The Plant Journal (2006) **46**, 723–735 doi: 10.1111/j.1365-313X.2006.02741.x

Gene duplication, exon gain and neofunctionalization of *OEP16*-related genes in land plants

Sinéad C. Drea[†], Nga T. Lao, Kenneth H. Wolfe and Tony A. Kavanagh[*]
Plant Molecular Genetics Laboratory, Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland

Received 30 November 2005; revised 25 January 2006; accepted 27 January 2006.
[*]For correspondence (fax +353 1 6798558; e-mail tkvanagh@mail.tcd.ie).
[†]Present address: Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520-8104, USA.

Summary

OEP16, a channel protein of the outer membrane of chloroplasts, has been implicated in amino acid transport and in the substrate-dependent import of protochlorophyllide oxidoreductase A. Two major clades of OEP16-related sequences were identified in land plants (OEP16-L and OEP16-S), which arose by a gene duplication event predating the divergence of seed plants and bryophytes. Remarkably, in angiosperms, *OEP16-S* genes evolved by gaining an additional exon that extends an interhelical loop domain in the pore-forming region of the protein. We analysed the sequence, structure and expression of the corresponding Arabidopsis genes (*atOEP16-S* and *atOEP16-L*) and demonstrated that following duplication, both genes diverged in terms of expression patterns and coding sequence. *AtOEP16-S*, which contains multiple G-box ABA-responsive elements (ABREs) in the promoter region, is regulated by *ABI3* and *ABI5* and is strongly expressed during the maturation phase in seeds and pollen grains, both desiccation-tolerant tissues. In contrast, *atOEP-L*, which lacks promoter ABREs, is expressed predominantly in leaves, is induced strongly by low-temperature stress and shows weak induction in response to osmotic stress, salicylic acid and exogenous ABA. Our results indicate that gene duplication, exon gain and regulatory sequence evolution each played a role in the divergence of *OEP16* homologues in plants.

Keywords: OEP16, seed, pollen, exon gain, gene duplication, G-box ABA-responsive elements.

Introduction

A defining characteristic of all plant cells is the presence of organelles collectively called plastids. In flowering plants, which show the greatest diversity of plastid types, these include the undifferentiated proplastids of meristematic tissues, chlorophyll-containing chloroplasts of green leaves, colourless amyloplasts and leucoplasts found in starch-storing and oil-storing organs, respectively, and yellow, orange or red chromoplasts found in flowers and fruits (Kirk and Tilney-Basset, 1978). All of these specialized plastids differentiate in a cell-, tissue- or organ-specific manner, either directly from proplastids or indirectly by redifferentiation of pre-existing types. The remarkable diversity of plastid form and function reflects their specialized metabolic roles in different tissues and organs during growth and development (Thomson and Whatley, 1980). Consequently, the regulation of plastid metabolic activities is complex and highly responsive to both cellular developmental programmes and external factors such as light, temperature and nutrient availability (Bauer *et al.*, 2001).

All plastids are surrounded by two membranes (the inner and outer envelopes) that regulate the import of the protein constituents of the plastid and, in addition, the import and export of metabolites. The plastid genome encodes a mere 2% of the estimated 3500–4000 proteins comprising the plastid proteome. Thus, the vast majority of plastid proteins are encoded by nuclear genes and are imported into the organelle following synthesis in the cytoplasm (Abdallah *et al.*, 2000). The protein import apparatus comprises two well-characterized multisubunit complexes termed translocon of the outer chloroplast (Toc) and translocon of the inner chloroplast (Tic) envelope. The Toc complex, comprising the core components Toc33/Toc34, Toc75 and Toc159, recognizes and binds plastid proteins and either inserts them into the outer membrane or passes them to the Tic

complex for import across the inner membrane (reviewed in Soll and Schleiff, 2004).

The spectrum of metabolic activities carried out by plastids also requires the import and export of inorganic cations (K^+, Na^+, Mg^{2+} and Ca^{2+}), anions (nitrite, sulphate and phosphate) and a variety of metabolites such as acetate, phosphoenolpyruvate, dicarboxylic acids, maltose, amino acids and ATP. This exchange of ions and metabolites with the surrounding cytoplasm is regulated by the activities of specific carrier and channel proteins located in the outer and the inner plastid envelope membranes (reviewed in Neuhaus and Wagner, 2000). It is generally accepted that the inner envelope plays the major role in this process. This conclusion is based on solute transport studies confirming its role as the key permeability barrier, and the identification of carrier proteins that show clearly distinct substrate selectivity and specificity (Flugge, 1998). The role of the outer membrane in regulating the flow of metabolites and ions between the plastid and the cytoplasm is less clear (Bolter and Soll, 2001; Flugge, 2000; Soll *et al.*, 2000). Electrophysiological investigations initially indicated a largely non-specific and apparently free diffusion of low molecular mass solutes across the outer membrane via porin-like channels, suggesting that its role as a selective permeability barrier might be negligible (Flugge and Benz, 1984). However, evolutionary and other considerations suggest that retention of the outer envelope following the ancient endosymbiotic event that gave rise to modern plastids argues strongly for a role not only in the import of proteins into plastids but also in controlling specific metabolite and ion fluxes between the plastid and the cytoplasm (Soll *et al.*, 2000).

The majority of proteins in the plastid outer envelope membrane resemble the channel-forming porins of the outer membrane of Gram-negative bacteria. The *Escherichia coli* outer membrane, for example, contains three different porin classes: general porins that do not show specific solute binding, low-affinity solute-selective porins, and energy-dependent, ligand-gated, high-affinity porins (Klebba and Newton, 1998). Solute channel proteins showing similar transport characteristics have been identified in the outer membrane of pea chloroplasts. OEP16, the first isolated outer membrane channel protein (Pohlmeyer *et al.*, 1997), has been investigated in considerable detail. Recombinant OEP16 reconstituted into liposomes shows low-affinity, but selective, permeability to amino acids and other compounds with primary amino groups. It is not, however, permeable to either phosphorylated or unphosphorylated sugars, even though the pore is sufficiently large to allow their passage (Linke *et al.*, 2000).

Recent evidence suggests that in addition to its putative role in solute transport, OEP16 may also collaborate with subunits of the Toc complex in the import of the protochlorophyllide oxidoreductase A pre-protein (pPORA), a key enzyme of chlorophyll biosynthesis. The import of pPORA into greening plastids is unusual in that it is regulated by substrate availability (Reinbothe *et al.*, 1995). Chemical cross-linking experiments have revealed an association between pPORA and various plastid envelope proteins, including OEP16 and Toc33, which are proposed to function as part of a protochlorophyllide-dependent pPORA translocon (Reinbothe *et al.*, 2004a). Furthermore, anti-OEP16 but not anti-Toc75 antibodies inhibited pPORA import, thus suggesting a functional role for OEP16 in pre-protein translocation and a possibly specific interaction between OEP16 and Toc33 (Reinbothe *et al.*, 2004b).

Most investigations of plastid outer envelope proteins, such as OEP16, have focused almost exclusively on the role of individual proteins in transport processes and their biophysical properties. As a consequence, little is known about the structure, evolutionary origins or transcriptional regulation of the corresponding genes or their homologues. Here we show that the two most closely related Arabidopsis *OEP16* genes are members of two distinct conserved subgroups, *OEP16-L* and *OEP16-S*, which originated by a gene duplication event that occurred prior to the divergence of bryophytes and seed plants. Moreover, following duplication, neofunctionalization of *OEP16-S* occurred via an unusual process of lineage-specific exon gain as well as regulatory sequence evolution and development of organ-specific expression patterns.

Results

Arabidopsis homologues of pea OEP16

Genes encoding four putative homologues of pea OEP16 (designated *atOEP16-1,-2, -3, -4*) were recently identified in Arabidopsis (Reinbothe *et al.*, 2004b). In pairwise comparisons, we noted that atOEP16-1 shares the highest level of amino acid identity with pea OEP16 (63%) and is the Arabidopsis OEP16 orthologue. Of the other homologues, atOEP16-2 shares 32% identity with pea OEP16. However, atOEP16-3 and atOEP16-4 share only marginally higher identity with pea OEP16 (21–22%) than various members of the Translocase of the Inner Membrane (TIM) superfamily of mitochondrial transport proteins (e.g. AtTim17-3 with 16% identity). For this reason, we focused our analyses on the most closely related Arabidopsis genes, *atOEP16-1* and *-2*, which were renamed *atOEP16-L* and *atOEP16-S*, respectively, because of their predominant expression in leaves (L) and seeds (S), respectively (see below).

OEP16-L and -S proteins are conserved in all land plants

Genes or expressed sequence tags (ESTs) encoding proteins similar to OEP16-L and OEP16-S were identified in each of the major land plant lineages. A phylogenetic analysis

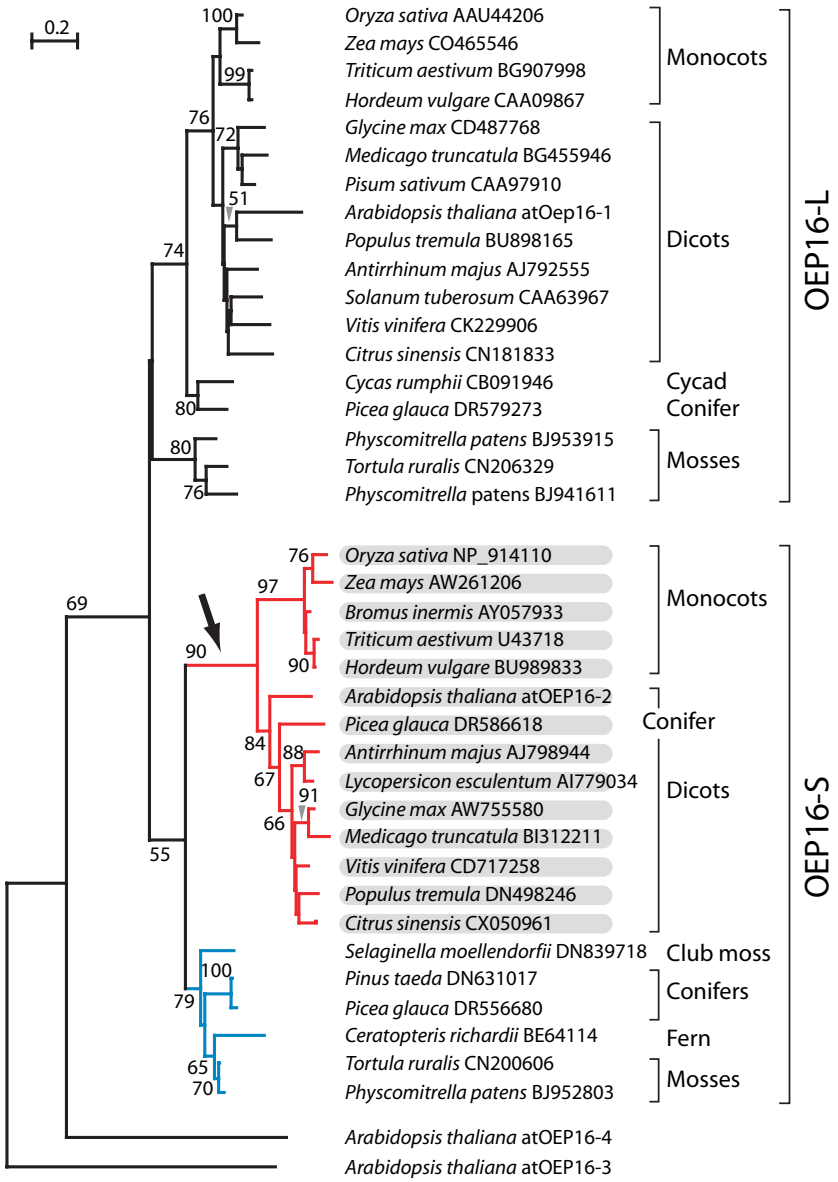

Figure 1. Gene duplication and S-domain gain during OEP16 evolution. The two major clades within land plants are labelled OEP16-L and OEP16-S. Grey highlighting and red branches indicate sequences that contain the additional S-domain, and the arrow indicates the inferred point of gain of this domain. The blue branches are OEP16-S sequences from non-angiosperm species that lack the S-domain. Bootstrap percentages are shown where they exceed 50%. The Arabidopsis OEP16 genes are named as in Reinbothe *et al.* (2004b). Their systematic names and National Centre for Biotechnology Information, USA (NCBI) protein accession numbers are: atOEP16-1 (renamed atOEP16-L) (At2g28900, AAC79594); atOEP16-2 (renamed atOEP16-S) (At4g16160, CAB10395); atOEP16-3 (At2g42210, AAB88646); atOEP16-4 (At3g62880, CAB83138). For other species, NCBI protein or DNA accession numbers are shown. The phylogenetic tree was constructed from amino acid sequences using the neighbour-joining method with atOEP16-3 and atOEP16-4 as outgroups.

demonstrated that there are two major clades corresponding to OEP16-L- and OEP16-S-like sequences (Figure 1). As expected, the OEP16-L group includes pea OEP16 and atOEP16-L. The gene duplication that gave rise to these two groups is seen to predate the divergence between bryophytes and seed plants. Parts of the tree are poorly resolved and have low bootstrap support, but in general the branching order seen within the OEP16-L and OEP16-S parts of the tree are in reasonable agreement with the expected species phylogeny. The most notable exception is the existence of two divergent types of OEP16-S sequence in the conifer *Picea glauca* (spruce), one of which falls in the expected position in the tree and the other falls among the dicot sequences. That a dicot-like OEP16-S sequence exists in conifers is confirmed by multiple independent EST sequences from spruce, and by a highly similar EST from *Pinus taeda* (accession number DN613327; not shown in Figure 1 because it is not full length). We speculate that this gene may possibly have originated by horizontal gene transfer, because otherwise the species distribution seen in the OEP16-S part of Figure 1 can only be explained by very large numbers of gene losses.

OEP16-S proteins of angiosperms contain an additional domain located between transmembrane helices 1 and 2

A multiple alignment of OEP-L and -S sequences revealed the presence of an additional 20–27 amino acid domain (which we refer to as the S-domain) in the OEP16-S proteins of angiosperms and in the dicot-like sequence from spruce

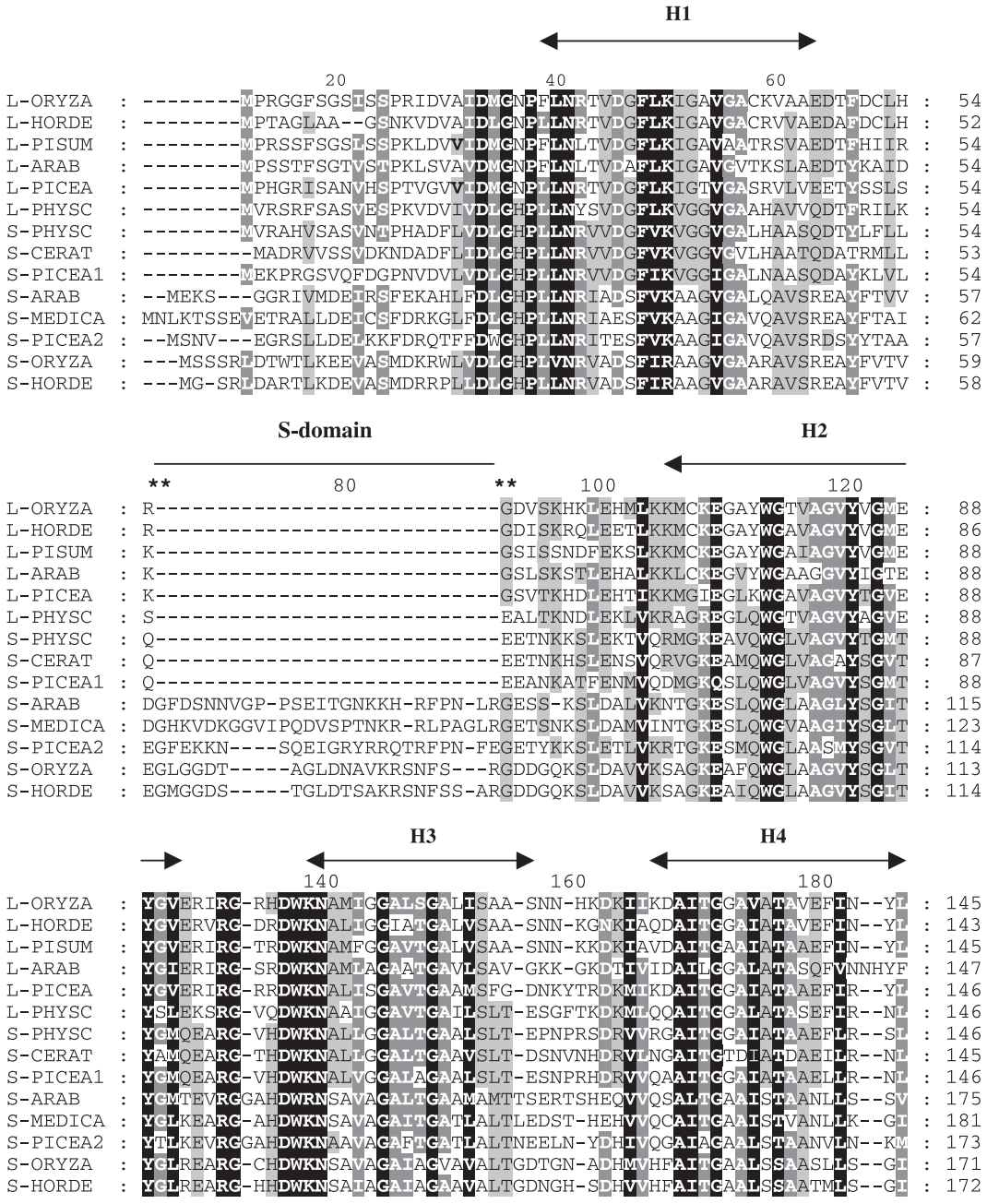

Figure 2. Multiple alignment and domain structure of OEP16 proteins. OEP16-L and OEP16-S amino acid sequences from *Arabidopsis thaliana* (ARAB), *Ceratophyllum richardii* (CERAT), *Hordeum vulgare* (HORDE), *Medicago truncatula* (MEDICA), *Oryza sativa* (ORYZA), *Pisum sativum* (PISUM), *Physcomitrella patens* (PHYSC), *Picea glauca* (PICEA), *Populus tremula* (POPUL) were aligned using CLUSTALW. The corresponding accession numbers are shown in Figure 1. S-PICEA1 and S-PICEA2 refer to accession numbers DR586618 and DR556680, respectively. Identical and similar residues, shared by five or more sequences, are boxed in black and grey, respectively. Regions corresponding to the four predicted transmembrane helical segments (H1–H4) identified by Linke *et al.* (2004) are indicated by double-headed arrows. The conserved amino acids flanking the S-domain insertion site are indicated by asterisks.

(Figure 2; grey shading in Figure 1). Because the OEP16-S proteins of bryophytes, gymnosperms, a pteridophyte and a club moss all lack this domain and group together in the phylogenetic tree (blue branches in Figure 1), we infer that the *OEP16-S* gene structure changed during evolution. The sequence encoding the S-domain was apparently inserted into the *OEP16-S* gene in an ancestor of angiosperms after it had diverged from gymnosperms, as indicated by the arrow in Figure 1.

The structure of pea OEP16, which lacks the S-domain, has been investigated in considerable detail. A deletion mutagenesis study identified amino acid residues 21–93 as

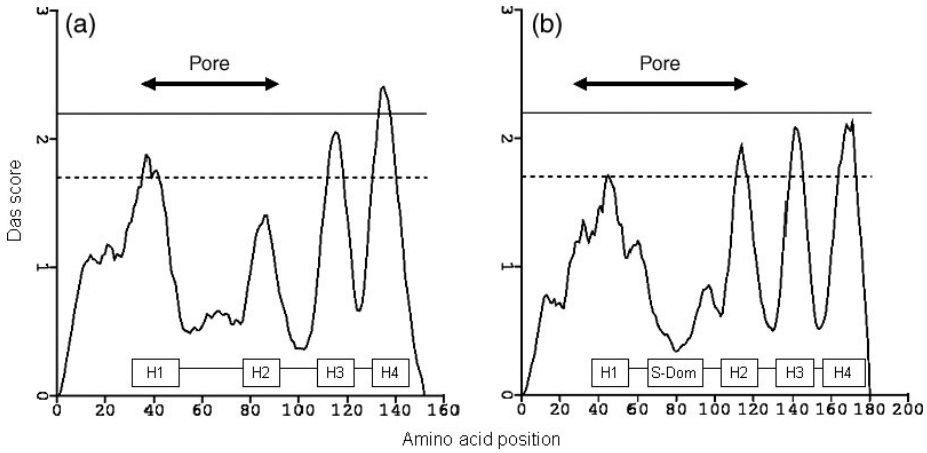

Figure 3. Secondary structure analysis of *Arabidopsis thaliana* OEP16 proteins. The dense alignment surface (DAS) method (Cserzo *et al.*, 1997) was used to predict transmembrane helices in OEP16-L (a) and OEP16-S (b). Amino acid positions corresponding to the transmembrane helical segments (H1–H4) and the interhelical loops identified in pea OEP16 (Linke *et al.*, 2004) are indicated by rectangles and connecting lines, respectively. The H1–H2 region, which contributes to pore formation (Linke *et al.*, 2004), is indicated by a double-headed arrow. S-dom denotes the S-domain region encoded by the additional exon (Ex-S) in *OEP16-S* homologues. The lower and upper horizontal lines represent the DAS loose and stringent cut-off points, respectively.

the minimal region capable of forming a channel in liposomes, thus defining the pore-forming region of the protein (Steinkamp *et al.*, 2000). More recently (Linke *et al.*, 2004), fluorescence and circular dichroism spectroscopy, hydrophobic cluster analysis and the prediction of transmembrane segments using the dense alignment surface (DAS) method (Cserzo *et al.*, 1997) have shown that OEP16, in common with other members of the TIM protein family, consists of four transmembrane α-helices (H1–H4). Based on these investigations, H1 is proposed to determine substrate specificity and, in collaboration with H2, channel or pore formation (Linke *et al.*, 2004).

We used the DAS method to compare the hydrophobicity profiles of *Arabidopsis thaliana* OEP16 proteins that lack and contain the S-domain (Figure 3). The atOEP16-L profile shows four strongly hydrophobic peaks (H1–H4), which correspond to the predicted transmembrane α-helices identified in pea OEP16 (Linke *et al.*, 2004). H1–H4 were also identified in the atOEP16-S profile. However, in this protein, the interhelical domain between H1 and H2 is expanded by more than twofold relative to atOEP16-L, by a region of relatively low hydrophobicity that corresponds to the S-domain (Figure 3). This suggests that the S-domain may fundamentally alter the structure and properties of the OEP16-S pore-forming region.

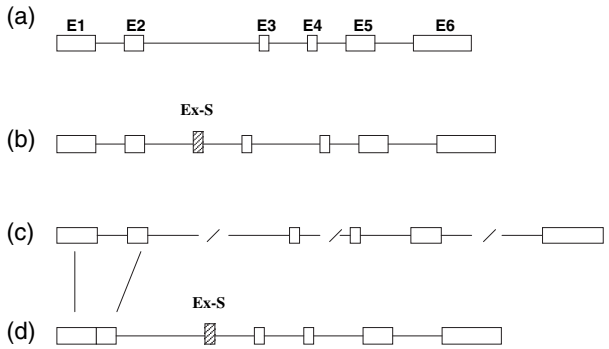

Figure 4. Intron–exon structure of Arabidopsis and rice *OEP16* homologues.
(a) *Arabidopsis thaliana OEP16-L* (At2g28900): expressed sequence tag (EST) accession AY084261 was aligned with the complementary genomic sequence NM128449.
(b) *A. thaliana OEP16-S* (At4g16160): EST accession AY088334 was aligned with the complementary genomic sequence AL161543.
(c) Rice *OEP16-L*: EST accession AU08982, isolated from an 8-day-old *Oryza sativa* (japonica group) green shoot library was aligned with the complementary genomic sequence (AC079356) located on rice chromosome 5.
(d) Rice *OEP16-S*: EST accession BI804632 isolated from an *O. sativa* endosperm cDNA library was aligned with the complementary sequence (AP002871) on chromosome 1. Exons (E1–E6) are denoted by rectangles and are drawn to scale. Horizontal lines indicate introns. In (c), introns were too large to be drawn to scale; hence the inclusion of forward slashes. Ex-S denotes the additional exon in *OEP16-S* homologues that encodes the S-domain.

The S-domain is encoded by an additional exon in Arabidopsis and rice

The amino acid residues immediately flanking the S-domain insertion junctions are well conserved among OEP16-S sequences corresponding to D/E-[G……R/E]-G-D/E, where the residues inside the brackets represent the first and last amino acids of the S-domain (Figure 2). Moreover, in OEP-L and OEP-S proteins that lack the S-domain, the amino acid residues homologous to those flanking the insertion site position also show a high degree of conservation, for example R/K|G-D/S in angiosperms (the vertical line represents the insertion site). This prompted an investigation of the exon–intron structure of *OEP16-L* and *-S* genes in Arabidopsis and rice (Figure 4). In both species, the *OEP16-L* genes comprise six exons (E1–E6), and although four of the

introns in rice are several-fold larger than in Arabidopsis, the coding capacity of the individual exons and the intron/exon junctions are conserved. The *OEP16-S* genes also contain E1–E6, although in rice E1 and E2 are fused, but differ from *OEP16-L* genes in that they contain an additional exon (Ex-S), which encodes the S-domain.

Surprisingly, we found two *OEP16-S* ESTs from *Citrus sinensis* (orange), with identical sequences except that one form contained Ex-S whereas the other lacked it, which suggests that the *OEP16-S* transcript in *Citrus* is alternatively spliced. The *Citrus* EST containing Ex-S was derived from a seed cDNA library (CX050961), whereas the EST without the exon was from a flower cDNA library (CX064493). We did not, however, find EST evidence for alternative splicing of *OEP16-S* transcripts in any of the other species shown in Figure 1.

AtOEP16-L *is expressed predominantly in leaves and is cold inducible*

An RNA gel blot analysis of total RNA extracted from flowers, roots, leaves and green siliques of mature Arabidopsis plants grown under greenhouse conditions indicated that *atOEP16-L* transcripts accumulated at high levels in leaves (Figure 5a). Because the *OEP16-L* orthologue in barley (*cor-tmc-ap3*) was shown to be cold regulated (Baldi *et al.*, 1999), we investigated whether *atOEP16-L* might be similarly regulated and, additionally, whether other abiotic stresses (dehydration and hyperosmotic stress) and mediators of biotic (salicylic acid; SA) and abiotic (ABA) stress might also influence *atOEP16-L* expression. Three-week-old seedlings showed a significant increase (approximately two- to four-fold) in transcript accumulation in response to SA and mannitol, but not in response to dehydration (Figure 5b). In addition, although *atOEP16-L* expression was induced by exogenous ABA, the methanol control treatment also resulted in significant induction. By far the greatest increase in *atOEP16-L* transcript levels (approximately 20-fold) occurred in response to a cold (4°C) treatment. The data suggest that *atOEP16-L* may play an adaptive role not only in the cold stress response but also in the general stress response pathway.

AtOEP16-S *is expressed predominantly in seeds*

The RNA gel blot analysis revealed that *atOEP16-S* mRNA accumulated at a high level in developing siliques, at a low level in flowers and was absent in leaves and stem tissue (Figure 5c). Because of the high-level accumulation of transcripts in siliques, we investigated the possibility that *atOEP16-S* transcription might be regulated by the ABA-insensitive loci *ABI3* and *ABI5*, which encode transcription factors of the B3 domain and basic leucine zipper (bZIP) families, respectively, that function as key regulators of seed development (Finkelstein *et al.*, 2002). Total mRNA was isolated from siliques of wild-type plants (ecotypes *Ler* and

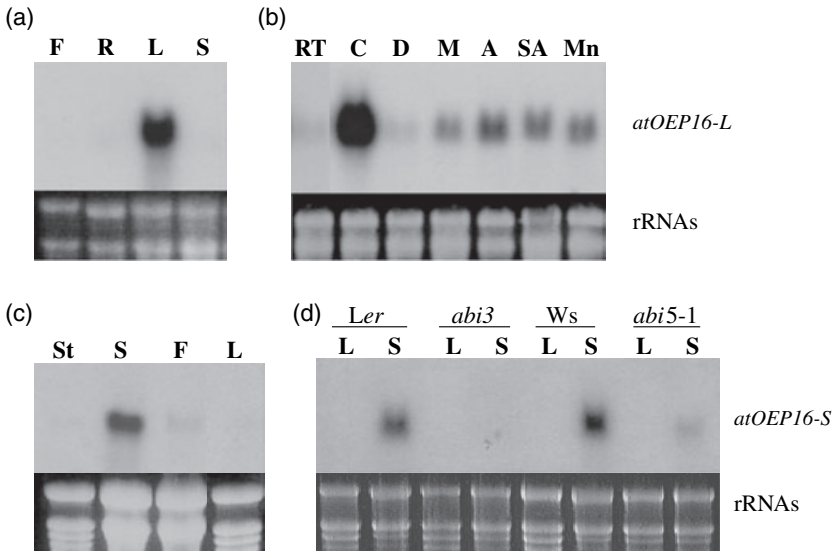

Figure 5. An RNA gel blot analysis of *atOEP16* expression.
(a) Total RNA samples from flowers (F), roots (R), leaves (L) and siliques (S) of mature greenhouse-grown *Arabidopsis thaliana* ecotype Columbia plants.
(b) Total RNA samples from 2-week-old *A. thaliana* ecotype Columbia seedlings, grown in sterile culture, which received the following treatments: room temperature, 20°C (RT); cold, 4°C (C); dehydration (D); methanol (M); ABA (A); salicylic acid (SA); mannitol (Mn). RNA gel blots (a) and (b) were hybridized with a radiolabelled *atOEP16-L* cDNA probe.
(c) Total RNA samples from stems (St), siliques (S), flowers (F) and leaves (L) of mature *A. thaliana* ecotype Columbia plants.
(d) Total RNA samples from leaves (L) and siliques (S) of *A. thaliana* ecotype Landsberg *erecta* (L*er*), the *ABA-insensitive3* mutant (*abi3*) in the L *er* background, *A. thaliana* ecotype Wassilewskija (Ws) and the *ABA-insensitive5-1* mutant (*abi5-1*) in the Ws background. RNA gel blots (c) and (d) were hybridized with a radiolabelled *atOEP16-S* cDNA probe. The lower panels show ethidium bromide-stained rRNAs.

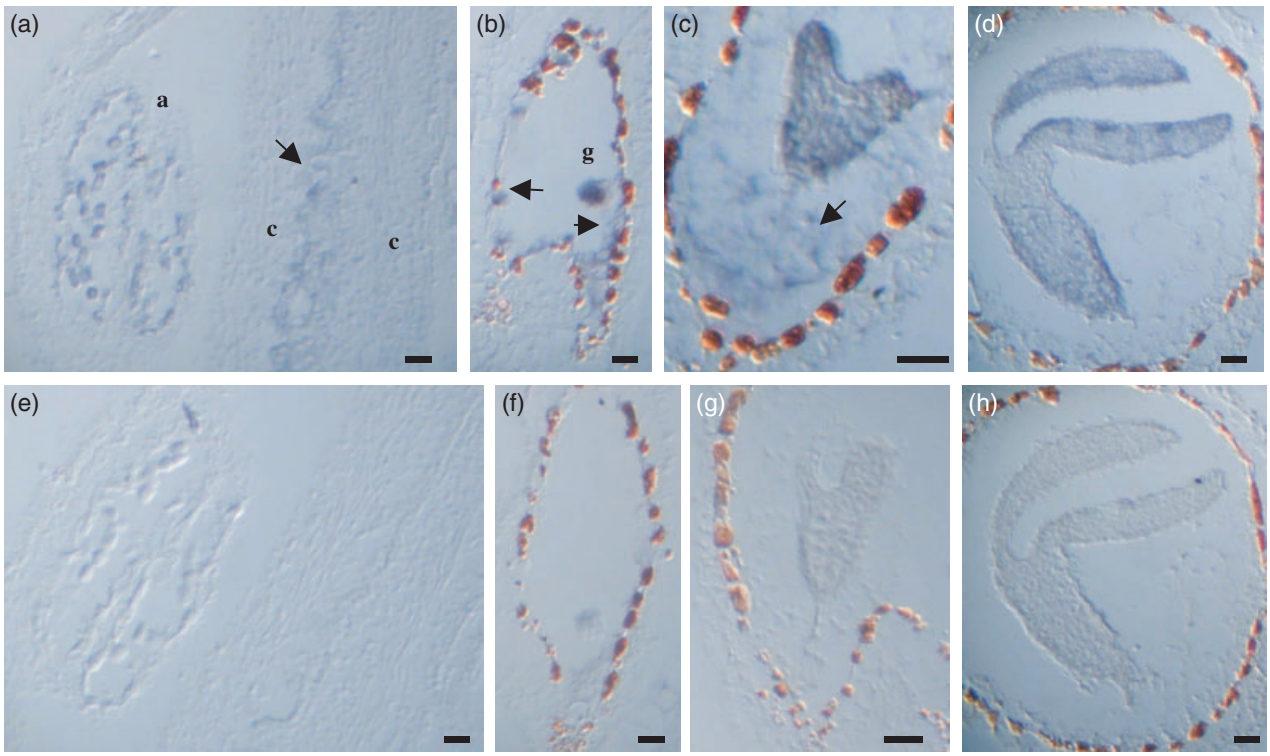

Figure 6. Localization of *OEP16-S* mRNA in *Arabidopsis thaliana* flowers and embryos by *in situ* hybridization.
(a) A section through anther (a) and carpel (c) shows expression in the pollen grains and the developing ovules (arrow).
(b) Expression in the globular embryo (g) but not in the suspensor cells (rightward-pointing arrow); a signal was also detected in the endosperm nuclei (leftward-pointing arrow).
(c) Expression throughout the heart-stage embryo and in the surrounding endosperm (arrow).
(d) Expression throughout the walking stick embryo.
(e)–(h) Corresponding sections hybridized with an *atOEP16-S* sense control probe. Scale bar = 25 μm.

Ws) and the mutants *abi3-1* and *abi5-1* and analysed for the presence of *OEP16-S* transcripts. In contrast to wild-type siliques, *atOEP16-S* transcripts were undetectable in *abi3-1* siliques and accumulated at very low levels in siliques of the *abi5-1* mutant (Figure 5d). In addition, *atOEP16-S* was not induced by abiotic stress treatments or exogenously applied ABA (data not shown), treatments that induced the expression of *atOEP16-L* (Figure 5b). This suggests that the expression of *atOEP16-S* is regulated in siliques by an ABA signalling pathway that requires the transcription factors ABI3 and ABI5.

The localization of *atOEP16-S* transcripts in developing flowers and siliques was investigated by *in situ* hybridization with sense and antisense probes. Sections through the anthers and carpels of immature flowers showed expression in developing pollen grains and also in young developing ovules (Figure 6). Following fertilization, transcripts were detected in the globular embryo (although not in the attached suspensor cells), in heart-stage embryos and in more advanced walking stick embryos (Figure 6). The hybridization signal was also detected in the developing endosperm. Genome-wide transcriptional profiling (Nakabayashi *et al.*, 2005) has recently shown that *OEP16-S* mRNA is one of the more abundant stored mRNA species in seeds of Arabidopsis.

The promoter region of atOEP16-S *(but not* atOEP16-L*) contains multiple G-box ABA-responsive elements*

An intergenic region of 410 bp separates the translational start codon (ATG) of *atOEP16-S* from the start codon of the adjacent upstream gene, *ptlpd2* (Drea *et al.*, 2001). Thus, all relevant *atOEP16-S* promoter and upstream regulatory sequences are probably located within this short intergenic region. A search for putative promoter *cis*-acting regulatory elements was carried out by interrogating the PLACE (plant *cis*-acting elements) database (Higo *et al.*, 1998). This analysis identified four ACGT-core elements belonging to the G-box family (5′CACGTG3′) of ABA-responsive elements (ABREs) (Busk and Pages, 1998; Nakabayashi *et al.*, 2005) (Figure 7). G-box ABREs mediate ABA-induced gene expression by binding members of the bZIP transcription factor family, for example ABI5, and are found in the promoters of many genes expressed during seed maturation

(a)

```
TGTGTACGAGAGAGTGAATGAGAGATCTGTGTAGAGAAAAAGGGAG   46

TGACGGAGATGACAAGTCTGTACTAGTCTCACTCGACGAGCGGAGG   92
            D1                  G1
CGGAGATATCAAAGCGGAGATGACACGTGTCACGTTTCTCGTCTTT  138

AAACCGGAGATTTCAAAATTCCCGGCTATAATATCCAACTCTGAGC  184

CCATTGGGCCTTTTATTTCTCGGTCTGCCTGACGTTATCTTGTGAG  230
      G2
AGCCACGTGGGGAAGGAGAACGGCGAACAGTTCAAACGGCAGACTC  276
      G3          D2          G4          D3
CGACACGTGTGGAAAAGCAGAACACGTGTCAAAAGAACGCTGTTCT  322

GTTTCACAATCTTCTCCTTACTTGTTGTTGAAGAGAGAAGTATTAA  368
      LP        *
CAGAGAAAGAGAGAAGCAACAAGTGAAGAAAGAAAGAAAAAAATG   413
```

(b)

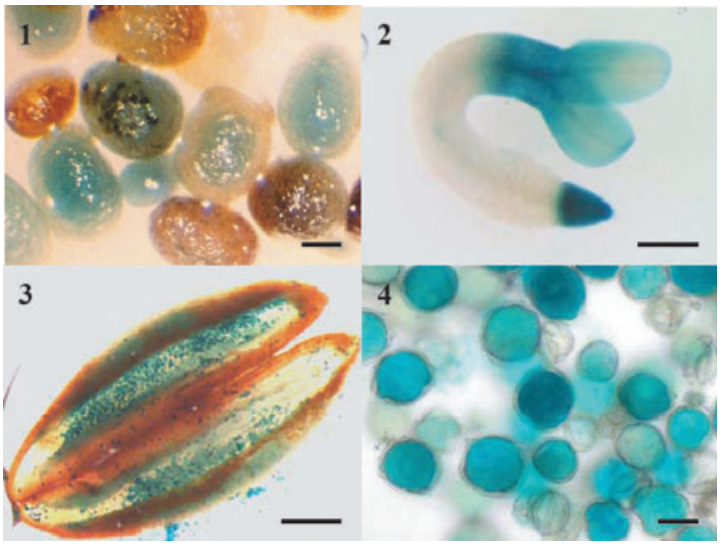

Figure 7. *AtOEP16-S* promoter *cis*-elements and localization of GUS activity in *OEP16-S–GUS* transgenic plants.
(a) The sequence of the 410 bp promoter region between *OEP16-S* and the adjacent upstream gene (At4g16155). Four G-box ABA-responsive element (ABRE) motifs are shown in bold, underlined and numbered G1–G4. Three DNA binding with one finger (Dof) binding elements are underlined and numbered D1–D3. A single late pollen (LP) element is underlined. The putative transcriptional start site, corresponding to the first nucleotide (G) in expressed sequence tag (EST) accession Z32604, is indicated by an asterisk. The *OEP16-S* translation initiation codon (ATG) is underlined and in bold.
(b) Histochemical detection of GUS activity directed by the *atOEP16-S* promoter in transgenic *Nicotiana tabacum* plants. GUS expression in: (1) embryo and endosperm of maturing seeds; (2) hypocotyl and root apex of a 4-day-old seedling; (3) a mature anther; and (4) pollen grains at dehiscence. Scale bars in (2) and (3) = 1 mm; in (1) = 400 μm; and in (4) = 50 μm.

and in genes whose expression is induced by dehydration (Nakabayashi *et al.*, 2005). G-box ABREs were notably absent in the promoter region of *atOEP16-L* (data not shown).

The *atOEP16-S* promoter region also contains three DNA binding with one finger (Dof) domain-binding elements (5′AAAG3′) and a single late pollen (LP) element (5′AGAAA3′). Dof domain proteins regulate many aspects of plant development, including endosperm-specific gene expression and seed germination (Yanagisawa, 2004), whereas the LP element has been shown to direct gene expression primarily during pollen maturation (Bate and Twell, 1998).

The atOEP16-S *promoter confers developmentally regulated seed- and pollen-specific GUS expression in tobacco*

An *atOEP16-S* promoter-*GUS* gene fusion, pOEP–GUS1, was introduced into tobacco in order to investigate temporal and spatial regulation of gene expression. Five independent transgenic lines containing the pOEP–GUS1 gene fusion were analysed for GUS expression during flower and seed development. Mature seeds contained the highest level of GUS activity [up to 2470 pmol methylumbelliferone (MU) μg^{-1} protein h^{-1}]. Anthers and petals contained approximately 10 and 1%, respectively, of the GUS activity found in seeds, whereas GUS activity was virtually undetectable in leaves, stems, roots, stigmas and unfertilized ovaries (data not shown). Histochemical staining further revealed that GUS activity in anthers was localized in pollen grains, whereas in seeds most of the activity was localized in the embryo (Figure 7), consistent with the *atOEP16-S* mRNA distribution pattern revealed by *in situ* hybridization in Arabidopsis and with the known functions of the putative *cis*-elements identified in the *atOEP16-S* promoter region. Finally, GUS activity was also detected in very young seedlings, where it was localized in the cotyledons, cotyledon–hypocotyl junction and root tips.

Because the intensity of GUS histochemical staining observed in developing pollen grains increased during the course of flower maturation, a quantitative analysis of GUS activity was carried out using anthers from four developmental stages (Twell *et al.*, 1993). This revealed an approximately 40-fold increase in GUS activity during pollen development, with maximal activity observed in mature pollen at anther dehiscence (Figure 8). Similarly, because mature seeds contained the highest levels of GUS activity and young seedling leaves contained little, the duration of GUS expression following germination was also investigated. Seedlings were germinated *in vitro* and total protein extracts subjected to quantitative GUS assays at 3-day intervals. This revealed a rapid decline in GUS activity, such that by day 9, less than 5% of the activity found in seeds remained (Figure 8). Thus, the *atOEP16-S* promoter appears to be most active during the maturation phase in both pollen and seed, during which desiccation tolerance is acquired.

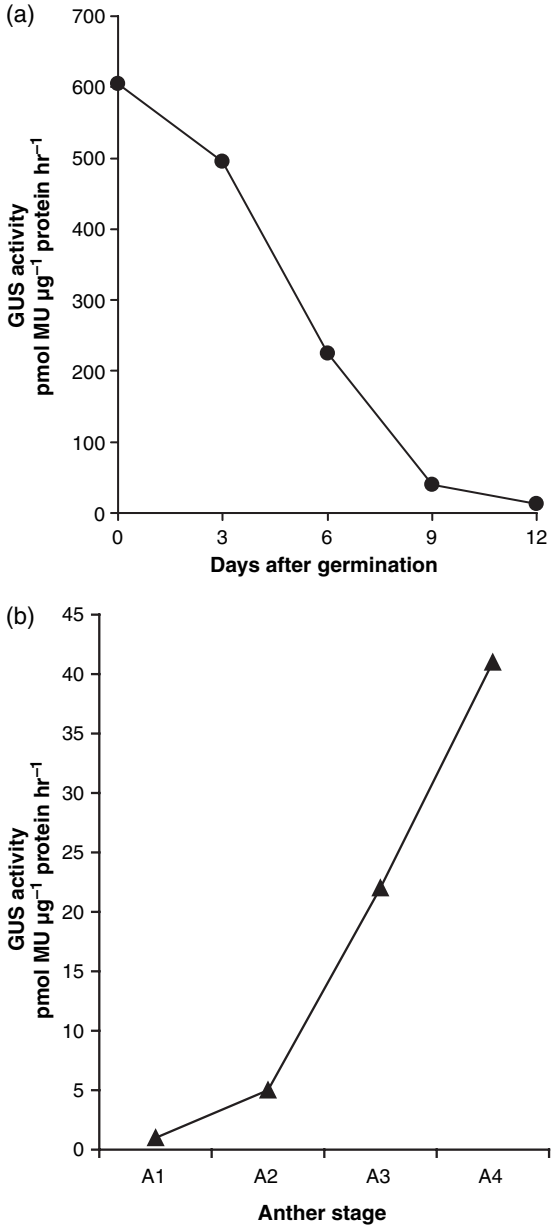

Figure 8. GUS expression directed by the *atOEP16-S* promoter during seedling and pollen development.
(a) GUS activity in seedlings of a transgenic tobacco line containing the *OEP16-S–GUS* transgene. Activity was measured at 3-day intervals post-germination.
(b) GUS activity in anthers during the four stages of development (A1–A4, where A4 is mature dehiscent anthers) in the same transgenic line. GUS activity is shown as pmol MU μg^{-1} protein h^{-1}.

Discussion

The most striking difference between OEP16-L and -S proteins is the presence of the S-domain in the OEP16-S proteins of seed plants. The more ancient OEP16-S proteins found in mosses and ferns uniquely lack this domain, whereas angiosperms encode only S domain-containing proteins. Surprisingly, conifers encode both OEP16-S types (as well as OEP16-L) and appear to have acquired a gene encoding a dicot-like OEP16-S protein by horizontal transfer. Although the S-domain shows considerable variation in both length and amino acid sequence, its hydrophobicity profile is remarkably conserved between species (data not shown), suggesting a high degree of functional conservation. By virtue of its location, the S-domain expands the loop region between transmembrane helices H1 and H2 (Linke *et al.*, 2004). Because these helices are probably involved in substrate recognition (H1) and pore formation (H1 + H2) (Linke *et al.*, 2004; Steinkamp *et al.*, 2000), the presence of the S-domain may alter the import characteristics of OEP16-S proteins (relative to OEP16-L and OEP16-S sequences lacking the S-domain), for example by altering the specificity of substrate recognition, the biophysical properties of the pore or interactions with other import complex components, for example Toc33, as proposed by Reinbothe *et al.* (2005).

Until recently, plastid import processes were thought to be largely conserved between different plastid and tissue types and across species. This appears to be the case for certain core components of import complexes, for example Toc75 and Tic110 (Dávila-Aponte *et al.*, 2003). However, recent investigations have not only revealed significant functional specialization among homologous components, but also highlighted the importance of developmental context in determining the specific composition and properties of import complexes, as originally proposed by Dahlin and Cline (1991). Arabidopsis contains two paralogous genes coding for Toc33/34 isoforms, *atToc33* and *atToc34*, of particular relevance in light of the recently proposed interaction between Toc33 and OEP16 in the import of PORA (Reinbothe *et al.*, 2004a,b, 2005). Both genes are developmentally regulated, with the highest expression levels in young leaves and infloresences, but *atToc33* is expressed at

levels several-fold higher than *atToc34* (Jarvis *et al.*, 1998; Kubis *et al.*, 2003). Knockout mutants of *atToc33* (*ppi1*; Gutensohn *et al.*, 2000; Jarvis *et al.*, 1998) and *atToc34* (*ppi3*; Constan *et al.*, 2004) indicate a substantial but not complete overlap in function. The *ppi1* mutant displays a chlorotic phenotype, whereas *ppi3* appears phenotypically normal, indicating that Toc33 can partially compensate for the lack of Toc34 (Constan *et al.*, 2004). Most significantly, however, an embryonic-lethal phenotype was observed in crosses that combined the *ppi1* and the *ppi3* homozygous knockout mutations, indicating an essential role for *atToc33* and/or *atToc34* during seed development. Because OEP16-L has been reported to interact specifically with Toc33, but not with Toc34 (Reinbothe *et al.*, 2005), it would be interesting to determine whether OEP16-S interacts with Toc33, Toc34 or indeed with other receptors and import channel components during embryo and/or seed development.

Neofunctionalization within the *OEP16-L* and *-S* gene subgroups is also suggested by their contrasting gene expression patterns. *atOEP16-L*, like its orthologue *cor-tmc-ap3* in barley (Baldi *et al.*, 1999), is expressed constitutively at a low level, predominantly in leaves, but is strongly induced by low-temperature stress. The upregulation of *cor-tmc-ap3* correlates positively with a cultivar-specific ability to adapt to low temperatures (cold hardiness): the lower the cold-hardening temperature, the higher the amount of *cor-tmc-ap3* mRNA. Significantly, the increase in gene transcripts in response to low temperatures parallels a corresponding increase in the levels of the encoded protein. Moreover, a barley albino mutant that does not show cold-induced upregulation of *cor-tmc-ap3* was unable to adapt to low temperatures (Baldi *et al.*, 1999; Dal Bosco *et al.*, 2003). There is considerable regulatory cross-talk between the various abiotic stress (i.e. cold, dehydration, and osmotic stress) response pathways. Thus, some cold-inducible genes may also be induced by dehydration or by the application of ABA (Yamaguchi-Shinozaki and Shinozaki, 2005). However, neither dehydration nor ABA was found to affect the expression of *cor-tmc-ap3* (Baldi *et al.*, 1999). Our data show that this was also the case for *atOEP16-L* expression. However, a search of the *atOEP16-L* promoter region for *cis*-acting elements known to direct cold-inducible gene expression, for example DRE/CRT/LTRE elements (Yamaguchi-Shinozaki and Shinozaki, 2005), failed to identify any such elements (data not shown). This suggests that some as yet unidentified cold response elements may be located in the *atOEP16-L* upstream region.

atOEP16-S, the expression of which essentially replaces that of *OEP16-L* during seed development, is clearly regulated by an ABA signalling pathway, as shown by its reduced expression in *abi* mutant backgrounds. Moreover, its strong expression in developing seeds can be correlated with the presence, in the promoter region, of multiple G box-type ABREs (Busk and Pages, 1998), which are not found upstream of *OEP16-L* (data not shown). A recent transcriptomics analysis of stored mRNAs in seeds of Arabidopsis identified G box-related sequences as the most common *cis*-element in the promoters of genes that are highly expressed in seeds (Nakabayashi *et al.*, 2005). Moreover, gene expression levels correlated positively with ABRE copy number; genes containing three or more ABREs, such as *atOEP16-S*, were among the most highly expressed. Our RNA gel blot, *in situ* hybridization and promoter GUS analyses indicate high-level expression of *OEP16-S* during the maturation/desiccation phase of seed development, consistent with the transcriptomics data of Nakabayashi *et al.* (2005), which shows that *atOEP16-S* belongs to the gene class that gives rise to the most abundant stored mRNAs in dry seeds (approximately 500 genes of 22 000 assayed). The rapid decline in GUS activity observed soon after germination of seeds of the OEP16-S–GUS line is also consistent with the rapid decrease in *atOEP16-S* transcript levels reported by Nakabayashi *et al.* (2005) following imbibition of Arabidopsis seeds, which in turn correlates with rapidly declining levels of ABA. Furthermore, on the basis that the *OEP16-S* ESTs identified in *Picea* and in both monocot and dicot angiosperm species were typically isolated from flower-, seed-, endosperm-, or caryopsis-specific cDNA libraries, it is probable that the ABA-dependent, seed-specific regulation shown by a*tOEP16-S* is conserved in these homologues.

The unique association of S domain-containing OEP16-S proteins with seed-bearing plants and the *atOEP16-S* gene expression data reported here, suggest a possible role during reproduction. Unlike reproduction in seed-bearing plants, reproduction in mosses and ferns, each of which produce motile male gametes, is heavily dependent on external water. The evolution of mechanisms that freed the process of male gametophyte and embryo development from the necessity of external water revolutionized the plant lifecycle (Steeves, 1983). In angiosperms and conifers, seed development proceeds through an initial phase of embryogenesis and embryo growth to a maturation phase characterized by growth arrest, the accumulation of lipid and protein reserves, the induction of dormancy and, crucially, the acquisition of desiccation tolerance followed by extensive dehydration (Goldberg *et al.*, 1994). This maturation phase is ABA dependent, as evidenced by the fact that Arabidopsis mutants containing severe alleles of ABA biosynthetic and response loci produce 'green' seeds that fail to mature correctly and are lethally desiccation sensitive under conditions of ambient humidity (Ooms *et al.*, 1993). Pollen development likewise involves a late maturation phase marked by the acquisition of desiccation tolerance and dehydration (McCormick, 1993). In contrast, both mosses and ferns lack equivalent embryo (sporophyte) and gametophyte maturation phases and reproduction is notably desiccation sensitive. Consistent with a possible role for

S domain-containing OEP16S proteins in the acquisition of desiccation tolerance during reproductive development, *atOEP16-S* expression levels were highest during the late maturation phases in both pollen and seed.

Plant *OEP16* genes provide an example of a duplication event that was followed by neofunctionalization (Force *et al.*, 1999). An early gene duplication in the ancestor of all land plants gave rise to *OEP16-S* and *OEP16-L*. One of the daughter genes, *OEP16-S*, later gained a new seed-specific function. This neofunctionalization did not occur immediately after the duplication, and so could not have been the immediate reason for the retention of two copies of the gene, but occurred in the progenitor of angiosperms after other lineages (bryophytes, seed ferns and probably gymnosperms) had already branched off. The new function of *OEP16-S* involves expression in a new tissue (seed) that did not exist at the time the gene duplication occurred, and a correlated change in the secondary structure of the protein (and hence possibly the pore region) was achieved by gaining an exon. Although exon shuffling has been reported in plants (Kubo *et al.*, 1999; Long *et al.*, 1996), we are not aware of any plant genes other than *OEP16-S* where the structure of an orthologue differs between species due to the apparent gain of a new exon internal to the gene. Some examples are known in mammals, and in most of these cases the new exon begins life as a variant that is only present in a minority of transcripts through alternative splicing (Cusack and Wolfe, 2005; Modrek and Lee, 2003). In this regard, it is interesting that we found *Citrus OEP16-S* to be alternatively spliced and it may represent an ancestral state that has been lost in other angiosperms, or perhaps occurs in other tissues.

Experimental procedures

Sequence alignment and phylogenetic analysis

OEP16-related sequences were identified by TBLASTN searches against the National Centre for Biotechnology Information, USA (NCBI) GenBank and dbEST databases (Altschul *et al.*, 1990). Predicted protein sequences were aligned using CLUSTALW (Thompson *et al.*, 1994), and phylogenetic trees were constructed using the neighbour-joining method (Saitou and Nei, 1987), ignoring gaps in the alignment and using Poisson correction of distances.

Plant material and stress treatments

Arabidopsis (Columbia) seedlings were grown axenically for 3 weeks at 23°C in a growth room with a 12 h/12 h light/dark cycle. Seedlings were grown on filter paper discs placed in contact with a solid medium comprising 0.7% agar containing 0.5 × MS salts, 0.5 g l^{-1} 4-morpholineethanesulfonic acid (MES), 0.5 × vitamins (Sigma-Aldrich, St. Louis, MO, USA), pH 5.7, and 0.5% (w/v) sucrose. To initiate stress treatments, plantlets were transferred to liquid medium containing either 200 mm mannitol (hyperosmotic stress), 2 mM SA, 10 μM ABA in 0.1% methanol or a 0.1% methanol control. Cold stress was initiated by transferring seedlings to liquid medium at 4°C in a cold room maintained at the same temperature. Dehydration stress was initiated by placing seedlings in an open Petri dish at room temperature. Fresh medium-only controls were conducted in parallel. Whole plantlet samples were taken for RNA extraction at 3 and 27 h time points.

RNA gel blot analyses

Total RNA was prepared from organs and tissues of Arabidopsis, as described in Drea *et al.* (2001). Total RNA samples were electrophoresed on 1% agarose-formaldehyde gels and blotted on to Hybond-N membranes (Amersham, Little Chalfont, Buckinghamshire, UK) by capillary transfer (Sambrook *et al.*, 1989). Radiolabelled double-stranded DNA probes were prepared using the random-priming method with PCR-generated fragments as the template. Northern blot hybridization with ^{32}P-labelled probes and subsequent membrane washing was carried out as described in Sambrook *et al.* (1989). Signal detection was carried out using autoradiography with X-ray film (AGFA, Mortsel, Belgium) and Hyperscreen intensifying screens (Amersham).

Construction of an AtOEP16-S promoter–GUS gene fusion

A 520 bp DNA fragment comprising the entire promoter region located between AtOEP16-S and the adjacent upstream gene, ptlpd2 (Drea *et al.*, 2001), was generated using PCR with *A. thaliana* DNA and the oligonucleotide primers POgus1 (5′-GCT CTA Gaa gct tGG ATG AAC AAA GCG TAC G-3′; the lowercase nucleotides were added to generate a *Hin*dIII site to facilitate cloning) and POgus2 (5′-CGg gat ccA TTA CAA TTC TTC CTC C-3′; the lowercase nucleotides were added to generate a *Bam*HI site to facilitate cloning). In addition to promoter sequences, the amplified fragment also encodes the N-terminal 11 amino acids of AtOEP16-S. This was used to create an in-frame translational fusion with the GUS gene in the vector pBI101 (Jefferson *et al.*, 1987) and was shown by DNA sequencing to be free of PCR-induced mutations and to be in-frame with the GUS gene. The resulting plasmid, pOEP–GUS1, was transferred into *Agrobacterium tumefaciens* LBA4044 by triparental mating and introduced into *Nicotiana tabacum* (var. Samsum NN) (Drea *et al.*, 2001).

GUS activity assays

Quantitative GUS assays were performed using pooled seedling and anther extracts prepared in GUS extraction buffer and the substrate 4-methylumbelliferyl glucuronide (MUG), as described by Jefferson *et al.* (1987). Histochemical visualization of GUS activity in whole seeds and seedlings was performed using the substrate X-Gluc, as described by Jefferson *et al.* (1987).

Localization of OEP16-S transcripts by in situ *hybridization*

Plant material (siliques and flowers), grown under standard greenhouse conditions, was fixed in 4% paraformaldehyde, dehydrated and embedded in paraffin wax. *In situ* hybridization was performed based on the protocol described in Coen *et al.* (1990). Paraffin sections (8 μm) were hybridized with digoxigenin (DIG)-UTP-labelled antisense and sense RNA probes produced by *in vitro* transcription of the cDNA template with T7 and T3 RNA polymerase, respectively. Hybridized probe RNA was detected using anti-DIG–alkaline phosphatase with nitro-blue tetrazolium chloride/5-bromo-4-chloro-3 indolyl phosphate (NBT/BCIP) colorimetric detection.

Images were captured on a Nikon Microphot-SA microscope using Nomarski optics.

Acknowledgements

NTL was supported by the Irish Council for Science Engineering and Technology (IRCSET award SC/2003/38). TK was supported by EU Framework 6 Contract LSHG-CT-2003-503238 (PLASTOMICS). KHW was supported by Science Foundation Ireland.

References

Abdallah, F., Salamini, F. and Leister, D. (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci.* **5**, 141–142.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Baldi, P., Grossi, M., Pecchioni, N., Vale, G. and Cattivelli, L. (1999) High expression of a gene coding for a chloroplastic amino acid-selective channel protein is correlated to cold acclimation in cereals. *Plant Mol. Biol.* **41**, 233–243.

Bate, N. and Twell, D. (1998) Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Mol. Biol.* **37**, 859–869.

Bauer, J., Hiltbrunner, A. and Kessler, F. (2001) Molecular biology of chloroplast biogenesis: gene expression, protein import and intraorganellar sorting. *Cell. Mol. Life Sci.* **58**, 420–433.

Bolter, B. and Soll, J. (2001) Ion channels in the outer membrane of chloroplasts and mitochondria: open door or regulated gates? *EMBO J.* **20**, 935–940.

Busk, P.K. and Pages, M. (1998) Regulation of abscisic acid-induced transcription. *Plant Mol. Biol.* **37**, 425–435.

Coen, E.S., Romero, J.M., Doyle, S., Elliott, R., Murphy, G. and Carpenter, R. (1990) *Floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell*, **63**, 1311–1322.

Constan, D., Patel, R., Keegstra, K. and Jarvis, P. (2004) An outer envelope membrane component of the plastid protein import apparatus plays an essential role in *Arabidopsis*. *Plant J.* **38**, 93–106.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G. and Elofsson, A. (1997) Prediction of location of transmembrane α-helices in prokaryotic membrane proteins: the Dense Alignment Surface method. *Protein Eng.* **10**, 673–676.

Cusack, B.P. and Wolfe, K.H. (2005) Changes in alternative splicing of human and mouse genes are accompanied by faster evolution of constitutive exons. *Mol. Biol. Evol.* **22**, 2198–2208.

Dahlin, C. and Cline, K. (1991) Developmental regulation of the plastid protein import apparatus. *Plant Cell*, **3**, 1131–1140.

Dal Bosco, C., Busconi, M., Govoni, C., Baldi, P., Stanca, A.M., Crosatti, C., Bassi, R. and Cattivelli, L. (2003) *Cor* gene expression in barley mutants affected in chloroplast development and photosynthetic electron transport. *Plant Physiol.* **131**, 793–802.

Dávila-Aponte, J.A., Inoue, K. and Keegstra, K. (2003) Two chloroplastic protein translocation components, Tic110 and Toc75, are conserved in different plastid types from multiple plant species. *Plant Mol. Biol.* **51**, 175–181.

Drea, S.C., Mould, R.M., Hibberd, J.M., Gray, J.C. and Kavanagh, T.A. (2001) Tissue-specific and developmental-specific expression of an *Arabidopsis thaliana* gene encoding the lipoamide dehydrogenase component of the plastid pyruvate dehydrogenase complex. *Plant Mol. Biol.* **46**, 705–715.

Finkelstein, R.R., Gampala, S.S. and Rock, C.D. (2002) Abscisic acid signaling in seeds and seedlings. *Plant Cell*, **14**(Suppl.), S15–S45.

Flugge, U.I. (1998) Metabolite transporters in plastids. *Curr. Opin. Plant Biol.* **1**, 201–206.

Flugge, U.I. (2000) Transport in and out of plastids: does the outer envelope membrane control the flow? *Trends Plant Sci.* **5**, 135–137.

Flugge, U.I. and Benz, R. (1984) Pore forming activity in the outer membrane of the chloroplast envelope. *FEBS Lett.* **169**, 85–89.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531–1545.

Goldberg, R.B., De Paiva, G. and Yadegari, R. (1994) Plant embryogenesis: zygote to seed. *Science*, **266**, 605–614.

Gutensohn, M., Schulz, B., Nicolay, P. and Flugge, U.I. (2000) Functional analysis of the two *Arabidopsis* homologues of Toc34, a component of the chloroplast protein import apparatus. *Plant J.* **23**, 771–783.

Higo, K., Ugawa, T., Iwamoto, M. and Higo, H. (1998) PLACE: a database of plant *cis*-acting regulatory DNA elements. *Nucleic Acids Res.* **26**, 358–359.

Jarvis, P., Chen, L.J., Li, H., Peto, C.A., Fankhauser, C. and Chory, J. (1998) An *Arabidopsis* mutant defective in the plastid general protein import apparatus. *Science*, **282**, 100–103.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Kirk, J.T.O. and Tilney-Basset, R.A.E. (1978) *The Plastids: Their Chemistry, Structure, Growth and Inheritance*. Amsterdam/New York: Elsevier North-Holland Biomedical Press.

Klebba, P.E. and Newton, S.M. (1998) Mechanism of solute transport through outer membrane porins: burning down the house. *Curr. Opin. Microbiol.* **1**, 238–247.

Kubis, S., Baldwin, A., Patel, R., Razzaq, A., Dupree, P., Lilley, K., Kurth, J., Leister, D. and Jarvis, P. (2003) The *Arabidopsis ppi1* mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *Plant Cell*, **15**, 1859–1871.

Kubo, N., Harada, K., Hirai, A. and Kadowaki, K. (1999) A single nuclear transcript encoding mitochondrial *RPS14* and *SDHB* of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl Acad. Sci. USA*, **96**, 9207–9211.

Linke, D., Frank, J., Holzwarth, J.F., Soll, J., Boettcher, C. and Fromme, P. (2000) *In vitro* reconstitution and biophysical characterization of OEP16, an outer envelope pore protein of pea chloroplasts. *Biochemistry*, **39**, 11050–11056.

Linke, D., Frank, J., Pope, M.S., Soll, J., Ilkavets, I., Fromme, P., Burstein, E.A., Reshetnyak, Y.K. and Emelyanenko, V.I. (2004) Folding kinetics and structure of OEP16. *Biophys. J.* **86**, 1479–1487.

Long, M., de Souza, S.J., Rosenberg, C. and Gilbert, W. (1996) Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc. Natl Acad. Sci. USA*, **93**, 7727–7731.

McCormick, S. (1993) Male gametophyte development. *Plant Cell*, **5**, 1265–1275.

Modrek, B. and Lee, C.J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* **34**, 177–180.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. and Nambara, E. (2005) Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination; epigenetic and genetic regulation of transcription in seed. *Plant J.* **41**, 697–709.

Neuhaus, H.E. and Wagner, R. (2000) Solute pores, ion channels and metabolite transporters in the outer and inner envelope

membranes of higher plant plastids. *Biochim. Biophys. Acta*, **1465**, 307–323.

Ooms, J., Leon-Kloosterziel, K.M., Bartels, D., Koornneef, M. and Karssen, C.M. (1993) Acquisition of desiccation tolerance and longevity in seeds of *Arabidopsis thaliana* (a comparative study using abscisic acid-insensitive abi3 mutants). *Plant Physiol.* **102**, 1185–1191.

Pohlmeyer, K., Soll, J., Steinkamp, T., Hinnah, S. and Wagner, R. (1997) Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane. *Proc. Natl Acad. Sci. USA*, **94**, 9504–9509.

Reinbothe, S., Runge, S., Reinbothe, C., van Cleve, B. and Apel, K. (1995) Substrate-dependent transport of the NADPH:protochlorophyllide oxidoreductase into isolated plastids. *Plant Cell*, **7**, 161–172.

Reinbothe, S., Quigley, F., Gray, J., Schemenewitz, A. and Reinbothe, C. (2004a) Identification of plastid envelope proteins required for import of protochlorophyllide oxidoreductase A into the chloroplast of barley. *Proc. Natl Acad. Sci. USA*, **101**, 2197–2202.

Reinbothe, S., Quigley, F., Springer, A., Schemenewitz, A. and Reinbothe, C. (2004b) The outer plastid envelope protein OEP16: role as precursor translocase in import of protochlorophyllide oxidoreductase A. *Proc. Natl Acad. Sci. USA*, **101**, 2203–2208.

Reinbothe, S., Pollmann, S., Springer, A., James, R.J., Tichtinsky, G. and Reinbothe, C. (2005) A role of Toc33 in the protochlorophyllide-dependent plastid import pathway of NADPH:protochlorophyllide oxidoreductase (POR) A. *Plant J.* **42**, 1–12.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbour, NY, USA: Cold Spring Harbor Laboratory Press.

Soll, J. and Schleiff, E. (2004) Protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* **5**, 198–208.

Soll, J., Bolter, B., Wagner, R. and Hinnah, S.C. (2000) Response: the chloroplast outer envelope: a molecular sieve?. *Trends Plant Sci.* **5**, 137–138.

Steeves, T.A. (1983) The evolution and biological significance of seeds. *Can. J. Bot.* **61**, 3550–3560.

Steinkamp, T., Hill, K., Hinnah, S.C., Wagner, R., Rohl, T., Pohlmeyer, K. and Soll, J. (2000) Identification of the pore-forming region of the outer chloroplast envelope protein OEP16. *J. Biol. Chem.* **275**, 11758–11764.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Thomson, W.W. and Whatley, J.M. (1980) Development of nongreen plastids. *Annu. Rev. Plant Physiol.* **31**, 375–394.

Twell, D., Patel, S., Sorensen, A., Roberts, M., Scott, R., Draper, J. and Foster, G. (1993) Activation and developmental regulation of an Arabidopsis anther-specific promoter in microspores and pollen of *Nicotiana tabacum*. *Sex. Plant Reprod.* **6**, 217–224.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (2005) Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* **10**, 88–94.

Yanagisawa, S. (2004) Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell Physiol.* **45**, 386–391.

