGE---NGQLRNRGEIPVIYRR TPCKYP-GKNIA---FHVNEGST-DYWLSLVVEFEDGEGDGVGSMHIREAGGTE-----WLEMNHVWGANWCIIIRG---PLKGPF  
PtEXPB3 AAFGHMAISGE---NGQLRNRGEIPVIYRR TPCKYP-GKNIA---FHVNEGST-DYWLSLVVEFEDGEGDGVGSMHIREAGGTE-----WLEMHLWGANWCVIRG---PLKGPL  
Sl\_07g054170 AAFGHMAVSGY---GGSLDRRGVISILYRR TLPCKYP-GKNIA---FHVNEGST-PYWLSLVVEFEDGEGDGVGSMHIREARSNQ-----WIEMSHLWGANWAIIGG---PLQGP  
Sl\_10g008440 SAFGRMAVSGR---NTNLRNRGKISIIYRKTPCKYP-GQNIA---FHVNEGST-AYWLSLVVEFEDGEGDIGSMQIK EARSNQ-----WLTMTHLWGANWCIIGG---PLQGP  
Vv\_01017505001 TAFGAMAIPG---QEEKLRNVGVLEIRYARVACDYS-GKTIA---FHVDLGAN-PNSFSVLVEFEGDGLAGVALKET-----LKSGSKWRAMQQSWGAVWKLDAG---YELKPPF

|                     |                                                                                                                            |
|---------------------|----------------------------------------------------------------------------------------------------------------------------|
| Ma_randomP06310     | AAFGRMAVAGE---AGQLRDRGEISVIFCRTPCKYP-GKNIA---FHVNEGST-NFWLSLLVEFEGDDGDIGSMHIKQANSMK-----WLEMKHVWGASWCIIGG---PLQGPF         |
| Ma_5P12080          | AAFGRMAVAGE---AGQIRDRGEISVFRRTPCCKYP-GKNIA---FHVNEGST-NYWLISLLVEFEDDDGDIGAMHIKQANSVE-----WLEMKHIWGGNWCNFNGG---PLQGPF       |
| Ma_1P00120          | AAFGRMAVPGK---ASRLRDRGEMPVFRRTPCCKYP-GKSIA---FHVNEGST-NFWLSLLVEFEDDDGDIGSMHIKQANSVE-----WLEMKHIWGANWNCIIGG---PLHGPF        |
| Eg_4P60075          | AAFSRMAVAGS---ASQLRDRGEISVLFRRTPCKYR-GKNIA---FHVNEGST-EYWLSLLVEFENGDDGVGSMHIKQSNSVE-----WLEMKHIWGANWNCIIGG---PLRGPF        |
| Bd3_g49850          | AAFGR LAVAGH---GGQLRN RGEISVFRRTPCCKYR-GKNIA---FRVVEGST-SFWLSLLVEFEDGDGDIGSMQLKQAKSAQ-----WQDMKHIWGATWSLTPG---PLIGPF       |
| Bd_5g16497          | AAFSRMAVAGA---GAHLRDRGQLSVVYRR TACKY G-GKNIA---FHVNEGST-NFWLSLLVEFEDGEDGDISMQLKQASSAQ-----WLDMKHVWGATWCMYGG---PLAGPF       |
| Sb_004G227900       | AAFGR LAVAGA---GGQLRN RGEIN VVFRRTACRYG-GKNIA---FHVNEGST-SFWLSLLVEFEDGDGDIGSMQLKQANSAQ-----WRDMQHVWGATYSLTPG---PLVGP       |
| Sb_006G156900       | AAYSRMAVAGA---GGRLRDRGQLNVVYRR TACKY G-GKNIA---FHVNEGST-NFWLSLLVEFEDGEDGDISMQLKQANSVE-----WLDMRHVWGATWCLVRG---PLVGP        |
| OsEXPB16            | AAFAR LAVAGH---GGQLQN RGEISVVYRR TACKY G-GKNIA---FHVNEGST-TFWLSLLVEFEDGDGDIGSMQLKQANSAQ-----WQDMKHIWGATWSLTPG---PLVGP      |
| OsEXPB17            | AAFSRMAVAGA---GGHLRDRGQLSVVYRR TACKY G-GKNIA---FRVNEGST-NFWLSLLVEFEDGQGDIGSMQIKQANSVE-----WLDMKHVWGATWCLVRG---PLVGP        |
| Conservative EXPB-I |                                                                                                                            |
| AtEXPB2             | TAFGAMAISG---QDSQLRN VGELQILYKKVECN YI-GKTVT---FQVDKGSN-ANSFAVLVAYVNGDGEIGRIELKQALD-----SDKWLSMSQSWGAVWKLDVS--SPLRAPL      |
| AtEXPB4             | KAMGALAKPG---QADQLRSAGVIRVNYKRAACLYR-GTNIV---FRMDAGAN-PYYISFVVEYENGDDLSNVEIQP-AGG-----SFISMQEMRSAVWKNVSGS-ALR-GPF          |
| AtEXPB5             | KAFGALAKRG---KGDQLRN LGELKVS YK RACCKHP-KTMIA---IHVDAGAN-PYMSFAV KFANGDGNFACIEVQP-AGG-----QYMKMEEMRSAVWRLSPGV-PLK-GPF      |
| PtEXPB1             | TAFGAMAISG---KEDQLRNAGVLQIQHQRVPCNWP-GKTVT---FHVDSGSN-PYYFATVVEYEDGDGELKSVELKQALD-----SDSWVPMQKSWGAVWKLDAG--SLLRAPL        |
| Sl_10g018500        | TAFGALAKPG---QADLLRGDGIINIGYKR RVACNYP-QTTIA---FRMDQGSN-PYFVSCVIEFENGEGDLGLVLDKQALD-----NKWPPMQHWSWGANWKIEFP-QIK-PPF       |
| Sl_03g093390        | TAFGAMAIPG---QEQLR DAGV LQIRYARVACDYS-RKNIV---FHVDQGSN-SEYFAV VIEFEQGDGLAKVELKEKRSSSIISRGKNKNYKWRQMQQSWGAVWKLDAG--SKLHPPF  |
| Sl_07g049540        | IAFAKLAKPG---EQAELHKAGRIPIYYKR RVACN YN--RNIL---FKVDKGSN-PNFFAVVSEAVDGDGDL SLVEIQT-GGKK-----NTWNSMNRMIGSTWSVGIDP-NTQKPPF   |
| Sl_05g052330        | NAFGAMAKFG---QADQLRSLGRIDIYYKR VSCD HK--QNIM---FKVEKGSN-PYFLAIAIEAQNGDGDLSLVEIKH-TNS-----NEWLQMQQMFGATWSINIYP-DTQIPPF      |
| Sl_01g090810        | TSFGAMAISG---QADQLRNAGVIQIQYKR RVECNYP-GVKLT---FHVDSGSN-PNYFAALVEYEDGDGDLNSVDLKQALD-----SESQPMQQSWGAVWKLDG--SSLVAPF        |
| Sl_01g090820        | TAMGALAKPG---QANALRNIGNIPISYQRVPCHYQ-NWKIA---FKVDQGSN-PNFLSVNVEFQNGDGDLSLVEFLP-ARS-----NQAIRANHVFGATWSSNINP-AAQPAPY        |
| Vv_01018163001      | TAFGAMAKSG---QDDQLRNAGVLQIQYRRVECEYP-GTSVV---FHVDSGSN-SNYFASLIEYEDGDGDLARVYLQQALD-----SDMWVPMQKSWGAVWKLDYG--SELRAPF        |
| Vv_01018168001      | TAFGAMATPG---QADQLRNAGILQIQYRRVECEYP-GASVA---FHVDSGSN-PNYFASLIEYEDGDGELAGVDLQQALH-----SNSWVPMQKSWGAVWKLDG--SGLRAPF         |
| Vv_01030375001      | AAFGRMAVAGN---NNQLRDRGEISVIFRRTPCKYP-GKNIA---FHVNEGST-DYWLSLLVEFEDGDGDVGSMHIREASSTE-----WLEMRHVWGASWCITGG---PLKGP          |
| Ma_7P07040          | TAFGAMAKPG---QSDQLRNAGVLQIQYTRVECN YA-GIDVA---FHVDSGSN-PYYIAVLIEYEGGDGLACVDIKEGASSSS---STPSSSWIPMTQSWGAEWRLN--SGPKLRPPF    |
| Ma_randomP23770     | TAFGAMAKPG---QADALRN VGSLQIQYSRVHCKYS-SVNIS---FKMDAGSN-PYYFAV VIEFEDGDGDL SAVD VQAGSGF-----WIPTQQSWGAVWKLN SG--WPLQAPL     |
| Ma_2P13740          | AAFAMAKPG---QAAALRSAGSIHIQYARVPCSY P-GFHVT---FRVDNGSN-PNYFAVLPEFVNGDGEISAAEVGQGSS-----WTPMQSSYGA VWKLNA---PVHGPA           |
| Ma_2P13750          | AAFAMAKPG---EAAALRSAGFIHQYARVPCSY P-GFHIA---FRVDNGSN-PNYFAVLPEFVNGDGEISAAEVGQGSS-----WSPMQFSWGA VWKLNA---PVGPA             |
| Ma_2P13730          | AAFAMAKPG---QADALRSVGSIQIQYARVPCSY P-GFHVA---FRVDDGSN-SNYLAVLPEFVNGDGEISAVEVGQGSS-----WTPMQNSWGA LWKLNA---PVPGPA           |
| Ma_4P26910          | TAFGAMASSG---RADEIRNAGVMTVQYARVQCSYP-GFDLA---FRVDAGSN-PYYFAV VIEYEGDGLA AVELMQASSAD-----SSQWLSMQKSWGAVWKLN SA--WQLQAPF     |
| Eg_4P156078         | TAFGAMAKPG---MADELRSAGYIYIQYSRVACEYP-GRVT---FHVEHGSN-PYYFAV VVEFEDGDGDL SAV EIKG-----SGEWRAMRQSWGAVWRLDTS--SPVHGPF         |
| Eg_4P79642          | TAFGAMAKPG---MADELRNAGILPIQYSRVACEYP-GKTI---FHVDSGSN-PYYMAV VVEFEDGDGDL SAV ELKDG-----SAEWRAMAQSWGAVWRLDSG--STLHGPF        |
| Pd_30s825291g003    | TAFGAMARPG---MADELRNAGILPIQYSRVACEYP-GKTI---FHVEAGCN-PYYFAMVVEFEDGDGDLA AVEIKDG-----SNAWRAMTQSWGAVWRLDSG--SVLHAPF          |
| OsEXPB5             | TAFGAMANRG---MGDRLRSAGVLKIQYKR VP CRF--AMNVA---FKVDAGSN-PYYLAILVQYANGDGD LA AVHIMK-----ARGG--WKAMQQSWGATWRLNSN--KPLSPPF    |
| OsEXPB11            | TSMGAMAKPG---MADKLRAAGILQVQYRRVPCKYS-GVNIA---FRVDQGAN-PFYFEVLIEFEDGDGDLNAVDLMEAGC-----G-WTPMVQNWGA LWRYNSN--KALKAPF        |
| OsEXPB12            | TAFGAMARPG---QADQLRGAGLLQIQYTRVECEWT-GVGLT---FVVDSGSN-PNYLALLVEYDDNDSDLA AVDIMP I G-A-----GASGSWIPMQKSWGAVWRLN-SGSALQ-GPF  |
| OsEXPB14            | TAFGAMAKPG---QDDQLRNAGKLVPQYARVPCKWQ-GVDIA---FRVDAGSN-QYYLAVLVEDEDDGDGDL SAV DLMQSGGS-----GGGGSWAAMQQSWGAVWKYN-SGAPLQ-APM  |
| OsEXPB15            | TAFGAMANPG---QADQLRAAGVLQIQYNRVPCN WG-GVKLT---FVVDVGSN-PNYFAVLVKYENGDDGLSGVELMQTG-----AG-AAWTQMQQSWGAVWKLN-AGSALQ-APF      |
| OsEXPB18            | TAFGAMANPG---QADQLRAAGVLQIQYNRVPCN WG-GVMLT---FAVDAGSN-PSYFAVLVKYENGDDGLSGMDLMQTG-----AG-AAWTPMQQSWGAVWKLS-AGAAIQ-APL      |
| Bd_3g50730          | TAFGAMAKPG---QADQLRAAGRLQIQYTRVPCNWR-GTDIA---FKVDAGSN-SYYLAMLIEDESGDGL SAV DLKQSGA-----GAGGWAPMARSWGAVWRFN-SGPMQLQ-PPF     |
| Bd_4g41110          | RAFAMAKPG---LAEALRHAGNIKRVN RVACNWKY-GVDMA---FKVDVGSN-PYYLAVLVEDESGDGL SAV DLMQSHGG-----SGGGWAPMQRSWGA VWKYN--GGPLQ-GPM    |
| Bd_4g41117          | KAFGALAKPG---LAGSLRN VGNIKVQFN RVACNWH-GVNVA---FRVDAGSN-PNYLAVLVEDEAGDGNLSA VELQQRGG-----AGG-WAPMQRSWGA TWKYN--G-AVK-APV   |
| Bd_5g17760          | TAFGAMAKPG---QADQLRGAGVLQIQYTRVP CSWA-GAELT---FVVDAGSN-PHYFAVLVQYQNGDGLSGVELMQGG-----AGGAAWTPMQHWSWGA VWKLD-AGSALQ-APL     |
| Bd_5g17780          | TAFGAMAKPG---MADKLRSNGVLKIEFDRVPCKYN-GKKIS---FKMDSGAN-PYYLAMLIEY EAGDGL SAV EVMEAGG-----SKGSAKWMPMRQSWGALWCLDSKSGKPLQAPF   |
| Bd_5g17770          | TAFGAMANPG---MADRLRSAGELRIHYARVPCN YP-GMNVA---FRVDQGSN-PYYFAVLVQYQAGDGLSS V EIMQGGCRAGQH--GRNGEPWRMRMQSWGATWFLNSNDGKPLRAPF |
| Bd_3g50740          | TALGAMAKPG---KADALRAGGVLR IQYKR VPCKYP-GVNIA---FRVDQGSN-PFYFKALIEFQDGDGDLKR VALKQAGS-----NQWTNMVQDWGA LWRLNNG--QRLRGPF     |
| Bd_3g50750          | TSMGAMAKPG---MADKLRSAGILKIQYKR VPCKYS-GVNIA---FRVDQGSN-PFYLEVLIEFEDDDGDGLNAVDLMEANS-----GTWTPMVQNWGA LWRLNSNTGKPLRGPF      |
| Sb_004G294300       | TSMGAMAKPG---MADRLRAAGILKIQYKR VPCKYS-GVNIA---FRVDQGSN-PFYLEVLVEFEDGDGDL SAV DLM EAGC-----GTWTPMVQNWGA LWRYNSNTGKALKAPF    |
| Sb_004G294400       | TSMGAMAKPG---MADKLRAAGILKIQYKR VPCKYS-GVNIA---FRVDQGSN-PFYFEVLVEFEDGDGDL SAV DLM EAGC-----GTWTPMVQNWGA LWRYNSNTGKALKAPF    |
| Sb_002G124400       | TAMGAMAKPGR-GMADKVRAGGVVRVQYRRVPCKYP-GVNVA---FKVDQGAN-PFYFNVLIEFEDDDGDGL SAV DLM EAGS-----SVWTPMAHNWGA TWRLNNG--RRLRAPF    |

Sb\_002G124500 TAMGAMARPG---MADKLRLAGVLRILYRRVPCKYP-GVNVVA---FKVDQGAN-PFYFDVLIEFEDDDGDLKAVDLMEAGS-----SVWTPMAHNWGATWRLNNG--RRLRAPF  
 Sb\_006G171100 TAFGAMANRG---MADRLRAAGILKIYQYKRVSCNYN-NGAMGGIAFKVDRGSN-PFYLAVLIIQYLNNGDGLDAAVHIMQQG-----GAWTPMQHSWGAMWRANSNTGKPLRAPF  
 Sb\_006G171200 TAFGAMANPG---QADKLRLNAGVLQIYQYTRVPCNWN-GVKIT---FVVDAGSN-PEYFAVLIIKYVNGDGLSAVDLMQTG-----AG-ALWASMQPSWGAVWKFN-AGSALQ-APL  
 Sb\_006G171300 TAFGAMANRG---MADRLRSAGILKIYQYKRVPCNYN-GRTVN---FKVDAGSN-PYYLAVLIEYVAGDGEISAVDIMQAGY-----NSWTPMQQSWGAVWRVNSNNGQPLRVPF  
 Sb\_001G155600 TAFGALAKPG---QESQLRGAGAIQIYQYTRVQCEWP-GVDVT---FSVESGSN-PNYLAVLIEYEDSDSDLYSVDIMQGGG-----GTTGQWVPMQQSWGAVWKIN-SGAVLQ-GPF  
 Sb\_001G155700 TAFGAMAKPG---MADQLRAAGLLQIYQYTRVQCEWP-GVDVT---FSVDSGSN-PNYLAVLIEYEDSESDLSAVDIMQN-----SAGQWVPMQHSWGAIWRLD-SGSALH-GPF  
 Sb\_004G294500 TAFGALAKPG---QADQLRSAGLLKIYQYTRVPCNWN-GMDIA---FKVDGGSN-ANYLAVLIEYESGDGELRSVELMQR-----GGAAWAPMQRSWGAVWRYN-SGAALQ-APF  
 Sb\_010G120100 TAFGAMAKPG---QADQLRNVGRLPVQYQYTRVPCNWN-GVDVA---FRVDAGSN-ANYLAMAIEYEAGDGLRAVELQN-----GAAAWAPMERSWGAVWRYQSSGSSSLQAAPL  
 Divergent EXPB-I  
 CsEXPB1 IAFGSLALPG---KELDLHKAGKLAVLFKRVPCDYK-GRNIN---FHVEEGSN-PNYLALLTEYQNGEGDLSNLEIKEAGSST-----WTPLKQSWGKIWRLDS--PHPLKGPI  
 CsEXPB2 TAFGSMAYPG---RANELRHAGRLAIDFARVPCQYP-GRTIS---FHVEQGSN-PNYLAVLIEYEDGEDGLNMVEIMESNSRN-----WVQMKESWGAIYRMDT--NRPLHAPF  
 OsEXPB1 LAFGAMAKDG---KDEELRKAGIIDQFRRVCKYP-DTKVT---FHIEKASN-PNYLALLVKYVAGDGDVVEIKEKGSEE-----WKALKESWGAIWRIDT--PKPLKGGF  
 OsEXPB2 TAFGAMARPG---LNDQLRHAGIIDQFRRVPCYHR-GLYVN---FHVEAGSN-PVYLAVLVEFANKDGTVVQLDVMESLPSK-----PTRVWTPMRRSWGSIWRLDA--NHRLQGGF  
 OsEXPB3 TAFGAMAKPG---QNDQLRHAGIIDQFRRVPCNFP-GLKVT---FHVEEGSN-PVYFAVLVEYEDGDGDVVQVDLMEANSQ-----SWTPMRESWGSIWRLDS--NHRLTAPF  
 OsEXPB4 TAFGRLAKPG---LNDKLRLHSGIIDIEFTRVPCFP-GLKIG---FHVEEYSN-PVYFAVLVEYEDGDGDVVQVDLMESKTA-----PPTGKWTPMRESWGSIWRLDT--NHRLQAPF  
 OsEXPB6 TAFGAMAKPG---LNDKLRLHSGIIDQFRRVPCNYP-GLKIN---FHVEEGSN-PVYFAVLVEYEDLDGDGVVQVDLMESKSAGA-----TGVWTPMRESWGSIWRLDS--NHRLQAPF  
 OsEXPB7 TAFGAMAKDN---RNDELRLHAGIIDQFRRVPCQYP-GLTVT---FHVEQGSN-PVYMAILVEYENGDDGDVVQVDLMESRYS-----TPTGVWTPMRESWGSIWRLDT--NHPLQGGF  
 OsEXPB8 IAFGRLAKPG---RADDLRRAGIIDVQFARVPCFP-GLKVG---FHVEEGSS-PVYLAVLVEYENGDDGVAQVDLKEAGAG-----GGRWTPMRESWGSIWRLDS--NHRLRAPF  
 OsEXPB9 KAFGAMACPG---KETELRKAGIIDQFRRVCKYP-GQKVT---FHVEKGSN-PNYLAVLVKFVADDGDVIMQMDLQEAGLPA-----WRPMKLSWGAIWRLDT--ATPLKAPF  
 OsEXPB10 HAFGAMAKEG---KDEELRKAGIIDQFRRVCKYP-ETKVT---FHVEKGSN-PNYLAVLVKYVAGDGDVVEIKEKGSEE-----WKPLNESWGAIWRLDT--NRPLKGGF  
 Bd\_2g35007 TAFGALAMPG---RESELRRYGIIDLRFRVRCKLAPKMKIA---FHVEEGSN-PEYLAVLVKFVAGDGDIVQMDLKQEGWPE-----WKPMRESWGAIWRLDMDVYHTPLQGGF  
 Bd\_4g00360 KAFGSMAKKG---QEQLRGCGEVEIQFRRVCKYYPGLTKVT---YHVEKGSN-PNYLALLVKFVGGDGDVVAVEVQEKKGYN-----WIPLKESWGAIWRIDT--AKPLKGPL  
 Bd\_1g78120 HAFGSMAKKG---KEQDLRSAGEVEIQFRRVCKYPEGTKVT---FHVEKGSS-PNYLAILVKYVGGDGDVVAVDVKEKGKDE-----WVALKESWGAVWRLDT--AKPLKGPL  
 Bd\_1g78340 TAFGAMAKDQ---RNDELRLHAGIIDQFRRVACQYP-LTKVT---FHVEHGSN-PYYMAILVEYENGDDVDQVDLIMESTPDGSG---EPTGQWVPMKESWGSIWRLDT--RRPLMHGF  
 Bd\_1g78350 TAFGKLAKPG---RNDELRLHAGIIDQFTRVPCFP-GLKVG---FHVEEGSN-AVYMAILVEYENGDDGDVVQVDLMESGRG-----GGRWTRMKESWGSIWRLDS--NHRLQAPF  
 Bd\_3g33110 TAFGAMANYG---QNDQLRHAGIIDQFRRVPCNYP-GINVI---FHVEAGSN-PNYLAVLVLFANRDGNVVQMDVMESRYG-----GPTGYWPMRRSWGAIWRLDS--GHPLQGGF  
 Bd\_3g33150 TAFGALAKPG---LNDKLRLHSGIIDQFRRVPCNFP-GLKVT---FHVEQGSN-PVYFAVLVEYEDGDGDVVQVDLMEANSQG-----AWTPMRESWGSIWRLDS--GHRLQAPF  
 Bd\_3g33130 TAFGAMAKYG---QNNELRRAGIIDQFRRVPCNYQ-GMNVN---FRVERGSN-PNYLAVLVQHANKNDGNVQVDLKDSD-----GSYGSWTPMKRSWGAIWRLDT--GRPLKAPF  
 Bd\_3g33140 TAFGAMANPG---QNDQLRHAGIIDQFRRVPCNKP-GLNVN---FRVQHGSN-PNYLAVLVEYANVDGTVVQMDLRESRRGH-----PSGPWRPMRRSWGSIWRLDT--YRPLQGGF  
 Bd\_3g32810 TAFGLMAQPG---KEQLLRNCGELQLQFRRVCKLPPGTKIT---FHIEKGSN-PNYLAVLVKFASDDGDIVQMDLQDKISPE-----WKPMKESWGAVWRMDS--IKPLKGPY  
 Bd\_3g33160 TAFGRLAKPG---LNDKLRLHSGIIDIEFTRVPCFP-GLKIG---FHVEEYSN-PVYFAVLVEYEDGDGDVVQVDLMEASKTAHG---PLTGQWTRMRESWGSIWRLDS--NHRLQAPF  
 Bd\_3g33120 TAFGAMANNNG---QNDRLRHAGIIDQFRRVPCNYP-GMNVN---FHVERGSN-PNYLAVLVQHANKNDGNVVLMEIMESRYG-----RPTGQWTAMTRSWGAIWRLDT--NYPMQGPF  
 Sb\_004G1901600 TAFGAMAQPG---RNNELRHAGIIDQFRRVPCIYP-GQTVT---FHIEHGSN-PNYLAVLVEFEDGDGDVVQVDLMEANSGW-----WTPMRESWGSIWRLDS--KRRLQGGF  
 Sb\_001G539700 TAFGAMAKKG---EEEKLRKAGIIDMKFRRVCKY---GEKVT---FHVEKGSN-PNYLALLVKYVDGDGDVVGVDIKEKGGDA-----YQPLKHSWGAIWRKDS--DKPIKFPV  
 Sb\_001G301000 TAFGSMARPG---LNDKLRLHAGIIDQFRRVACNNR-GLTVN---FHVQAGSN-PNYLAVLVEYANKAGTVVQMDAMEAN-----SGHWMPMRRSWGSIWRLDS--YRALRGPF  
 Sb\_001G306200 KAFGSLAKPG---LNDKLRLHCGIMNVEFRRVRCKL---GGKIM---FHVEKGSN-PNYLAVLVKNVADDGNIVLMELEDKASPG-----FKPMKLSWGAVWRFD--PKPIKGGF  
 Sb\_001G300500 TAFGAMAKPG---RNNELRHAGIIDQFRRVPCNYP-GQKVT---FHVEEGSN-PVYFAVLVEFEDGDGDVVQVDLMEANSGY-----WTPMRESWGSIWRLDA--NHRLQAPF  
 Sb\_001G301300 TAFGAMAKTG---LNDKLRLHAGIIDQFRRVPCNYP-GLSVN---FRVQVGSN-PNYFAVLVQYAGKDGAVVQLDLMETSKATG---KPTGVWTPMRVSWGAVWRLDT--KRPLQPPF  
 Sb\_001G301500 TAFGAMASWG---LNDKLRLHAGIIDQFRRVPCNFP-GLTIT---FRVQHGSN-PNYLAVLIEHENQDGDVVQADIMQNN-----SGYWEPMHESWGSIWRIDP--NRPLRGPF  
 Sb\_001G306400 KAFGSLAKPG---LNDKLRLHCGIMNVEFRRVRCKL---GGKIM---FHVEKGSN-PNYLAVLVKNVADDGNIVLMELEDKASPG-----FKPMKQSWGAVWRFD--PKPIKGGF  
 Sb\_001G300400 TAFGRLAKPG---LNDKLRLHSGIIDIEFTRVPCFP-GLKIG---FHVEEYSN-PVYFAVLVEYEDGDGDVVQVDLMEASKTAHG---PPTGRWTPMRESWGSIWRLDT--RRPLQPPF  
 Sb\_001G306500 KAFGSLAKPG---LNDKLRLHCGIMNVEFRRVRCKL---GGKIM---FHVEKGSN-PNYLAVLVKNVADDGNIVLMELEDKASPG-----FKPMKQSWGAVWRFD--PKPVKGGF  
 Sb\_001G539600 TAFGAMAKKG---EEEKLRKAGIIDQFRRVCKY---GEKVT---FHVEKGSN-PNYLALLVKYVDGDGDVVGVDIKEKGGEA-----YQPLKHSWGAIWRKDS--DKPIKFPV  
 Sb\_001G301400 TAFGAMAKPG---LNGKLRLNAGALSIIYRRVPCNYP-GLNVR---FHVMGGSN-PFYFAVIVYAGSDGAVVQVELKEANSR-----TWRLYESWGAVWRIDP--GHPLRAPL  
 Sb\_001G542200 TAFGKLAKPG---RNDELRLHAGIIDIEFTRVPCFP-GLTVT---FHVEEGSS-QVYFAVLVEYEDGDGDVVQVDLMEASKTAHG---S-----SRRWTPMRHSWGSIWRLDT--NRPLQPPF  
 Sb\_001G542100 TAFGAMAKDD---RNDELRLHAGIIDQFRRVPCQYP-GLTVT---FHIERGSN-PNYLAVLVEYNGDGDVVQVDLMEASRTDD-G---EPTGVWEPMRESWGSIWRLDT--RRPLQGGF  
 Sb\_001G540000 TAFGAMAKKG---EEEKLRKAGIIDQFRRVCKY---GEKVT---FHVEKGSN-PNYLALLVKYVDGDGDVVGVDIKEKGGDA-----YQPLKHSWGAVWRKDS--DKPIKFPV  
 Sb\_001G300700 TAFGAMAKPG---RSDRLRHAGIIDQFRRVPCNYP-GQKVT---FHVEEGSN-SVYLAVLVEFEDGDGDVVQVDLMEANSGY-----WTPMRESWGSIWRLDS--NHRLQAPF  
 Sb\_001G300800 TAFGAMAKPG---RSDRLRHAGIIDQFRRVPCNYP-GQKVT---FHVEEGSN-PVYFAVLVEFEDGDGDVVQVDLMEANSGS-----WTPMRESWGSIWRLDS--GHRLTAPF  
 Sb\_001G300900 TAFGAMAKPG---RNDELRLHAGIIDQFRRVPCYYP-GQKVT---FHVEVGSN-PVYFAVLVEFEDGDGDVVQVDLMEANSGS-----WTPMRESWGSIWRLDS--DHRLTAPF

|                     |                                         |
|---------------------|-----------------------------------------|
| AtEXPL1             | RFVVTGGYDGKMIWSQ-SVLPSNWEAGKIYDAGVQI    |
| OsEXPL2             | RVVVTGGYDGKVVWADGEVLP RRWTAGRVYDAGVQI   |
| EXPB-II             |                                         |
| Pp_022G002600       | TIKITTKLNDHTVVAE-DVIPDWFGPGVQYESNVQI    |
| Pp_022G064000       | DVMITAKLNGHVLI AQ-DAIPEYFQPGGIYESNVQL   |
| Pp_017G043100       | SVKINCLMNGHTVVAK-DVIPAGFAPGQEYESNVQI    |
| Pp_017G043200       | DVKIVSKLNGHTVIAE-KAIP EFFFEPGKLYESNVQM  |
| Pp_001G152100       | DIRIVSRLNRHTV LAR-GVIPAYFRPGTSYRSRVQM   |
| Sm_36179            | SVRVTTVGGEVATCVL-TGITS-WSPGETST-----    |
| Sm_171427           | SIRLTALQTGEKIIAH-NVIPANWAPQHSYSTGVNF    |
| Sm_234188           | SFRVTTLTSETVVAR-NVIPRNWASNTCYRSRVNF     |
| AtEXPB1             | SVKLTTLSNNKTL SAT-DVIPSNWVPKATYTSRLNF   |
| AtEXPB3             | SIKLTTLSAGKTL SAT-DVVPRNWAPKATYSSRLNF   |
| PtEXPB2             | SVKLTTTLTTGR TLSAR-EVIPRNWAPKATYTSRLNF  |
| PtEXPB3             | SVKLTTTLSTGR TLSAR-DVIPRNWVPKATYTSRLNF  |
| Sl_07g054170        | SVKLTTTLSTGR ALSAR-DVIPGNWSPKATYTSRLNF  |
| Sl_10g008440        | SIKLTTTLTKKRTLSAR-DVIPSKWTPKATYTSRLNF   |
| Vv_01017505001      | SIQLTSQYSGQTLVAK-NVIPDGWKPGSTYRSLVNY    |
| Ma_randomP06310     | SVKLTTTLTTQR TLSAR-DVIPRNWSPKATYTSRLNF  |
| Ma_5P12080          | SVKLTTTLTTQKTFSAR-DVIPSNWSPDVITYTSRLNF  |
| Ma_1P00120          | SVKLATLTTTRKTFSAR-DVIPRNWSPKATYTSRLNL   |
| Eg_4P60075          | SVKLTTTLSTQR TLSAR-DVIPKNWSPKATYTSRLNF  |
| Bd3_g49850          | SVRLTTLT TTKKTL SAQ-DVIPRNWTPKATYTSRLNF |
| Bd_5g16497          | SVRLTTLSAPKTLTAR-DVIPRNWAPRGTYTSRLNF    |
| Sb_004G227900       | SVRLTTLSSKQTLTAQ-DVIPKNWAPKATYTSRLNF    |
| Sb_006G156900       | SVRLTTLSARKTLTAR-DVIPRNWTPKATYTSRLNF    |
| OsEXPB16            | SVRLTTLTTRQTLSAQ-DVIPKNWTPKATYTSRLNF    |
| OsEXPB17            | SVRLTTLSAQKALTAR-DVIPRNWKPTATYTSRLNF    |
| Conservative EXPB-I |                                         |
| AtEXPB2             | SLRVTS-ESGKT VVAS-NVIPANWQPGAIYKSNVNF   |
| AtEXPB4             | NIRLTSE-SHKVIVAY-NVIPANWKPD ESYRSIVNF   |
| AtEXPB5             | NIRLTSV-SGKKIIAK-GV IPEKWSPGAIYHSKVNF   |
| PtEXPB1             | SIKLTS-ESGKTIVAS-GVIPAGWAPGQTYRSLVNF    |
| Sl_10g018500        | SIRLTTLDSNKILVAN-NVIPLDWAPGLIYHSLVNF    |
| Sl_03g093390        | SIRLTSQYSDQILIAN-NVIPTGWQPGATYRSLVNY    |
| Sl_07g049540        | SLRLTSG-TKQSVTAL-NVIPDGWQPTQFYNSNVNF    |
| Sl_05g052330        | SIRLTSQ-NKHQVEAN-NVIPINWQAREIYYSNVNF    |
| Sl_01g090810        | SLKLTG-ESGQTVVAN-GVIPAGWQPGKTYRSVVNF    |
| Sl_01g090820        | SIRVTTG-SNKKLTAT-NVIPVGWQPGRTYRSNVNF    |
| Vv_01018163001      | SIKLTALESGNTLVAK-NVIPAGWQPGKTYRSVVNF    |
| Vv_01018168001      | SIRLTALESGNTLVAK-NVIPAGWQPGKTYRSLVNF    |
| Vv_01030375001      | SVKLTTTLSTGR TLSAR-EVIPRNWAPKATYTSRLNF  |
| Ma_7P07040          | SFRLTSGLSR KVLVAN-NVIPVGWQPGATYRSLVNY   |
| Ma_randomP23770     | SIRLTSGLSGKTLVAT-NVIPVDWKPGATYNSTVNF    |
| Ma_2P13740          | SIRLTSAVSRKTIVAT-NVIPAGWKPGATYSSNVNF    |
| Ma_2P13750          | SIRLTSAVSRKTIVAT-NVIPAGWRPGATYYSNVNF    |
| Ma_2P13730          | SIRLTSAVSSKTIVAT-NVIPAGWRPGATYYSNVNF    |
| Ma_4P26910          | SIRLTTLLSNRTIVAD-DVIPSTWLPGETYRSTTH-    |
| Eg_4P156078         | SIRLTTQYSGQKLVAN-NVIPTGWQAGTTYRSFVEQ    |

|                  |                                       |
|------------------|---------------------------------------|
| Eg_4P79642       | SIQLTSQYSGKKLVAN-NVIPAGWKPGATYRERRTR  |
| Pd_30s825291g003 | SIQLTSQYSGKKLVAN-NVIPAGWKPGATYRSFVNY  |
| OsEXPB5          | SIRLTSG-SGKVLVAN-NVIPSGWQAGLTYRSTVNY  |
| OsEXPB11         | SLRLTSD-SGKVLVAN-NVIPASWKPGVITYRSLVNY |
| OsEXPB12         | SVRLTFS-SGQMFVAS-NAIPAGWNPGMAYQPG-VA  |
| OsEXPB14         | SIRLTSG-SGRTLVAS-NVIPAGWQPGGTYRSIVNF  |
| OsEXPB15         | SIRLTSS-SGKTLVAS-NVIPSGWKPGMSYISTVNF  |
| OsEXPB18         | SIRLTSS-SGKTLVAS-NVIPSGWKPGASYTSTVN-  |
| Bd_3g50730       | SVRLTSG-SGKTLVAD-NVIPAGWKPGGTYHSVVNY  |
| Bd_4g41110       | SIRLTSG-SGRKLVAA-DVIPAGWQPGRTYRSVVNY  |
| Bd_4g41117       | SIRLTSS-SGKKLVAA-NVIPVGWQPGRTYRSLVNY  |
| Bd_5g17760       | SLRLTSS-SGKKLVAS-NVIPVGWKAGAAYSQSAVNY |
| Bd_5g17780       | SFRLTSG-SGKVLVAN-NVVP TGWNAGKAYQANVNY |
| Bd_5g17770       | SFRLTSG-SGKVLEVT-NAIPAGWRAGTSYRSSVNY  |
| Bd_3g50740       | SIKLTSD-SNRVLVVN-NVIPANWKAGATYRSLVNY  |
| Bd_3g50750       | SLRLTSD-SNRKLVVN-NVIPATWKAGQTYRSLVNY  |
| Sb_004G294300    | SLRLTSD-SGKVLVAN-NVIPAGWNPATYRSLVNY   |
| Sb_004G294400    | SLRLTSD-SGKVLVAN-NVIPAGWNPATYRSLVNY   |
| Sb_002G124400    | GLRLTSD-SGRVLVAS-NAIPAAWTPGKTYRSLVNY  |
| Sb_002G124500    | GLRLTSD-SGRVLVAN-NAIPAAWKPGKAYRSLVNY  |
| Sb_006G171100    | SVRLISG-SGKVLVVR-NAIPAGWRAGMTYWSTVNY  |
| Sb_006G171200    | SIRLTSS-SGKQLVAS-NVIPAGWTPGATYQSSVNY  |
| Sb_006G171300    | SVRITSG-AGKVLVAR-NAIPAGWSAGRTYRSVNF   |
| Sb_001G155600    | NIRLTFS-SGRVLVAG-NAIPAGWNAGVAYRSGGVA  |
| Sb_001G155700    | HLRLTFS-SGRVLIAS-NAIPAGWTPGVAYRAGGVA  |
| Sb_004G294500    | SVRLTSG-SGRTVVAT-NVIPAGWSPGATYRSVVNF  |
| Sb_010G120100    | SVRLTSG-SGKTLVAS-GVIPAGWQPGNTYRCVVNF  |
| CsEXPB1          | SVKVTTL-SGKSIVAN-DAIPADWKPNKYDSKVNF   |
| CsEXPB2          | SIRMTDD-KNRKLVAR-NAIPANWRPNTFYRSVVQY  |
| OsEXPB1          | SVRVTTT-GGEKIIAE-DAIPDGWKADSVYKSNVQA  |
| OsEXPB2          | SLRMVSE-SGQTVIAH-QVIPANWRANTNYGSKVQF  |
| OsEXPB3          | SLRITNE-SGKQLVAS-QVIPANWAPMAVYRSFVQY  |
| OsEXPB4          | SIRIRNE-SGKTLVAN-NVIPANWRPNTFYRSFVQY  |
| OsEXPB6          | SLRIRSD-SGKTLVAN-NVIPANWSPNSNYRSIVQF  |
| OsEXPB7          | SLRITNE-SGKTLIAD-QVIPADWQPNNTVYSSIVQF |
| OsEXPB8          | SIRIRSD-SGKTLVAP-DVIPLNWPNTFYRSFVQY   |
| OsEXPB9          | SIRVTTE-SGKSIIAK-DVIPVNWMPDAIYVSNVQF  |
| OsEXPB10         | SLRVTTT-SDQKLVAN-DVIPDNWKANALYKSEIQV  |
| Bd_2g35007       | SIRLTSE-SGKVLVAP-HVIPADWKPKTV-----    |
| Bd_4g00360       | SVRYTTD-GGTKAVSP-DVIPEKWKPD TMYVAKY-- |
| Bd_1g78120       | TVRYTTD-GGTKGESE-DVIPEDWKPD TMYESK--- |
| Bd_1g78340       | SLRITNE-SGQTLVAD-QVIPADWEPNAIYSSIIQF  |
| Bd_1g78350       | SLRIRNE-SGKTLVAR-NVIPSNWRPNTFYRSIVQY  |
| Bd_3g33110       | SLRIRSE-FGATLVAH-QAIPADWKPD TNYWSNIQY |
| Bd_3g33150       | SLRITNE-SGKTLVAD-KVIPANWAPS AVYRSFVQY |
| Bd_3g33130       | SFRMRSE-FGATKVAY-QVIPVNWKGNNYWSNVQY   |
| Bd_3g33140       | SLRIRSE-SGQTLTAK-QVIPANWKPD TDRSRVQF  |
| Bd_3g32810       | SIRLTSE-SGKKIIAK-DIIPLNWKPD TFLKSNIQF |
| Bd_3g33160       | SIRITNE-SGKRLIAR-EVIPANWRPNTFYRSFVQY  |
| Bd_3g33120       | SLRIRSE-SGSTLVAN-QAIPADWKPNARYWSNIQY  |

|               |                                       |
|---------------|---------------------------------------|
| Sb_004G191600 | SLRITNE-SGERLVAY-QVIPANWAPNTYYRSNIQY  |
| Sb_001G539700 | TVQITTE-GGKTAYE-DVIEGWKADTTYTAK---    |
| Sb_001G301000 | SMRIRSE-NGRTL VAN-NVIPANWRANTDYRSYVQF |
| Sb_001G306200 | SIRLTSE-SGKKLVAP-NVIPATWKPD TLYNSNIQF |
| Sb_001G300500 | SLRITNE-SGRKLVAN-RVIPANWAPNNYYRSIIQY  |
| Sb_001G301300 | SLRVRNG-SGKTLVAS-NVIPANWKPM TDYPSNVQF |
| Sb_001G301500 | SIRITNE-SGKQLVAK-NIIPANYVADVDYRSYVQY  |
| Sb_001G306400 | SIRLTSE-SGKKLVAP-NVIPASWKPD TLYNSNIQF |
| Sb_001G300400 | SIRIRNE-SGKTLVAR-NVIPANWRPNTFYRSFVQY  |
| Sb_001G306500 | SIRLTSE-SGKKLVAP-NVIPATWKPD TLYNSNIQF |
| Sb_001G539600 | TIQITTE-GGKTAYE-DVIEGWKADTTYTAK---    |
| Sb_001G301400 | SLRVRSD-GGKTLVAY-DVIPINWRGNADYRTIAQF  |
| Sb_001G542200 | SIRTRSD-SGKTLVAR-DVIPLNWRPNTFYRSIVQY  |
| Sb_001G542100 | SLRVTNE-SGKTLVAD-QVIPADWQPDNVYSSIVQF  |
| Sb_001G540000 | TVQITTE-GGKTAYE-DVIEGWKADTTYTAK---    |
| Sb_001G300700 | SLRITNE-SGRTL VAN-QVIPANWVPNTYYRSIIQY |
| Sb_001G300800 | SLRITNE-SGKTLVAN-QVIPANWVPNTYYRSIVQY  |
| Sb_001G300900 | SLRITNE-SGKTLVAN-QVIPANWTPNTYYRSMVQY  |

Figure S2

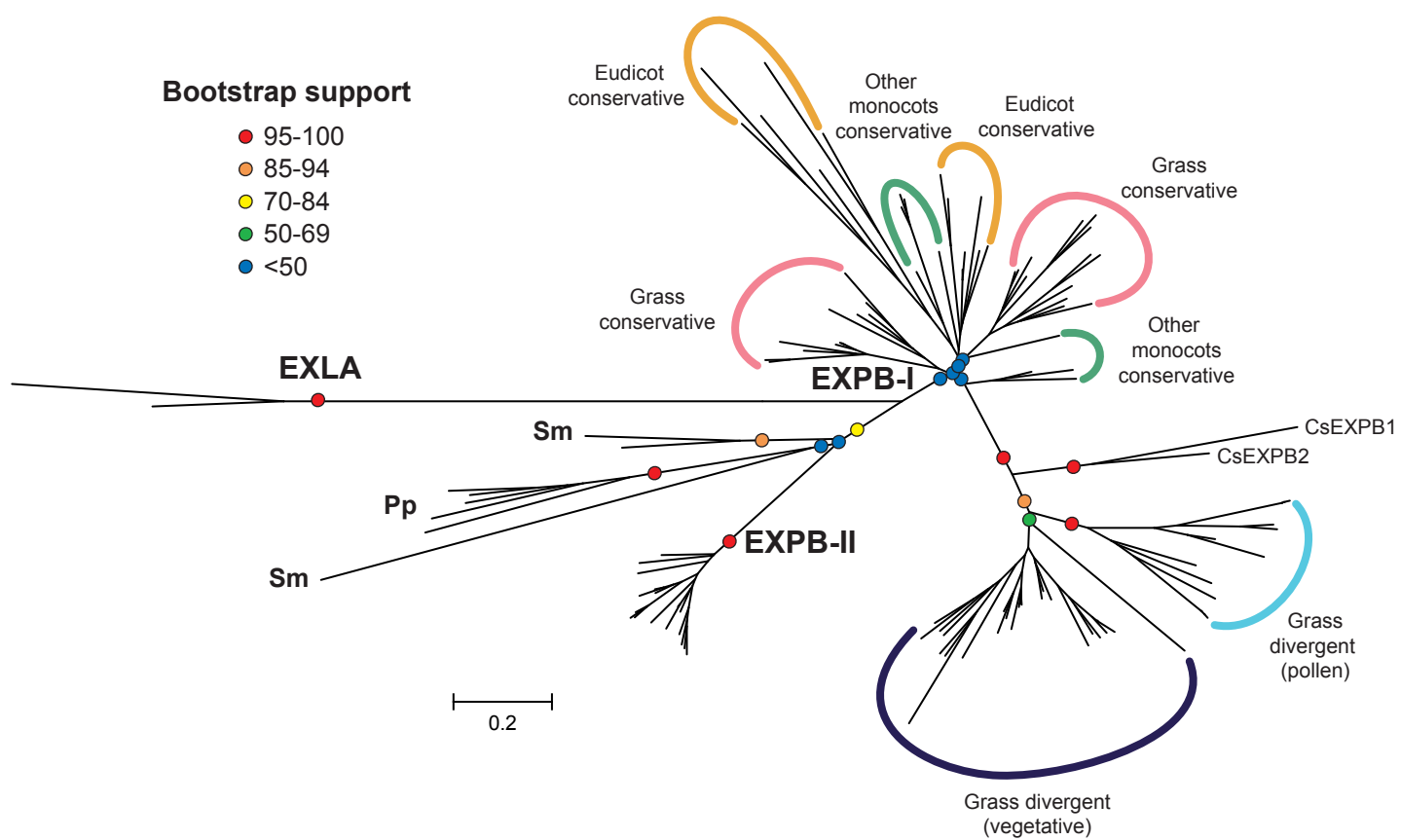


**Table S1.** Primers used in this work.

| Name  | Sequence                          |
|-------|-----------------------------------|
| 5POL  | 5'-GGGCAAAGCCAGTGAGCCAAA-3'       |
| 3POL  | 5'-GGCATTGCATTTGGCTCACTG-3'       |
| 5VEG  | 5'-GATATGAGACGGCCTGCCTGAGC-3'     |
| 3VEG  | 5'-GCCTCCTTTCAATTCCATGACATC-3'    |
| 3VEGN | 5'-GGAGGAAAAGGATGCGGTTCTTG-3'     |
| POLFL | 5'-ATCTTTCTTTCATTTGCCACCTT-3'     |
| POLFR | 5'-TTGTTTACCAAATATTTCAAACGATCT-3' |
| VEGFL | 5'-AAATCAGAACCAAATCACCCCTTTT-3'   |
| VEGFR | 5'-GATGACCATACCCAGCCCTCA-3'       |
| POLCL | 5'-GCTGTTCTTAAGGGCGATGT-3'        |
| POLCR | 5'-TCACAAGGGACCCTCTTAAACA-3'      |
| VEGCL | 5'-AATTACAACACCTCTGCGGTT-3'       |
| VEGCR | 5'-TACTGGCATGGCACACGAG-3'         |

**Table S2.** Analysis of partial EXPB sequences from banana, date palm and oil palm genomes.

| Species                    | Sequence                      | Clade   |
|----------------------------|-------------------------------|---------|
| <i>Elaeis guineensis</i>   | EG4P79726                     | EXPB-II |
|                            | EG4P63786                     | EXPB-II |
|                            | EG4P63698                     | EXPB-II |
|                            | EG4P108181                    | EXPB-II |
|                            | EG4P63687                     | EXPB-II |
| <i>Musa acuminata</i>      | GSMUA_Achr2T11930_001         | EXPB-I  |
|                            | GSMUA_Achr1P04260_001         | EXPB-I  |
|                            | GSMUA_Achr5P15440_001         | EXPB-I  |
|                            | GSMUA_AchrUn_randomP23760_001 | EXPB-II |
|                            | GSMUA_AchrUn_randomP23780_001 | EXPB-II |
| <i>Phoenix dactylifera</i> | PDK_30s868001g002             | EXPB-II |
|                            | PDK_30s893441g001             | EXPB-I  |