Planta (2006) 223: 618–625
DOI 10.1007/s00425-005-0174-z

RAPID COMMUNICATION

Madoka Gray-Mitsumune · Daniel Philippe Matton

The *Egg apparatus 1* gene from maize is a member of a large gene family found in both monocots and dicots

Received: 3 October 2005 / Accepted: 21 October 2005 / Published online: 9 December 2005

Abstract The maize ZmEA1 protein was recently postulated to be involved in short-range pollen tube guidance from the embryo sac. To date, EA1-like sequences had only been identified in monocot species. Using a more conserved C-terminal motif found in the monocot species, numerous ZmEA1-like sequences were retrieved in EST databases from dicot species, as well as from unannotated genomic sequences of *Arabidopsis thaliana*. RT-PCR analyses were produced for these unannotated genes and showed that these were indeed expressed genes. Further structural and phylogenetic analyses revealed that all members of the EA1-like (*EAL*) gene family shared a conserved 27–29 amino acid motif, termed the EA box near the C-terminal end, and appear to be secretory proteins. Therefore, the EA box proteins defines a new class of small secretory proteins, some of which being possibly involved in pollen tube guidance.

Keywords *Arabidopsis* · Fertilization · Peptide signaling · Pollen tube guidance

Abbreviations *ZmEA1*: *Zea maize* egg apparatus gene 1 · EST: Expressed sequence tag · GFP: Green fluorescent protein

Introduction

Plant reproduction is by far the most sophisticated process of plant life. Pistils can selectively retain, guide and accept pollen from the same or related species, while rejecting pollen from other species or, in the case of self-incompatible species, reject pollen from the same species when it shares identical alleles at the S-locus, thus maintaining the integrity of the species. This process requires multiple levels of communication between the pollen and the pistil (Swanson et al. 2004). After successful adhesion and germination of the pollen on the stigma surface, the pollen tube follows guidance cues provided by the pistil through the style and then finally to the ovule. While the pistil provides the chemical and physical environment to nourish pollen tube growth, it also produces chemotropic agents through the style. The TTS proteins in tobacco (Cheung et al. 1995), gamma-amino butyric acid (GABA) in *Arabidopsis* (Palanivelu et al. 2003) and chemocyanin in lily (Kim et al. 2003) have all been shown to exert an effect on pollen tube growth or guidance. Once the pollen tube arrives at the ovule, it relies on signals produced from the embryo sac for its guidance. Studies using the in vitro fertilization system of *Torenia fournieri* have shown that the synergid cells produce a short-range pollen tube attractant(s) that is species specific (Higashiyama et al. 2003; 2001), although the identity of the attractant is still unknown.

A maize protein produced by the egg apparatus, termed ZmEA1, has recently been postulated as a candidate for a short-range pollen tube attractant (Marton et al. 2005). ZmEA1 is a small protein of 94 amino acids and is secreted from the egg apparatus to the micropylar region of the ovule integument. Half of the transgenic plant lines carrying either an RNAi or antisense RNA construct exhibit reduced fertility. Close examination of these transgenic plants revealed defects in the short-range pollen tube guidance. The potential role of ZmEA1 as a species-specific signal was further supported by the high conservation within the same species and the low conservation between maize and rice. Marton et al. further identified ZmEA1 homologues from several other Gramineae species, including two rice homologues. Since no homologous sequences could be identified from *Arabidopsis* or other dicot species, Marton et al. (2005) proposed that EA1-like sequences

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00425-005-0174-z and is accessible for authorized users.

M. Gray-Mitsumune · D. P. Matton (✉)
Institut de Recherche en Biologie Végétale (IRBV),
Département de Sciences Biologiques, Université de Montréal,
4101 rue Sherbrooke est, H1X 2B2 Montréal, QC, Canada
E-mail: dp.matton@umontreal.ca
Tel.: +1-514-872-3967
Fax: +1-514-872-9406

may be specific to Gramineae. The absence of EA1-like sequences from the *Arabidopsis* genome was also noted by a recent commentary by McCormick and Yang (2005). In this commentary, the authors also point out several weaknesses in the Marton et al. (2005) study, the main one being the absence of correlative data between the phenotype observed and the expression level of the *ZmEA1* transcripts in transgenic lines. Therefore, the definitive proof for ZmEA1 as a pollen tube attractant is yet to be presented and will require a thorough biochemical analysis of the ZmEA1 protein as a *bona fide* pollen tube attractant. Nevertheless, the protein encoded by the *ZmEA1* gene has arisen as a candidate for a short range pollen tube attractant molecule.

Recently, we have constructed an EST library from *Solanum chacoense* ovules in order to identify signaling components involved in fertilization and early seed development (Germain et al. 2005). Since EA1-like sequences had only been found in monocot species, we were surprised to find three EA1-like sequences among our ovule EST library. In addition, several EA1-like sequences were found in unannotated regions of the *Arabidopsis* genome. To determine the authenticity of our sequence search, we conducted a survey of EA1-like sequences in monocot and dicot sequence databases. Here, we report the phylogenetic and structural characterization of EA1-like (EAL) sequences. We show that *ZmEA1* is a member of a large gene family that shares a highly conserved C-terminal domain, the EA box found in both monocot and dicot species. Subfamily members that shared a similar protein structure with ZmEA1 were only found in monocot species, and were expressed in various tissues. Other subfamily members that exhibited a slightly different protein architecture also showed a wide range of expression patterns, implying a broader biological function for EA1-like proteins in plants.

Materials and methods

Sequence retrieval and annotation

EA1-like sequences were identified via TBLASTN search of WU-BLAST2 program (*E* threshold = 1,000, Cutoff score = 30, protein weight matrix = BLOSUM62), using either the full length ZmEA1 amino acid sequence or the conserved motif "MKAPGxxGxxISRxxFxANPxLYFxxLRTxG". To identify *EAL* ESTs, WU-BLAST2 sequence searches were conducted between June 2005 and August 2005 against the Higher Plant EST database from the *Arabidopsis* Information Resource (TAIR; http://www.arabidopsis.org/wublast/index2.jsp). EST hits derived from same genes were identified by subjecting each EST nucleotide sequence to BLASTN search. Sequences exhibiting the nucleotide sequence identity of 98% or more after the removal of low quality sequences were clustered and treated as a UniGene cluster and were given arbitrary nomenclatures. Several databases were also consulted for the determination of UniGenes (GenBank, http://www.ncbi.nlm.nih.gov; SOL Genomics Network http://www.sgn.cornell.edu; Plant Genome Network, http://pgn.cornell.edu; The Institute for Genomic Research (TIGR), http://www.tigr.org). Similarly, rice EAL sequences were determined via a TBLASTN search against the rice genome (http://tigrblast.tigr.org/euk-blast/index.cgi?project = osa1). For *Arabidopsis EAL* sequences, a TBLASTN search was performed against the AGI Whole Genome database (BAC clones) using WU-BLAST2 program (http://www.arabidopsis.org/wublast/index2.jsp). Coding sequences and predicted peptides sequences of *Arabidopsis* EALs were determined using the GENSCAN program (Burge and Karlin 1997). The complete list of the gene annotations with GenBank references is supplied as a supplementary table (Supplementary Table 1).

Sequence analysis

The sequence alignment of EAL proteins was performed using the Clustal X program (Higgins et al. 1996) using its default parameters. Since sequences at the N-terminal end were highly divergent, only the sequences beginning 50 amino acids upstream of the EA box were included for the alignment. A phylogenetic tree was generated using the neighbor-joining method using 1,000 bootstrap replicates (Saitou and Nei 1987). Molecular weights and pIs were predicted using the ExPASy Compute pI/MW tool (Gasteiger et al. 2005). Transmembrane domains were predicted using the TMHMM program (Krogh et al. 2001) and the SOSUI program (Hirokawa et al. 1998). The presence of N-terminal signal peptide was predicted using the SignalP 3.0 program (Bendtsen et al. 2004b). The SecretomeP 1.0 program was also used for prediction of protein secretion (Bendtsen et al. 2004a).

RT-PCR analysis

Plant tissues were ground into fine powder in liquid nitrogen. Powdered tissues were resuspended in the following RNA extraction buffer (2% CTAB, 2% PVP, 100 mM Tris–HCl, 25 mM EDTA, 2 M NaCl, 0.5 g l^{-1} spermidine, 2% β-mercaptoethanol, pH8.0) and incubated at 65°C for 5 min (Chang et al. 1993). The solution was mixed with an equal volume of chloroform / isoamyl alcohol (24:1), vortexed, and subjected to a centrifugation at > 10,000 g for 5 min to separate the organic and aqueous phases. Nucleic acids in the aqueous phase were precipitated by adding 1.2 volume of isopropanol, followed by incubation at −20°C for 30 min, and were collected by centrifugation at 17,000 g for 20 min at 4°C. The precipitate was dissolved in RNeasy® RNA extraction buffer (Qiagen, Mississauga, Canada) and RNA was purified using RNeasy® columns according to the manufacturer's recommendation. The purification procedure included the treatment with RNase-free DNase I to remove contaminating DNA.

The final RT reactions contained 50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.5 mM dNTP mix, 0.1 μg μl^{-1} oligo dT_{10}, 0.5 U μl^{-1} RNase inhibitor (Roche Diagnostic, Laval, Canada), 10 U μl^{-1} M-MLV reverse transcriptase (Invitrogen Canada, Burlington, Canada) and 0.1 μg μl^{-1} RNA. The enzyme mixture was added to denatured RNA sample and the tubes were incubated at 37°C for 60 min, followed by 90°C for 5 min. The first strand cDNAs were immediately used for PCR reactions or stored at −20°C. The PCR mixture contained 1/20 vol of cDNA sample, 0.4 μM forward and reverse primers (See Supplementary Table 2 for the sequence information), 200 μM dNTP, 1×Taq buffer (Qiagen) and 0.025 U μl^{-1} HotTaq DNA polymerase (Qiagen). PCR reaction was performed according to the following thermal cycling regime: 1 cycle of 95°C/15 min; varying number of cycles of 94°C/20 s, 60°C/20 s, 72°C/30 s; 1 cycle of 72°C/10 min.

Results and discussion

Identification of EA1-like (EAL) sequences

A TBLASTN search was initially conducted using the full-length ZmEA1 amino acid sequence on an ovule EST library from *Solanum chacoense*, a dicot species, and retrieved three different clones sharing significant sequence similarity with the ZmEA1 deduced protein sequence (Germain et al. 2005). This prompted us to perform a wider search against all plant ESTs, again with the full-length ZmEA1 amino acid sequence. The search results contained many sequences that did not contain a highly conserved motif found in the ZmEA1-like sequences (Marton et al. 2005; McCormick and Yang 2005) but that showed weak sequence similarities with other parts of the protein (mainly hydrophobic regions, Fig. 1b). We thus conducted a TBLASTN search against plant ESTs using the conserved motif of EA1-like sequences "MKAPGxxGxxISRxxFxANPx-LYFxxLRTxG", which was identified from the sequence alignment by McCormick and Yang (2005). Since the query sequence was short and we expected high *e*-values for candidate sequences, the WU-BLAST criteria were changed to less stringent values (*E* threshold = 1,000, Cutoff score = 30). Over 200 ESTs were identified from the Higher Plant EST data set at The Arabidopsis Information Resource (http://www.arabidopsis.org/wublast/index2.jsp). Despite of the high e-values some of the BLAST hits (up to 940), only one EST appeared false positive. Contrary to the two previous studies which found EA1-like (EAL) sequences only in monocot species (Marton et al. 2005; McCormick and Yang 2005), the EST hits were derived from both monocots and dicots (Figs. 1a, 2). This included two species from magnoliids, *Persea americana* and *Liriodendron tulipifera*, and two species from basal Magnoliophyta, *Illicium parviflorum* and *Nuphar advena*. The three *Solanum chacoense* genes from our ovule EST library, termed *ScEAL1 to 3* were also found from this search. The majority of the predicted proteins contained one or more transmembrane domains located upstream of the conserved motif of the *ZmEA1*-like sequences. Although the amino acid sequences outside the motif are highly divergent, one would expect such sequence divergence if the N-terminal domain merely serves as the mean for protein targeting and the conserved motif is cleaved off from the N-terminal transmembrane domain, as suggested by Marton et al. (2005). Precursors of peptide hormones may exhibit low sequence conservation outside of mature bioactive peptides. For example, N-terminal amino acid sequences of the rapid alkalinization factor (RALF) proteins are highly divergent, whereas the C-terminal region that encodes the mature RALF peptide are well conserved (Pearce et al. 2001b). Similarly, rice and *Arabidopsis* phytosulfokine (PSK) precursor proteins exhibit limited sequence conservation outside of mature PSK-α peptides (Yang et al. 2001).

Using a similar approach, we identified five *EAL* sequences in the *Arabidopsis* genome, termed *AtEAL1 to 5*. Two genes in Chromosome 2, *AtEAL2* and *AtEAL3*, are annotated as a single gene with two highly similar exons (At2g30560) according to the version 5.0 of the *Arabidopsis* genome annotation. However, the GENSCAN gene prediction program (Burge and Karlin 1997) predicted two separate genes from the genomic sequence of this region. This was confirmed experimentally using RT-PCR analyses. Primers designed to amplify the two putative exons together failed to give an amplification product (data not shown), while RT-PCR analyses using primers designed to amplify each exon separately gave products of the expected size (Fig. 3a). The three other *AtEAL* genes (*AtEAL1, 4* and *5*) were found in unannotated intergenic regions. This is not surprising since genes coding for small proteins are often missed in whole genome annotations. Wen et al. (2004) reported that out of 21 members of the small protein DVL family, only four were properly annotated (Wen et al. 2004). One of the reasons why the *AtEAL* genes were not fully annotated can be linked to the absence of corresponding ESTs in EST databases. Yet, these genes appear to be transcribed, judging from the significant overlaps between the predicted genes and high-density tilling array signals (Fig. 3b). RT-PCR analyses further confirmed the transcription of all five *AtEAL* genes (Fig. 3a).

The rice genome contains eleven *EAL* sequences, that include gene clusters at chromosome 1, 7 and 12 (Fig. 2). Gene clusters may have resulted from recent gene duplication events, as sequences within clusters were highly conserved. Finally, we identified *PsAD1*, a gene expressed in dormant axillary buds of pea (Madoka and Mori 2000b), as a *ZmEA1* homolog among non-redundant GenBank entries. At this point, *EAL* sequences appear specific to angiosperm species, as no *EAL* sequences were found among any other eukaryotic/prokaryotic genomes or among 327,484 pine ESTs.

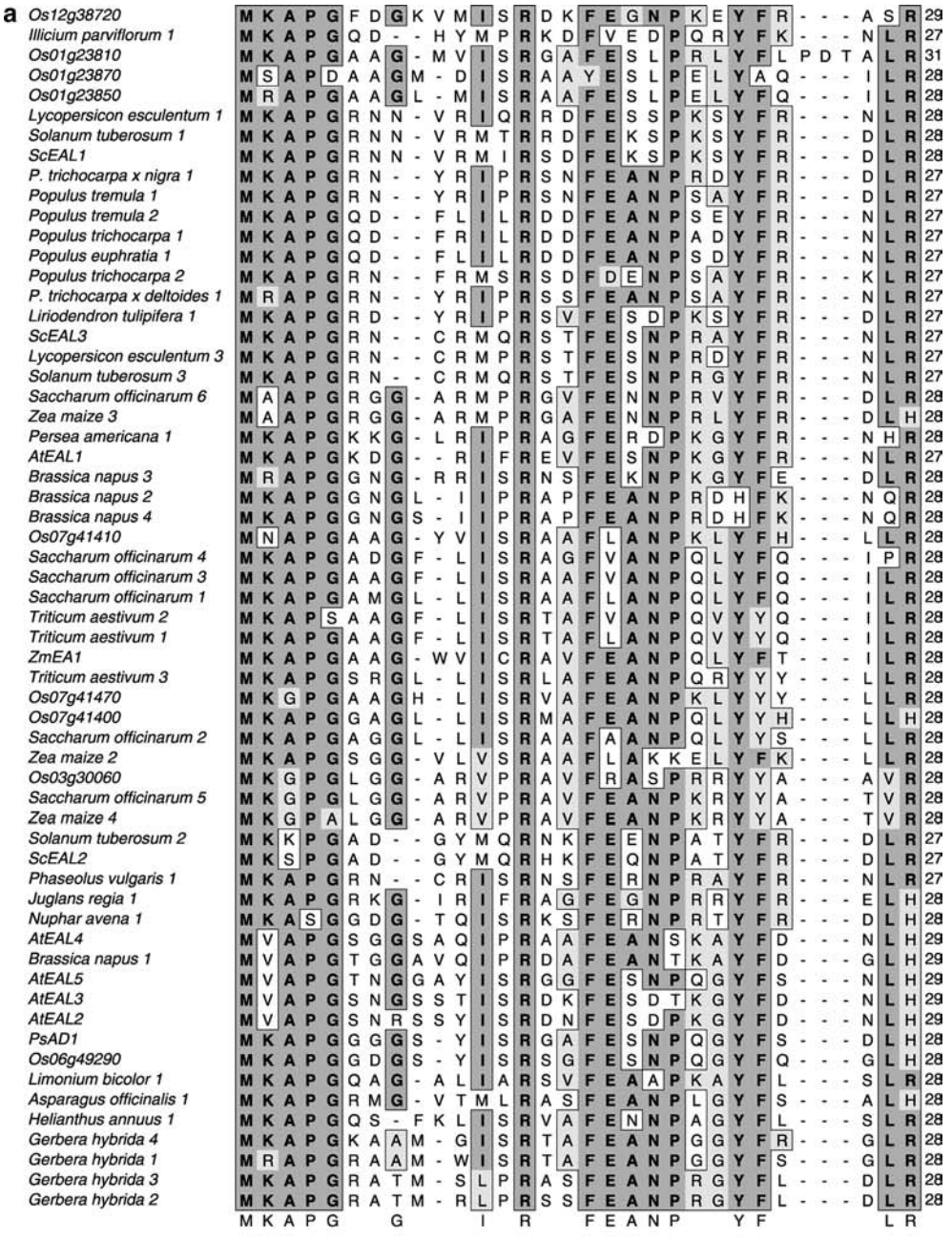

Fig. 1 Structure of EAL proteins. **a.** Sequence alignment of the EA1-like domain (*EA box*). *Arabidopsis* and rice sequences were derived from their genome sequences. Chromosome coordinates of *Arabidopsis* genes are as follows: *AtEAL1*, chr2 C/1135985-1136212; *AtEAL2*, chr2 W/13024424-13024690; *AtEAL3*, chr2 W/13025629-13025930; *AtEAL4*, chr4 C/15984747-15984941; *AtEAL5*, chr5 C/24703987-24704220. Rice genome annotations are according to Yuan et al. (2005). All other sequences were derived from EST databases (See Supplementary Table 1 for the GenBank IDs). The consensus sequence represents residues conserved in ≥51% of the sequences. **b.** Domain organization of the EAL family. *Gray circles* indicate hydrophobic regions

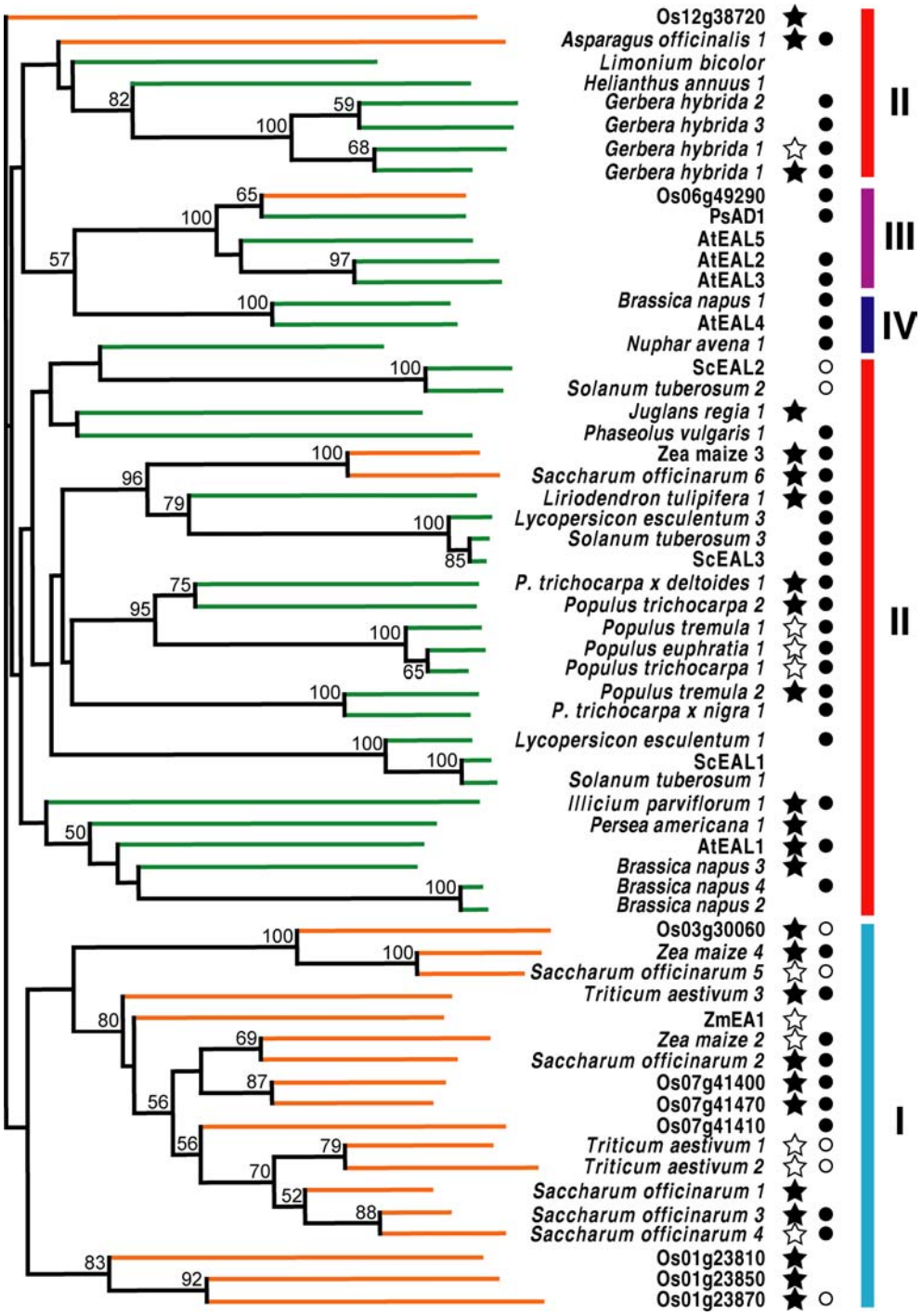

Fig. 2 Phylogenetic tree of the EAL family. Phylogenetic tree generated using the neighbor-joining method (Saitou and Nei 1987). The numbers along the branches are the percentages of the bootstrap support after 1,000 replicates. Only the values higher than 50% are shown. The colors of the branches show either monocots (*orange*) or dicots (*green*). *Symbols* at the right side of gene names indicate the result of secretary protein prediction algorithms. *Filled stars* are the ones predicted to have signal peptides. *Open stars* are those predicted to have signal peptides only when alternative translation initiation sites are used. *Filled circles* are the ones with significant NN-scores (>0.6) in the SecretomeP prediction (Bendtsen et al. 2004a). *Open circles* are the ones that give significant NN-scores only when alternative translation initiation sites are used. The *colored bars* at the right end show the four different classes of EALs as described in Fig. 1b

Structures of EAL proteins

The sequence alignment of the ZmEA1-like motif, termed the EA box thereafter, is shown in Fig. 1b. The EA box consists of 27–29 amino acids in length with the conserved residues (present in ≥51% of the sequences) "MKAPGx_2G$x_{1\text{-}3}$IxRx_2FEANPx_2YFRx_2LR" and lies near C-terminal ends of EAL proteins. EAL proteins exhibit structural similarities in addition to the EA box. They are highly basic with the pI range of 9–11, relatively small (<200 amino acids) and contain one or more hydrophobic regions. About half of EALs are predicted to have a signal peptide at the N-terminal end of the protein (Fig. 2). There are proteins that do not fit these criteria, including ZmEA1. However, these genes often contained multiple potential initiation codons and their profiles fit better to the criteria when downstream AUGs are used. In such cases, the first AUG(s) are often not located in an optimal context for translation initiation, possibly allowing ribosomes to scan further

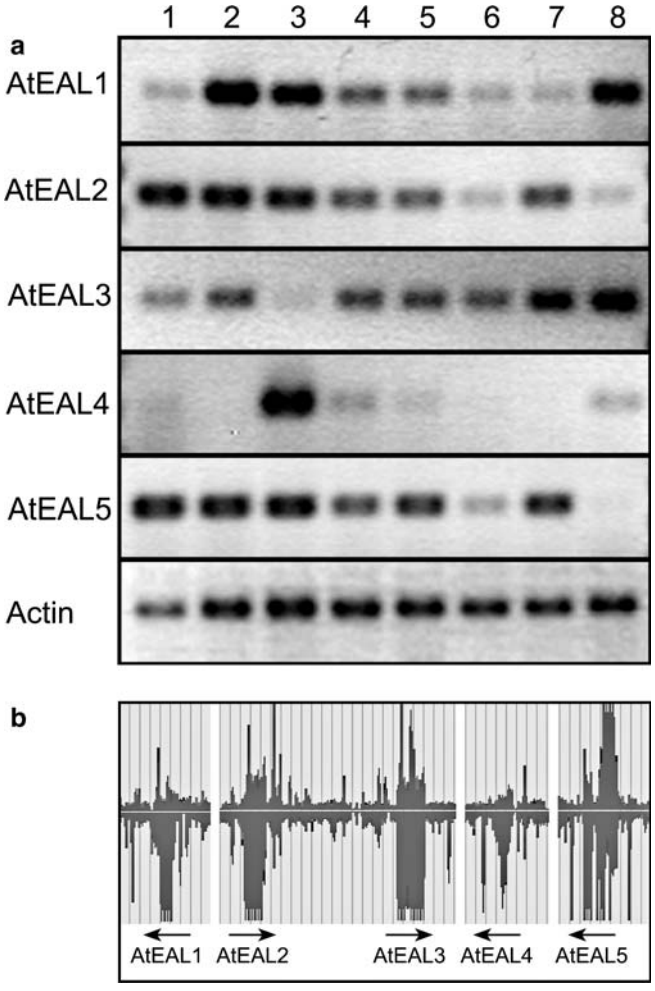

Fig. 3 Expression of Arabidopsis *EAL* genes. **a.** RT-PCR analysis. The inverse images of ethidium bromide stained gels are shown. PCR reactions were performed using either 35 (*AtEAL1, 2, 3, 4* and *5*) or 26 (*Actin 2*) thermal cycles. At least two more different PCR cycle numbers were tested to determine if the reactions were saturated. PCR reactions using RNA alone as template produced no amplification products (data not shown). Tissues used are as follows: *1* flowers before anthesis; *2* flowers after anthesis; *3* young siliques (less than 1 cm in length); *4* inflorescence stem upper half; *5* inflorescence stem, lower half; *6* cauline leaves; *7* rosette leaves; *8* roots. **b.** Transcriptome map of *Arabidopsis EAL* genes. Predicted gene sequences were aligned against the transcriptome map using Arabidopsis Tiling Array Transcriptome Express Tool (http://www.signal.salk.edu/cgi-bin/atta). Zoom-in figures of SALK tilling array signals are shown (Yamada et al. 2003). Alignment and transcriptional orientation of *AtEAL* genes are indicated by *arrows*

downstream (Kozak 2002). For example, *ZmEA1* cDNA contains six in frame potential initiation codons. The predicted pI of ZmEA1 is 5.9 if the first codon is used. However, the protein becomes more basic (pI 8.8) if the third or fourth AUG is used. The context of the third or fourth initiation codon is more similar to the monocot AUG consensus of (A/G)ccAUGG (Joshi et al. 1997) and, therefore, may be the preferred initiation sites. The shift of the initiation codon would also allow the first putative transmembrane domain to become a signal peptide for protein secretion. Similarly, several other proteins are predicted to have signal peptides only if a downstream AUG is used (Fig. 2).

Although only half of EALs contain signal peptides, other EALs may still be secreted into the extracellular space by unknown mechanisms. This was suggested by the high NN-scores of EALs in the SecretomeP prediction program (Bendtsen et al. 2004a) (Fig. 2). The SecretomeP program was designed to identify secretory proteins that lacked a signal peptide in mammals by comparing general protein features instead of specific sequence motifs. These proteins can be secreted despite the lack of a leader sequence, a process referred to as leaderless secretion or non-conventional/classical secretory pathway. The high NN-scores indicate that there is a higher likelihood of EALs localized extracellularly than intracellularly.

EAL proteins can be grouped into four classes, depending on the number and position of hydrophobic regions. Domain organizations of the four classes of EAL proteins are shown in Fig. 1b. Class I EAL proteins contain hydrophobic regions upstream and downstream of the EA box. They contain one or more transmembrane domains in the N-terminal side, which may serve as signal peptides (Fig. 2). The C-terminal hydrophobic tail may also form a transmembrane domain. Class I EAL proteins are exclusively derived from monocot species and are most closely related to ZmEA1. Class II EAL proteins contain hydrophobic regions only on the N-terminal side, about half of which are predicted to be signal peptides (Fig. 2). Class II EALs are of phylogenetically diverse origin and are derived from both monocot and dicot species, including two basal angiosperm species (Fig. 2). Class III and IV EAL proteins have no hydrophobic regions, yet they score high in the SecretomeP prediction algorithm (Fig. 2). Class III EAL proteins contain glycine-rich regions on the N-terminal side and are present in both monocot and dicot proteins. The sequence conservation of Class III EAL extends to the glycine-rich region, where basic residues (arginine, lysine or histidine) appear every 7–10 amino acid intervals, forming repetitive units. The glycine-rich repeat unit also contains one or two alanine and/or serine residues. Finally, Class IV EALs consist of soluble proteins that do not contain any recognizable domains other than the EA-Box.

EALs as mobile peptide signals

Most Class I EALs contain signal peptides and could presumably be exported via the ER/Golgi-dependent secretory pathway. However, the presence of the hydrophobic tail could make it difficult for them to move freely in the extracellular space. Marton et al. (2005) have shown that a ZmEA1-GFP C-terminal fusion protein was secreted into the extracellular space. However, transmembrane domain prediction softwares predict two transmembrane domains within the ZmEA1-GFP fusion protein, one at the N-terminal and

one at the C-terminal end of ZmEA1. Even after the cleavage of the N-terminal transmembrane domain, the processed ZmEA1-GFP fusion should retain the C-terminal transmembrane domain. The presence of this second hydrophobic domain immediately upstream of the GFP reporter construct should instead keep the protein anchored to the plasma membrane, unless the transmembrane domain is cleaved from GFP. If this is the case, the localization of GFP would not correspond to the localization of the mature ZmEA1 peptide. Perhaps, ZmEA1 is also cleaved from the C-terminal transmembrane domain and may be secreted to the extracellular space after all. One would expect peptide hormones to go through an extensive post-translational processing, as observed for the systemin and RALF peptides (McGurl et al. 1992; Pearce et al. 2001a, b). However, this point cannot be concluded from the available data.

Class II–IV EALs, on the other hand, lack the hydrophobic tail, and, therefore, are likely to move freely in extracellular spaces once they are secreted and their N-terminal transmembrane domain is removed. Although not all Class II–IV EALs contain signal peptides, they all appear to be secretory proteins (Fig. 2). An analogous situation exists with the mammalian fibroblast growth factor (FGF) family, key activators of tumor-induced angiogenesis. The majority of FGF members are exported through the ER/Golgi secretory transport. However, FGF-1 and its isoform FGF-2 lack signal peptides and are secreted through an alternative pathway (Nickel 2003). Is this the case for the EAL family? This point will need to be established experimentally.

Class III EALs are particularly interesting because of the high level of conservation between monocots and dicots and because of the glycine-rich domain. A glycine-rich protein AtGRP-3 has been identified as an extracellular interacting partner of the *Arabidopsis* wall-associated receptor kinase 1 (AtWAK1) (Park et al. 2001). Furthermore, a downstream target, the oxygen-evolving enhancer protein 2 (OEE2), of WAK1 becomes phosphorylated in a GRP-3 dependent manner (Yang et al. 2003). These observations suggest that glycine-rich proteins could serve as signaling molecules. A member of Class III EAL, PsAD1, is found abundantly in dormant axillary buds but it disappears upon release of dormancy (Madoka and Mori 2000a). Although the function of PsAD1 has not been determined, it has been proposed that PsAD1 is involved in the maintenance of the dormant state in axillary buds. It would be interesting to see if other members of Class III EALs are also involved in the axillary bud dormancy.

EAL genes are expressed throughout plant development

Figure 3a shows the expression of *Arabidopsis EAL* genes in different tissues as detected by RT-PCR. Four out of five *AtEAL* genes were transcribed throughout plant development. *AtEAL1*, *AtEAL2* and *AtEAL5* have a similar expression pattern, their transcripts being more abundant in reproductive tissues (flowers and young siliques, lane 1 to 3) than in vegetative tissues (stems, leaves and roots). On the other hand, *AtEAL3* transcripts were less abundant in reproductive tissues than vegetative tissues. *AtEAL4* transcripts were almost exclusively found in young siliques (Fig. 3a). The expression patterns of the *Arabidopsis EAL* genes are dissimilar to that of *ZmEA1*, which was exclusively detected in egg apparatus cells and proembryos (Marton et al. 2005). This may be explained by the fact that the egg apparatus occupies only a small fraction of flower tissues. However, one would at least expect decrease in the transcript level in flowers after fertilization. Yet, we observe no such evidence in our RT-PCR analyses (Fig. 3a).

RT-PCR analysis was also performed to detect transcripts levels of *Solanum chacoense EAL* genes. For all three *ScEAL* genes, amplification products were detected from both ovule and leaf RNA samples (data not shown), suggesting that they are transcribed in both reproductive (as expected from their EST origin) and vegetative tissues. This observation is further supported by the fact that EST hits from closely related genes in potato and tomato (*Solanum tuberosum 1-3*, *Lycopersicon esculentum 1* and *3*) were found in both reproductive and vegetative tissues (Supplementary Table I).

Our RT-PCR results are consistent with the previous observation made by McCormick and Yang (2005), in which they noted that EA1-like monocot genes had EST hits in both reproductive and vegetative tissues. Similarly, the EST hits from other classes of *EAL* genes (Figs. 1b, 2) were derived from both reproductive and vegetative tissues (Supplementary Table 1). The presence of *EAL* transcripts in various plant tissues suggests that each EAL may have a unique function during plant growth and development. However, expression of EALs in vegetative tissues does not definitively exclude the possibility that some of these peptides may also play roles in pollen tube guidance. Studies in animals have shown that stimulation via growth factors can lead to different developmental outcomes depending on the cells that receive signals. For example, transforming growth factor β1 (TGF-β1) can induce variety of cellular responses such apoptosis, cell survival and cell differentiation, depending on cell types (Sanchez-Capelo 2005). Neurotrophins such as nerve growth factor (NGF) and brain-derived neurotrophic factor (BDNF) are involved in development and function of neural cells as well as regulation of inflammatory processes (Hoyle 2003). Similarly, EAL proteins could play different functions depending on where they are expressed.

Conclusion

The recent identification of the pollen tube guidance gene *ZmEA1* and of similar genes in our *S. chacoense* ovule EST database has led us to uncover a new family

of small secretory proteins. The failure to identify dicot EA1-like sequences in previous studies illustrates the challenge we face in the Post-genomic era. Homology-based sequence searches may fail due to the short protein length and the low sequence conservation. The situation is further complicated by the fact that small proteins are poorly annotated in the genomic sequences. Careful structural analyses are necessary to identify homologous sequences. Newly identified dicot EALs belong to phylogenetically and structurally different clades from ZmEA1 and no biological roles have been determined for any of them. Are they involved in different developmental processes? If so, do they act in signaling pathways ? Functional analyses should help decipher the roles of this extended family of EA1-like proteins.

Acknowledgement We thank Hugo Germain for useful discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Canada Research Chair Program. D. P. Matton holds a Canada Research Chair in Functional Genomics and Plant Signal Transduction.

References

Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S (2004a) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 17:349–56

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004b) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–95

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113–116

Cheung AY, Wang H, Wu HM (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. Cell 82:383–93

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, Totoya, pp 571–608

Germain H, Rudd S, Zotti C, Caron S, O'Brien M, Chantha SC, Lagace M, Major F, Matton DP (2005) A 6374 unigene set corresponding to low abundance transcripts expressed following Fertilization in solanum chacoense bitt, and characterization of 30 receptor-like kinases. Plant Mol Biol 59:515–532

Higashiyama T, Yabe S, Sasaki N, Nishimura Y, Miyagishima S, Kuroiwa H, Kuroiwa T (2001) Pollen tube attraction by the synergid cell. Science 293:1480–1483

Higashiyama T, Kuroiwa H, Kuroiwa T (2003) Pollen-tube guidance: beacons from the female gametophyte. Curr Opin Plant Biol 6:36–41

Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383–402

Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14:378–379

Hoyle GW (2003) Neurotrophins and lung disease. Cytokine Growth Factor Rev 14:551–558

Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol 35:993–1001

Kim S, Mollet JC, Dong J, Zhang K, Park SY, Lord EM (2003) Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. Proc Natl Acad Sci USA 100:16125–16130

Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene 299:1–34

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

Madoka Y, Mori H (2000a) Acropetal disappearance of PsAD1 protein in pea axillary buds after the release of apical dominance. Plant Cell Physiol 41:556–564

Madoka Y, Mori H (2000b) Two novel transcripts expressed in pea dormant axillary buds. Plant Cell Physiol 41:274–281

Marton ML, Cordts S, Broadhvest J, Dresselhaus T (2005) Micropylar pollen tube guidance by egg apparatus 1 of maize. Science 307:573–576

McCormick S, Yang H (2005) Is there more than one way to attract a pollen tube? Trends Plant Sci 10:260–263

McGurl B, Pearce G, Orozco-Cardenas M, Ryan CA (1992) Structure, expression, and antisense inhibition of the systemin precursor gene. Science 255:1570–1573

Nickel W (2003) The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. Eur J Biochem 270:2109–2119

Palanivelu R, Brass L, Edlund AF, Preuss D (2003) Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels. Cell 114:47–59

Park AR, Cho SK, Yun UJ, Jin MY, Lee SH, Sachetto-Martins G, Park OK (2001) Interaction of the Arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. J Biol Chem 276:26688–26693

Pearce G, Moura DS, Stratmann J, Ryan CA (2001a) Production of multiple plant hormones from a single polyprotein precursor. Nature 411:817–820

Pearce G, Moura DS, Stratmann J, Ryan CA Jr (2001b) RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. Proc Natl Acad Sci USA 98:12843–12847

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

Sanchez-Capelo A (2005) Dual role for TGF-beta1 in apoptosis. Cytokine Growth Factor Rev 16:15–34

Swanson R, Edlund AF, Preuss D (2004) Species specificity in pollen-pistil interactions. Annu Rev Genet 38:793–818

Wen J, Lease KA, Walker JC (2004) DVL, a novel class of small polypeptides: overexpression alters Arabidopsis development. Plant J 37:668–677

Yang H, Matsubayashi Y, Nakamura K, Sakagami Y (2001) Diversity of Arabidopsis genes encoding precursors for phytosulfokine, a peptide growth factor. Plant Physiol 127:842–851

Yang EJ, Oh YA, Lee ES, Park AR, Cho SK, Yoo YJ, Park OK (2003) Oxygen-evolving enhancer protein 2 is phosphorylated by glycine-rich protein 3/wall-associated kinase 1 in Arabidopsis. Biochem Biophys Res Commun 305:862–868

Yuan Q, Ouyang S, Wang A, Zhu W, Maiti R, Lin H, Hamilton J, Haas B, Sultana R, Cheung F, Wortman J, Buell CR (2005) The institute for genomic research Osa1 rice genome annotation database. Plant Physiol 138:18–26

Supplementary Table 1a. Description of EA1 homologs

Gene name	Organism	Amino Acid Sequence	GenBank ID (cDNA, EST or UniGene)	EST sources	Other references
Basal Magnoliophyta					
Illicium parviflorum 1	yellow anise	MSSSDRNGAPAKGLAY	CV190938	flower buds	
Nuphar avena 1	water lilly	MGGGASKEGGLSGSK	CK750727, CD476094	flower buds	
Magnoliids					
Liriodendron tulipifera 1	yellow poplar	MNWWDFLLQKLNELLV	CK748156	flower buds	
Persea americana 1	avocado	MGAGESKESTALAVVV	CO999761		
Liliopsida (monocot)					
Asparagus officinalis	asparagus	MNFVESIANFFQKILEYI	CV290130	male inflorescence	
Os01g23810	rice	MAVLADSAARKLLHGA	no EST	no EST	BAD53199
Os01g23810 (3rd Met)	rice	MGHAIQWFFAGVVAGL	-	-	-
Os01g23850	rice	MVVLADSAARKLLHGA	CX117134, CB964846	drought stress leaf, whole plant	BAD53205
Os01g23870	rice	MVALADSAAARKLLLHC	UniGene Os.35110	whole plant, leaf, panicle	BAD53208
Os03g30060	rice	MVVPIANLSAVVAGAAH	AK100586	unknown	AAS07064
Os06g49290	rice	MGGKGGGGGGGGKG	AU082628	callus	BAD53815
Os07g41400	rice	MVAVGYIVGAIASVAVG	no EST	no EST	BAC83883, XP_479095
Os07g41410	rice	MVGVSEFVGGLLNSAK	UniGene Os.10373	callus, leaf, root, panicle, leaf bud	BAC83885, XP_506453, XP_479097
Os07g41470	rice	MVSLGFVVGAAAAAVV	BAD31137, CX106950	whole plant	BAD31137
Os12g38660	rice	MKRNEVKEKEKTDKPS	UniGene Os.11518	leaf, callus, flower, root, endosperm, panicle, seedling, leaf bud	identical to Os12g38720
Os12g38670	rice	MKRNEVKEKEKTDKPS	UniGene Os.11518	leaf, callus, flower, root, endosperm, panicle, seedling, leaf bud	transposon containing Os12g38720
Os12g38720	rice	MKRNEVKEKEKTDKPS	UniGene Os.11518	leaf, callus, flower, root, endosperm, panicle, seedling, leaf bud	
Saccharum officinarum 1	sugarcane	MMPGIVTVAAVVGGAV	CA150619, CA130220	leaf, shoot-root transition	
Saccharum officinarum 2	sugarcane	MALGGAAAVASLLAGA	276704, CA117670, CA163296, CA182390, CA109893, CA2899	developing seeds, leaf, seedling, internod, shoot-root transition	
Saccharum officinarum 3	sugarcane	MVSTREGAVKTAAAVV	CA230227, CA257120, CA298808	internode	
Saccharum officinarum 4	sugarcane	MVYPSDLIGGLDRLRQ	CA257045, CA224376	inflorescence	
Saccharum officinarum 4 (2nd Met)	sugarcane	MVSTREGAVKTAAAVV	-	-	
Saccharum officinarum 5	sugarcane	MAIPIANLSAVVTGALEA	CA186377, CA216707, CA150179, CA190174, CA114564	apical stalk, shoot-root transition, leaf, lateral buds	
Saccharum officinarum 6	sugarcane	MAFPGTPARKLLPGGL	078750, CA158837, CA185368, CA224660, CA117163, CA1838	leaf, apical meristem, shoot-root transition, apical stalk, stalk bark	
Triticum aestivum 1	wheat	MNTGNIVAGLGYVKNH	UniGene Ta.6606	flower, root, leaf, inflorescence	
Triticum aestivum 1 (3rd Met)	wheat	MVSTVSLAKPYLPAKLS	-	-	
Triticum aestivum 2	wheat	MTFSLFRDPLEQVFPKI	UniGene Ta.16319	crown, root, inflorescence	
Triticum aestivum 2 (5th Met)	wheat	MVSTRDGAVGTAAMVI	-	-	
Triticum aestivum 3	wheat	MAVPAGFIFGAAAAAVI	CV973548, CV973567	proembryo	
Zea maize 2	corn	MATATTEAEVRESLGTI	UniGene Zm.2519	flower, embryo, seed, root	
Zea maize 2 (2nd Met)	corn	MASRHPVAGAAAVIVV	-	-	
Zea maize 3	corn	MRVAVSVVIPALVVALV	UniGene Zm.21414	root	
Zea maize 4	corn	LTSHLWAWLAAARAVA	UniGene Zm.21588	root, shoot apical meristem	
ZmEA1	corn	MSSCPAIVMKDDDGIG	DR451979, DT535361, DR906761, DR906455, DR451979	egg apparatus	AAW58117
ZmEA1 (3rd Met)	corn	MGVFGIYFLWPVVGPT	-	-	

Eudicotyledons (true dicot)					
AtEAL1	Arabidopsis thaliana	MSDNDRNNNQKQDTS	no EST	no EST	chr2 C/1135985-1136212
AtEAL1 (2nd Met)	Arabidopsis thaliana	MAVIGVVGAIAGLFSGM	-	-	-
AtEAL2	Arabidopsis thaliana	MGGKGGSGSGGGGKG	no EST	no EST	chr2 W/13024424-13024690
AtEAL3	Arabidopsis thaliana	MEGKGGSGGGGKGGG	no EST	no EST	chr2 W/13025629-13025930
AtEAL4	Arabidopsis thaliana	MDKNVSAKASSVAQSS	no EST	no EST	chr4 C/15984747-15984941
AtEAL5	Arabidopsis thaliana	MSGKRSNVGGGKSGG	no EST	no EST	chr5 C/24703987-24704220
Brassica napus 1	canola	MDKKNDSAKSGNAAQ	CD812049	seed	
Brassica napus 2	canola	MNSTERKDDKDESSTF	CD843272	anthers	
Brassica napus 3	canola	MILEYFDTQTKVISLVIA	CD844293	anthers	
Brassica napus 4	canola	MAGVNKSKHSYSDHIN	CD845031, CD844183	anthers	
Gerbera hybrida 1	gerbera	MNTDNVAXDIFTKVISA	AJ752103, AJ752424, AJ760291, AJ760101, AJ765147	leaf, late petal, pappus bristles	
Gerbera hybrida 1 (2nd Met)	gerbera	MVLSFFRFLGRCLCCG	-	-	
Gerbera hybrida 2	gerbera	MNTENIAGDIFTKVISAV	AJ751894, AJ760051	leaf, late petal	
Gerbera hybrida 3	gerbera	MNTENIAGDIFTKVISAV	AJ760323, AJ760264,	leaf	
Gerbera hybrida 3 (2nd Met)	gerbera	MVLSFFRFLGRCLCCG	-	-	
Gerbera hybrida 4	gerbera	KGLYLLTNNSKSQTTST	AJ764586	pappus bristles	
Helianthus annuus 1	sunflower	GGIGHILVVCFNGLRPF	AJ828170	heart-shaped embryo	
Juglans regia 1	walnut	MVAGAVATGVVVLAVG	CV197263, CV195202, CV194696, CB303550	seed coat	
Limonium bicolor 1	Limonium bicolor	WIPRAQEFGTRGGSAA	CX263137	unknown	
Lycopersicon esculentum 1	tomato	MENIESVAKAAVEKLKE	397916, BW690796, BP902783, BP896499, BI209542, BG1351	crown gall, leaf, ovary, carpel, fruit	
Lycopersicon esculentum 1 (3rd Met)	tomato	MVENAKEIVLKTTPSFV	-	-	
Lycopersicon esculentum 3	tomato	MGIEEVKNYAIEKLKELI	I210587, BI210430, BI207151, BI203863, BG132009, BF11370	fruit, crown gall, suspension cells, root,	
Phaseolus vulgaris 1	bean	MDRLIDRGSHFLLRAST	CV535535	nodule	
Populus euphratia 1	Populus euphratia	MYRSLSMEPPTSLSTIS	AJ775609	shoot	
Populus euphratia 1 (5th Met)	Populus euphratia	MVHVVKWFTKQNVIAV	-	-	
Populus tremula 1	aspen	MEPPTSLSTISHQLSDL	7048, BU822402, BU819421, BU824253, CF235125(P. alba x tre	cambium, xylem	
Populus tremula 1 (5th Met)	aspen	MVHVVKWFKKQNVTA	-	-	
Populus tremula 2	aspen	MVAVVVIVLLFLSCCGC	CK095965, BU819048	cambium	
Populus trichocarpa 1	cottonwood	MEPPTSLSTISHQSSHL	CV243090, BU881326, CV266890, CV260508, CV238394, CV	cambium, leaf, vergitative and floral buds, bark, root	
Populus trichocarpa 1 (5th Met)	cottonwood	MASGVMVHVVKWFTK	-	-	
Populus trichocarpa 2	cottonwood	MGPPAPDVLEWLQKVH	CV227850, CV225449, CK319703 (P. deltoides), CV269272(P. t	bark, xylem, stem	
P. trichocarpa x deltoides 1	cottonwood hybrid	MKFLACETVVYVLQWF	CV273578	root	
P. trichocarpa x nigra 1	cottonwood hybrid	MASGAVSDFTYMPFDC	CV261169	bark	
PsAD1	pea	MGGKGGSGGGAKGGC	AB031227	dormant axillary bud	BAA90877
ScEAL1	Solanum chacoens	MNMVENAKEIVKRTTPI	DN976436	ovule	
ScEAL2	Solanum chacoens	KGNYSSLAKKKEVLEE	DN983899	ovule	
ScEAL2 (2nd Met)	Solanum chacoens	MGSEEENNTKMMKSP	-	-	
ScEAL3	Solanum chacoens	MGIEEVKNYAIEKLKEL	DN980854	ovule	
Solanum tuberosum 1	potato	EIVKRPTPFVSEYARILF	UniGene Stu.3514	callus, floral buds, sprouting eyes, axillary buds	
Solanum tuberosum 2	potato	MSGENNYSSLVKKKQV	UniGene Stu.7590	floral buds	
Solanum tuberosum 2 (2nd Met)	potato	MGSEEENNTKMMKKP	-	-	
Solanum tuberosum 3	potato	MGIEEVKNYAIEKLKEL	CK718030, CV430131, CN514484, DN590792, CN516880	tuber, swollen stolon	

Supplementary Table 1b. Protein prediction of EA1 homologs

Gene name	Amino Acid Sequence	# of AA	MW (kDa)	pI	AUG context[a]	SignalP	SecretomeP	EAL class
Basal Magnoliophyta								
Illicium parviflorum 1	MSSSDRNGAPAKGLA	72	7.8	9.6		Y	0.60	II
Nuphar avena 1	MGGGASKEGGLSGSK	57	6.2	9.6		N	0.78	IV
Magnoliids								
Liriodendron tulipifera 1	MNWWDFLLQKLNELL	109	12.9	9.0		Y	0.78	II
Persea americana 1	MGAGESKESTALAVV	69	7.3	10.0		Y	0.50	II
Liliopsida (monocot)								
Asparagus officinalis	MNFVESIANFFQKILEY	211	23.8	7.6		Y	0.64	II
Os01g23810	MAVLADSAARKLLHGA	281	29.5	11.3	**U**CCaug**G**	Y	0.34	I
Os01g23810 (3rd Met)	MGHAIQWFFAGVVAGL	228	24.0	11.5	**G**CCaug**G**	Y	0.38	I
Os01g23850	MVVLADSAARKLLHGA	571	60.2	7.1		Y	0.58	I
Os01g23870	MVALADSAAARKLLLH	472	49.4	6.4		Y	0.51	I
Os03g30060	MVVPIANLSAVVAGAAH	361	36.8	9.2		Y	0.34	I
Os06g49290	MGGKGGGGGGGGKG	116	9.6	9.7		N	0.79	III
Os07g41400	MVAVGYIVGAIASVAVG	73	7.3	9.5		Y	0.88	I
Os07g41410	MVGVSEFVGGLLNSAK	112	10.8	10.1		N	0.74	I
Os07g41470	MVSLGFVVGAAAAAVV	66	6.6	9.8		Y	0.79	I
Os12g38660	MKRNEVKEKEKTDKPS	102	11.2	8.9		Y	0.47	II
Os12g38670	MKRNEVKEKEKTDKPS	1133	128.0	6.7		Y	0.25	II
Os12g38720	MKRNEVKEKEKTDKPS	102	11.2	8.9		Y	0.47	II
Saccharum officinarum 1	MMPGIVTVAAVVGGAV	74	7.3	10.3		Y	0.57	I
Saccharum officinarum 2	MALGGAAAVASLLAGA	75	7.2	10.0		Y	0.60	I
Saccharum officinarum 3	MVSTREGAVKTAAAVV	79	7.7	10.4		Y	0.84	I
Saccharum officinarum 4	MVYPSDLIGGLDRLRQ	168	16.0	10.5	**G**AGaug**G**	N	0.84	I
Saccharum officinarum 4 (2nd Met)	MVSTREGAVKTAAAVV	65	6.7	9.5	**G**GCaug**G**	Y	0.82	I
Saccharum officinarum 5	MAIPIANLSAVVTGALEA	225	23.2	10.3		Y	0.40	I
Saccharum officinarum 6	MAFPGTPARKLLPGGL	156	16.7	9.9		Y	0.66	I
Triticum aestivum 1	MNTGNIVAGLGYVKNH	150	14.8	10.1	**C**GUaug**A**	N	0.21	I
Triticum aestivum 1 (3rd Met)	MVSTVSLAKPYLPAKLS	92	9.1	10.1	**G**CCaug**G**	Y	0.67	I
Triticum aestivum 2	MTFSLFRDPLEQVFPK	313	33.8	9.9	**C**GUaug**A**	N	0.32	I
Triticum aestivum 2 (5th Met)	MVSTRDGAVGTAAMV	78	8.0	11.3	**G**CCaug**G**	Y	0.65	I
Triticum aestivum 3	MAVPAGFIFGAAAAAVL	75	7.8	10.5		Y	0.75	I
Zea maize 2	MATATTEAEVRESLGT	101	10.3	9.9	**A**AAaug**G**	N	0.72	I
Zea maize 2 (2nd Met)	MASRHPVAGAAAVIVV	74	7.4	10.3	**G**ACaug**G**	Y	0.79	I
Zea maize 3	MRVAVSVVIPALVVALV	73	8.1	10.4		Y	0.86	II
Zea maize 4	LTSHLWAWLAAARAVA	incomplete	incomplete	incomplete		incomplete	0.86	I
ZmEA1	MSSCPAIVMKDDDGIG	92	9.2	5.9	**U**GCaug**U**	N	0.07	I
ZmEA1 (3rd Met)	MGVFGIYFLWPVVGPT	66	6.7	8.8	**G**CUaug**G**	Y	0.48	I

Eudicotyledons (true dicot)								
AtEAL1	MSDNDRNNNQKQDTS	75	8.2	10.1	**A**AAaug**A**	Y	0.60	II
AtEAL1 (2nd Met)	MAVIGVVGAIAGLFSGM	54	5.9	10.7	**G**CCaug**G**	Y	0.65	II
AtEAL2	MGGKGGSGSGGGGKG	88	7.7	9.7		N	0.78	III
AtEAL3	MEGKGGSGGGGKGGG	101	8.5	9.6		N	0.77	III
AtEAL4	MDKNVSAKASSVAQSS	64	6.3	9.6		N	0.81	IV
AtEAL5	MSGKRSNVGGGKSGG	77	7.0	10.5		N	0.55	III
Brassica napus 1	MDKKNDSAKSGNAAQT	62	6.2	9.2		N	0.72	IV
Brassica napus 2	MNSTERKDDKDESSTF	86	9.4	9.4		N	0.27	II
Brassica napus 3	MILEYFDTQTKVISLVIAL	61	7.0	10.4		Y	0.79	II
Brassica napus 4	MAGVNKSKHSYSDHIN	102	11.1	9.3		N	0.68	II
Gerbera hybrida 1	MNTDNVAXDIFTKVISAV	117	12.6	9.2	**C**AAaug**A**	N	0.80	II
Gerbera hybrida 1 (2nd Met)	MVLSFFRFLGRCLCCG	59	6.4	11.3	**A**CCaug**G**	Y	0.76	II
Gerbera hybrida 2	MNTENIAGDIFTKVISAV	159	18.0	9.2		N	0.60	II
Gerbera hybrida 3	MNTENIAGDIFTKVISAV	141	16.0	9.0	**C**AAaug**A**	N	0.73	II
Gerbera hybrida 3 (2nd Met)	MVLSFFRFLGRCLCCG	59	6.6	10.8	**A**CCaug**G**	Y	0.82	II
Gerbera hybrida 4	KGLYLLTNNSKSQTTST	incomplete	incomplete	incomplete		incomplete	0.13	II
Helianthus annuus 1	GGIGHILVVCFNGLRPF	incomplete	incomplete	incomplete		incomplete	0.13	II
Juglans regia 1	MVAGAVATGVVVLAVG	68	7.1	11.7		Y	0.52	II
Limonium bicolor 1	WIPRAQEFGTRGGSAA	incomplete	incomplete	incomplete		incomplete	0.56	N/A
Lycopersicon esculentum 1	MENIESVAKAAVEKLKE	166	18.9	10.3	**U**CAaug**G**	N	0.68	II
Lycopersicon esculentum 1 (3rd Met)	MVENAKEIVLKTTPSFV	147	16.8	10.7	**A**AUaug**G**	N	0.71	II
Lycopersicon esculentum 3	MGIEEVKNYAIEKLKELF	117	13.8	9.3		N	0.81	II
Phaseolus vulgaris 1	MDRLIDRGSHFLLRAST	158	17.7	8.5		N	0.85	II
Populus euphratia 1	MYRSLSMEPPTSLSTIS	151	17.0	9.3	**U**UAaug**U**	N	0.67	II
Populus euphratia 1 (5th Met)	MVHVVKWFTKQNVIAV	99	11.2	9.6	**G**UAaug**G**	Y	0.54	II
Populus tremula 1	MEPPTSLSTISHQLSDL	101	13.4	9.5	**U**CAaug**G**	N	0.74	II
Populus tremula 1 (5th Met)	MVHVVKWFKKQNVTA	72	8.3	9.8	**G**UAaug**G**	Y	0.71	II
Populus tremula 2	MVAVVVIVLLFLSCCGC	86	9.5	9.4		Y	0.72	II
Populus trichocarpa 1	MEPPTSLSTISHQSSHL	101	13.4	9.4	**U**CAaug**G**	N	0.62	II
Populus trichocarpa 1 (5th Met)	MASGVMVHVVKWFTK	77	8.6	9.6	**G**UAaug**G**	Y	0.68	II
Populus trichocarpa 2	MGPPAPDVLEWLQKVF	75	8.5	10.4		Y	0.71	II
P. trichocarpa x deltoides 1	MKFLACETVVYVLQWF	76	8.9	10.7		Y	0.63	II
P. trichocarpa x nigra 1	MASGAVSDFTYMPFDC	199	22.0	9.4		N	0.64	II
PsAD1	MGGKGGSGGGAKGGC	87	7.2	10.0		N	0.82	III
ScEAL1	MNMVENAKEIVKRTTPF	131	14.9	10.2		N	0.54	II
ScEAL2	KGNYSSLAKKKEVLEE	incomplete	incomplete	incomplete	incomplete	N	0.57	II
ScEAL2 (2nd Met)	MGSEEENNTKMMKSP	41	4.8	9.2	**A**UUaug**G**	N	0.73	IV
ScEAL3	MGIEEVKNYAIEKLKELF	113	13.3	9.3	**A**AGaug**G**	N	0.80	II
Solanum tuberosum 1	EIVKRPTPFVSEYARILF	incomplete	incomplete	incomplete		N	0.57	II
Solanum tuberosum 2	MSGENNYSSLVKKKQV	92	10.2	7.9	**A**GAaug**A**	N	0.46	II
Solanum tuberosum 2 (2nd Met)	MGSEEENNTKMMKKP	41	4.8	9.2	**A**UUaug**G**	N	0.72	IV
Solanum tuberosum 3	MGIEEVKNYAIEKLKELF	121	14.2	9.6		N	0.76	II

a - Gray letters indicate mismatches to the AUG context consensus (**A/G**)NNaug**G**. Only genes with alternative AUG sites are shown.

Supplementary Table 2. Primers used for RT-PCR analyses

Gene	Sequence	product size (bp)
AtEAL1	CCTGAAAATTCGCCCATCCT	210
	AGAATCTTGATAATCGTCACTCATC	
At2g30560*	GAGGAAAAGGTGGTGGTGGT	263
	ACTTTTCCCTCCTGCTCCAC	
AtEAL2	GTGGAGCAGGAGGGAAAAGT	247
	GACCAAACTCACTTTTTACCATGGAG	
AtEAL3	TCCTTTTCACTTTCGTTTCTCATTCA	338
	ACCCTGTCCACTACCATGCAAAT	
AtEAL4	TTGCCTTCTCCTTGGCATGA	207
	TTCAAGAGTTAAAACGTAAAATGGACA	
AtEAL5	CACTCCGAATCGGGTCTCTTT	252
	CCCATTTGTCCCTGGTGCTA	
Actin 2	ATAACCATCGGAGCTGAGAGATTCC	334
	TTGAAATCCACatCTGTTGGAAGGT	

*- At2g30560 contains AtEAL2 and AtEAL3 as two exons. This primer pair was designed to determine if the two genes are transcribed as a single gene. The forward primer binds to the first exon and the reverse primer binds to the second exon. No RT-PCR products were detected using this primer combination.