the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2010) **63**, 952–959 doi: 10.1111/j.1365-313X.2010.04292.x

Parallel structural evolution of auxin response factors in the angiosperms

Cédric Finet, Chloé Fourquin, Marion Vinauger, Annick Berne-Dedieu, Pierre Chambrier, Sandrine Paindavoine and Charles P. Scutt*

Laboratoire de Reproduction et Développement des Plantes (UMR 5667 – CNRS, INRA, Université de Lyon) Ecole Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon Cedex 07, France

Received 1 April 2010; revised 23 June 2010; accepted 28 June 2010; published online 3 August 2010.
*For correspondence (fax +33 4 72 72 86 00; e-mail Charlie.Scutt@ens-lyon.fr).

SUMMARY

Here we analyze the structural evolution of the paralogous transcription factors ETTIN (ETT/ARF3) and AUXIN RESPONSE FACTOR 4 (ARF4), which control the development of floral organs and leaves in the model angiosperm Arabidopsis. ETT is truncated at its C terminus, and consequently lacks two regulatory domains present in most other ARFs, including ARF4. Our analysis indicates *ETT* and *ARF4* to have been generated by the duplication of a non-truncated ARF gene prior to the radiation of the extant angiosperms. We furthermore show that either *ETT* or *ARF4* orthologs have become modified to encode truncated ARF proteins, lacking C-terminal regulatory domains, in representatives of three groups that separated early in angiosperm evolution: Amborellales, Nymphaeales and the remaining angiosperm clade. Interestingly, the production of truncated *ARF4* transcripts in Amborellales occurs through an alternative splicing mechanism, rather than through a permanent truncation, as in the other groups studied. To gain insight into the potential functional significance of truncations to *ETT* and *ARF4*, we tested the capacity of native, truncated and chimeric coding sequences of these genes to restore a wild-type phenotype to Arabidopsis *ett* mutants. We discuss the results of this analysis in the context of the structural evolution of ARF genes in the angiosperms.

Keywords: auxin response factor, angiosperm, molecular evolution, alternative splicing, ETTIN, auxin response factor 4, amborella trichopoda, cabomba, aquatica, ephedra distachya.

INTRODUCTION

ETTIN (*ETT*/*ARF3*) and *AUXIN RESPONSE FACTOR 4* (*ARF4*) are paralogous members of the ARF family, which contains a total of 23 genes in Arabidopsis (Guilfoyle and Hagen, 2007). ARF transcription factors are known to regulate gene expression in response to the plant hormone auxin by binding to auxin response elements (AuxREs) in the promoters of their target genes (Ulmasov *et al.*, 1997). Strong *ett* mutations produce severe defects in carpel development and alterations in floral phyllotaxy. Effects of these mutations on carpel development include reduced development of the ovary and style tissues, excessive growth of the stigma, incomplete carpel fusion and the ectopic production of transmitting tissue in external positions (Sessions and Zambryski, 1995; Sessions *et al.*, 1997). This latter effect represents a disruption to abaxial–adaxial polarity in the carpel, in which the abaxial and adaxial poles correspond, respectively, to the outer and inner surfaces of the ovary wall.

Plants in which only *ARF4* is inactivated show no aberrant developmental phenotype. However, double *ett arf4* mutants show a breakdown of abaxial tissue specification in all lateral organs (Pekker *et al.*, 2005), indicating an extensive overlap in the functions of these genes. The activity of *ETT* and *ARF4* is restricted to the abaxial domain of lateral organs through post-transcriptional regulation by two *TAS3*-generated tasi-microRNAs (Garcia *et al.*, 2006). *ETT* is also known to be regulated by upstream open reading frames (uORFs) in its 5′ leader sequence, which block translation of the main ORF in the absence of specific ribosomal factors (Nishimura *et al.*, 2004a,b, 2005). Several further members of the ARF family including *ARF4* possess uORFs, suggesting that these genes may also be regulated at the translational level.

Despite their partially redundant functions and similar mechanisms of regulation, ETT and ARF4 show distinct molecular structures, as ETT lacks two C-terminal dimerization domains, termed domains III and IV, which are present

in ARF4 and most other ARF proteins. These domains are known to interact with similar sequences present in negative regulators of ARF activity, termed Aux/IAA proteins (Ulmasov *et al.*, 1999a; Tiwari *et al.*, 2001). Aux/IAA proteins are themselves negatively regulated by auxin-binding F-box proteins (Dharmasiri *et al.*, 2005; Kepinski and Leyser, 2005), thus explaining the positive effect of auxin on ARF activity. Domains III and IV are also known to facilitate homodimerization, which appears to be required in some cases for the efficient binding of ARFs to DNA (Ulmasov *et al.*, 1999b).

Molecular phylogenetic studies indicate the angiosperms, or flowering plants, and the extant gymnosperms, or non-flowering seed plants, to form two sister clades (Bremer *et al.*, 2009). Within the angiosperms, three extant lineages, Amborellales, Nymphaeales and Austrobaileyales, collectively termed the ANA grade, appear to have diverged in a basal position from a remaining common lineage, from which all other living angiosperms are descended. Amborellales contains the single species *Amborella trichopoda*, which is a shrub that is endemic to the Southern Pacific island of New Caledonia, whereas Nymphaeales and Austrobaileyales each contain three small families of aquatic and woody species, respectively. The most recent phylogenetic analyses support the sequential divergence of Amborellales, Nymphaeales and Austrobaileyales from a remaining angiosperm lineage (Bremer *et al.*, 2009). However, two earlier studies supported a slightly different phylogeny in which Amborellales and Nymphaeales together formed a first-diverging clade (Qiu *et al.*, 2001; Zanis *et al.*, 2002).

As the carpel is the major defining feature of the angiosperms, and as *ETT* and *ARF4* are essential to carpel development in Arabidopsis, we decided to investigate the evolution of these genes in early angiosperms, beginning by identifying their orthologs in representative ANA-grade angiosperms and gymnosperms. The results of our study strongly suggest *ETT* and *ARF4* to have been generated by a duplication event in the angiosperm lineage, which occurred after the separation of the angiosperm and gymnosperm lineages, but prior to the radiation of the extant angiosperms. Our studies also show one or other of *ETT* and *ARF4* to have become modified to encode a truncated ARF protein, lacking C-terminal dimerization domains, in each of three lineages that diverged early in angiosperm evolution: Amborellales, Nymphaeales and the remaining angiosperm clade. We discuss several alternative scenarios for this parallel evolution of ARF protein structure based on the different forms of truncation observed, which include an alternative splicing mechanism operating in *A. trichopoda*, the probable sister to all other extant flowering plants.

To test the significance of truncations to *ETT* and *ARF4* for plant development, we measured the capacity of wild-type, chimeric and artificially truncated coding sequences of these genes to complement the *ett-1* mutant phenotype in Arabidopsis. The results of these studies indicate that the truncation of ETT contributes to its role in carpel development, and also reveal an underlying difference in the capacity of ETT and ARF4 to function in the absence of C-terminal dimerization domains. We discuss the data obtained from these functional studies in the context of ARF gene evolution in the angiosperms.

RESULTS

ETT and *ARF4* originated in a gene duplication event prior to the last common ancestor of extant angiosperms

To gain insight into the early evolution of *ETT* and *ARF4* in angiosperms, we identified closely related homologs of these genes by database searching, and by screening cDNA libraries made from reproductive tissues of the ANA-grade angiosperms *A. trichopoda* (Amborellales) and *Cabomba aquatica* (Nymphaeales), and of the gymnosperms *Ephedra distachya* (Gnetales) and *Ginkgo biloba* (Ginkgoales) (Table S1). We performed phylogenetic analyses of the sequences identified using both an extensive data set containing all well-aligned sites, and a restricted data set corresponding to the 5′ regions of these molecules, including their predicted DNA-binding domains. Both nucleotide and amino-acid data sets corresponding to these alignments were used in phylogenetic analyses (Figure S1).

All of the phylogenetic analyses performed in this study revealed a pair of well-supported clades in sister positions, including *ETT* and *ARF4* from Arabidopsis (Figures 1 and S2). The internal structure of each of these clades broadly recapitulates angiosperm phylogeny, with genes from *A. trichopoda* and *C. aquatica* occurring in basally diverging positions. In common with *ETT* and *ARF4* from Arabidopsis, the angiosperm ARF genes identified in this study were found to possess sites of regulation by *TAS3*-generated tasi-microRNAs, in addition to uORFs in their 5′ leader sequences, which may function in translational regulation (Table S2). Also like *ETT* and *ARF4*, these genes were found to be expressed in both reproductive and vegetative tissues (Figure S3). *In situ* hybridization of *AtrARF4* from *A. trichopoda* and of *CaqARF4* from *C. aquatica* revealed clear hybridization signals in ovule tissues, similar to signals obtained using *ETT* and *ARF4* probes in Arabidopsis (Figure S4). Taken together, these data support the close relationship between the angiosperm genes identified in this study and *ETT* and *ARF4* from Arabidopsis. Furthermore, our phylogenetic reconstructions strongly suggest the orthology of the angiosperm genes present in each of the two sister clades containing Arabidopsis *ETT* and *ARF4*, as identified in these analyses (Figures 1 and S2). As both of these clades include genes from ANA-grade angiosperms, we propose the duplication that separated the *ETT* and *ARF4* lineages to have predated the initial radiation of the extant angiosperms. The *ARF4* clade contains no gene from the Poaceae monocot *Oryza sativa*, the complete genome sequence of

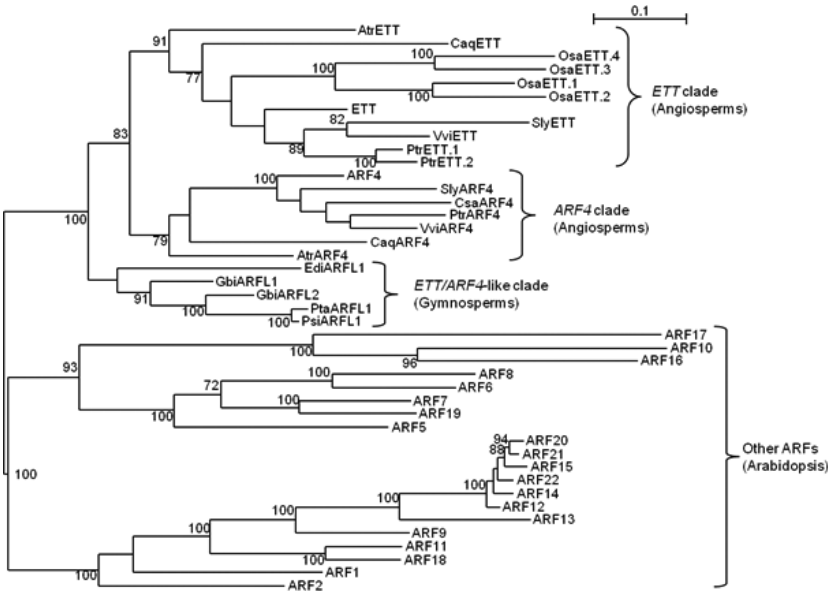

Figure 1. A maximum-likelihood phylogeny of *ETT*- and *ARF4*-like genes in seed plants from an alignment of full-length nucleotide sequences (see Figure S1).
Clade support values superior to 70% are indicated. Species of origin are indicated by gene name prefixes: *Atr*, *Amborella trichopoda*; *Caq*, *Cabomba aquatica*; *Csa*, *Cucumis sativus*; *Edi*, *Ephedra distachya; Gbi*, *Ginkgo biloba*; *Osa*, *Oryza sativa; Psi*, *Picea sitchensis*; *Pta*, *Pinus taeda*; *Ptr*, *Populus trichocarpa*; *Sly*, *Solanum lycopersicum*; *Vvi*, *Vitis vinifera*; no prefix, *Arabidopsis thaliana*. Accession numbers are given in Table S1.

which is now available (Goff *et al.*, 2002), suggesting that this gene was lost from the *O. sativa* lineage after the divergence of the monocots and eudicots. In certain cases, duplications have occurred to generate several paralogs of *ETT*, leading for example to four *ETT*-like genes in *O. sativa* and two in *Populus trichocarpa* (Figures 1 and S2).

Our phylogenetic analyses (Figures 1 and S2) also revealed a group of gymnosperm ARF genes positioned externally to the combined angiosperm *ETT* + *ARF4* clade. These genes formed a clade in the analysis shown in Figure 1, whereas they formed a grade of lineages in the three analyses shown in Figure S2. The internal structure of this gymnosperm clade or grade thus varied between analyses, making it difficult to conclude the precise relationships of orthology between the genes within it. Significantly, however, no gymnosperm genes grouped within the angiosperm *ETT* or *ARF4* clades, and no angiosperm genes grouped within this gymnosperm clade or grade. The genes of this clade or grade contain sites associated with post-transcriptional regulation by *TAS3*-generated tasi-miRNAs and uORFs in their 5′ leader sequences, as do putative angiosperm *ETT* and *ARF4* orthologs (Table S2). Taken together, these data suggest the gymnosperm genes identified here are orthologous to the combined *ETT* + *ARF4* clade of angiosperm genes (Figure 1). This interpretation implies the duplication event that generated *ETT* and *ARF4* to have occurred in the angiosperm lineage, subsequent to the last common ancestor of the extant angiosperms and gymnosperms. Of the gymnosperm *ETT/ARF4*-like genes identified here, we chose to investigate the expression of *EdiARFL1* from *E. distachya* because of the relative simplicity of performing *in situ* hybridization in this species. *EdiARFL1* was expressed both in reproductive and vegetative tissues (Figure S3), and showed a clear *in situ* hybridization signal in ovule tissues, as did *ETT* and *ARF4* in Arabidopsis (Figure S4).

Truncated transcripts of *ETT* or *ARF4* are present in three early diverging angiosperm lineages

The gymnosperm *ETT/ARF4*-like genes identified in the present study (Figure 1, Table S1) were found to encode proteins of typical ARF structure, including C-terminal interaction domains III and IV. In this respect, these genes resemble the *ARF4* orthologs of eudicots, and are distinct from the *ETT* orthologs of both eudicots and monocots, which encode truncated ARF proteins that lack C-terminal dimerization domains (Sato *et al.*, 2001; Guilfoyle and Hagen, 2007). Interestingly, we found the *ARF4* orthologs of the ANA-grade angiosperms *A. trichopoda* and *C. aquatica* to generate truncated transcripts, whereas the *ETT* orthologs of these species encoded full-length ARF proteins.

To further investigate this apparent switch in the molecular structure of *ETT* and *ARF4* between ANA-grade and later diverging angiosperms, we analyzed the genomic sequences of the novel ARF genes identified. We found the truncation of *CaqARF4* from *C. aquatica* to be generated by the presence of a stop codon that prematurely terminates exon 11 of this gene, whereas the truncation affecting *ETT* orthologs in eudicots such as *Arabidopsis thaliana*, and monocots such as *O. sativa*, is caused by a stop codon terminating exon 10 of these genes (Figure 2). *AtrARF4* from the ANA-grade angiosperm *A. trichopoda* was also found to encode a truncated ARF protein, lacking domains III and IV. However, we found the corresponding genomic sequence to possess the structure of a typical, non-truncated ARF gene. Further analysis revealed *AtrARF4* to be subject to an alternative splicing mechanism that generates both long

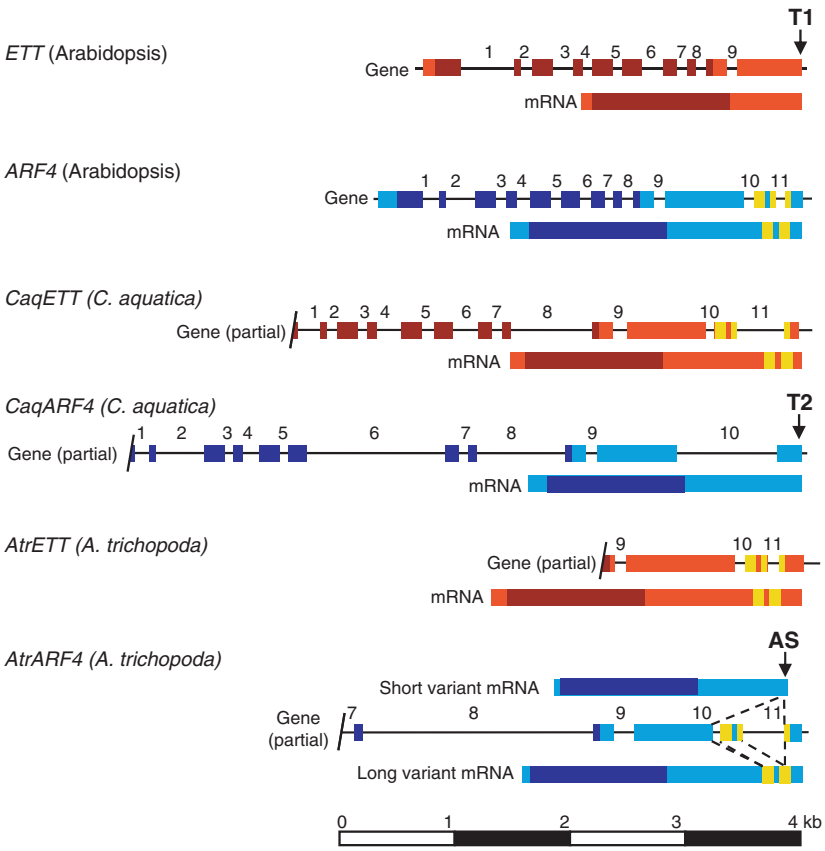

Figure 2. The structure of *ETT* and *ARF4* orthologs in Arabidopsis, *Cabomba aquatica and Amborella trichopoda*.
DNA-binding domains of *ETT* (red) and *ARF4* (blue) orthologs are shown in darker shading, and domains III and IV are colored yellow. Introns are numbered, and the C-terminal protein truncations caused by premature stop codons (T1 and T2) and alternative splicing (AS) are indicated.

and short variant transcripts. The short variant transcript of this gene is produced by the excision of a cassette containing intron 10, exon 11 and intron 11 (Figure 2), which also shifts the reading frame and completely eliminates domains III and IV from the encoded protein (Figure S5). The relative abundance of the short- and long-variant transcripts of *AtrARF4* was found to be 2.2:1 in an *A. trichopoda* female flower cDNA library, indicating the short variant transcript to predominate significantly, at least in female flowers. Both short- and long-variant forms of AtrARF4 were also observed in male flower RNA using RT-PCR analysis (data not presented). We also used RT-PCR to test for the presence of short-variant transcripts of *ARF4* in Arabidopsis, *EdiARFL1* in *E. distachya*, and *GbiARFL1* and *GbiARFL2* in *G. biloba*, but found no evidence of alternative splicing in these cases (results not presented).

Our analysis of ARF gene structure thus indicates that truncated transcripts of either *ETT* or *ARF4* are produced through distinct mechanisms in representatives of lineages that diverged at early stages in angiosperm evolution: a premature stop codon in exon 10 of the *ETT* orthologs of eudicots and monocots, a premature stop codon in exon 11 of *CaqARF4* in *C. aquatica* (Nymphaeales), and an alternative splicing mechanism operating on *AtrARF4* in *A. trichopoda* (Amborellales).

The truncation of *ETT* contributes to its function in carpel development

The discovery that truncations to either *ETT* or *ARF4* are widespread in the angiosperms led us to examine the functional significance of these structural modifications. To address this question, we assessed the ability of various wild-type, chimeric and artificially truncated versions of the *ETT* and *ARF4* coding sequences to complement the Arabidopsis *ett-1* mutation when expressed downstream of the *ETT* promoter and 5′ leader sequence (*proETT*). The strong *ett-1* allele was used in these experiments to provide a large range of phenotypic variation through which genetic complementation could be assessed. Chimeric constructs were made by the recombination of the *ETT* and *ARF4* coding sequences at a point of clear alignment between these, upstream of sequences encoding domain III in *ARF4*, whereas a truncated *ARF4* construct was generated by the introduction of a stop codon at this same point (Figure 3a).

A construct containing the wild-type *ETT* coding sequence, *proETT::ETT* (Figure 3a), completely restored wild-type gynoecium development and fertility to all *ett-1* homozygotes analyzed in a T_1 population (Figure 3b; Table S3). This positive control construct thus established that the *cis*-acting regulatory sequence used in these experiments

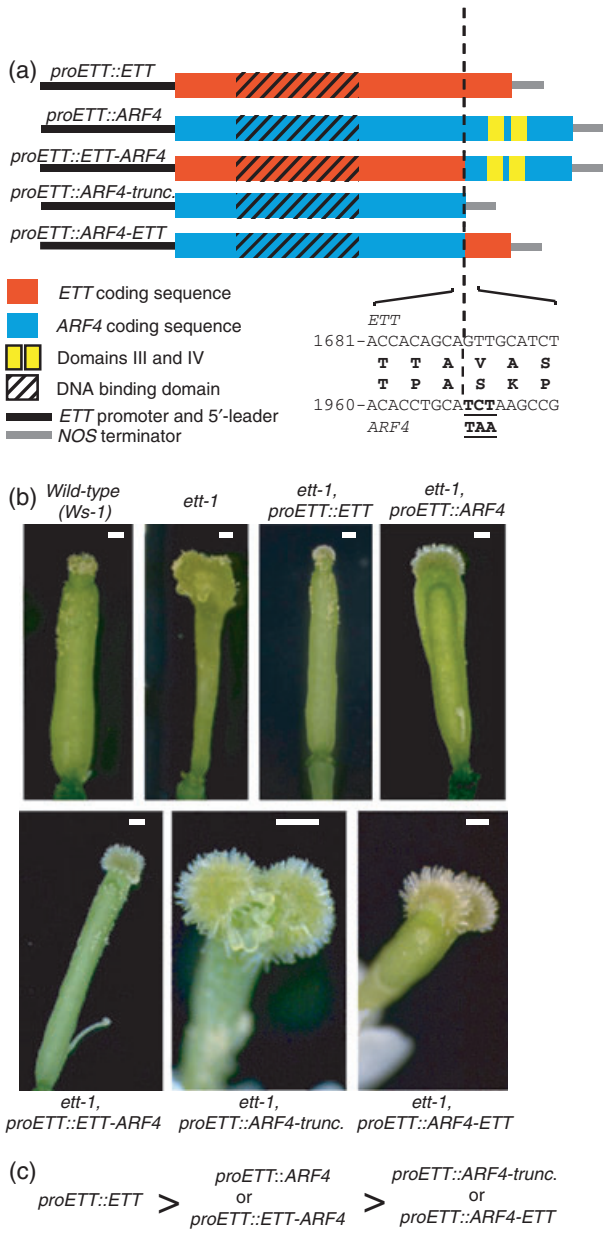

Figure 3. The effect of ARF truncations on Arabidopsis carpel development. (a) Transgene constructs containing wild-type, chimeric and truncated *ETT* and *ARF4* coding sequences, fused to the *ETT* promoter and 5′ leader. Nucleotides at the point of recombination (dashed line) or premature truncation (TAA stop codon) are indicated. (b) Gynoecia of wild-type Arabidopsis, *ett-1* mutants, and typical examples of *ett-1* mutants transformed with the constructs indicated. Scale bars: 100 μm. (c) Relative strengths of the constructs tested in complementing the *ett-1* phenotype.

was fully functional, and provided a standard against which the complementation achieved using other coding sequences could be measured.

Constructs containing either the wild-type *ARF4* coding sequence (*proETT::ARF4* in Figure 3a) or a translational fusion encoding the DNA binding domain of ETT, together with domains III and IV of ARF4 (*proETT::ETT-ARF4* in Figure 3a), were able to partially restore wild-type gynoecium development to *ett-1* mutants (Figure 3b; Table S3). Gynoecia of the majority of T_1 transformants containing either of these constructs showed valves and stigmatic tissues that were intermediate in size between those of wild-type plants and *ett-1* mutants. These partially complemented transformants showed reduced fertility in comparison with wild-type plants, although they did set some seed. In the T_2 generation, all transformants containing these constructs showed partial complemention, resembling the majority of T_1 transformants (Table S3).

The majority of T_1 plants transformed either with a chimeric construct encoding the DNA-binding domain of ARF4, translationally fused to the C terminus of ETT (*proETT::ARF4-ETT* in Figure 3a), or with an *ARF4* construct that had been truncated by the insertion of a premature stop codon (*proETT::ARF4-trunc* in Figure 3a), showed little if any restoration of wild-type gynoecium development (Figure 3b; Table S3). A minority of plants transformed with either of these constructs produced a proportion of flowers with less severe gynoecium phenotypes and that set some seed. In the T_2 generation, however, the progeny of these plants reverted almost completely to an *ett-1*-like phenotype.

The results of these analyses indicate a series of strengths in the ability of the different constructs tested to complement the *ett-1* mutation, as shown in Figure 3c. The addition of domains III and IV (from ARF4) to ETT thus reduced the capacity of this protein to complement the *ett-1* mutation. Chimeric *ETT-ARF4* and native *ARF4* coding sequences showed similar capacities to complement the *ett-1* mutation, suggesting a broad functional equivalence between the DNA-binding domains of ETT and ARF4, conditionally on the presence of domains III and IV. However, the removal of sequences encoding domains III and IV from *ARF4* greatly reduced the capacity of this coding sequence to complement the *ett-1* mutation. Thus, although *ETT* and *ARF4* show broad functional similarities, ETT is able to function with or without domains III and IV, whereas ARF4 appears to require these domains for its activity, at least in the context of the present analysis, in which an *ARF4* coding sequence was used to complement a mutation in its paralog *ETT*.

DISCUSSION

A reconstruction of the evolution of *ETT* and *ARF4* in the angiosperms

The phylogenetic and structural analysis of *ETT* and *ARF4*-like genes in present-day angiosperms and gymnosperms (Figures 1 and 2) has permitted a partial reconstruction of their molecular evolution since the last common ancestor of the living seed plants. The results of this analysis support the generation of *ETT* and *ARF4* by a gene duplication event in

the angiosperm lineage, prior to the last common ancestor of the extant angiosperms (Figure 4). Interestingly, a large-scale duplication event has been mapped to approximately this evolutionary stage (De Bodt *et al.*, 2005), which may explain the origin of several gene clades of importance to flower development through duplications that occurred prior to the radiation of the extant angiosperms. These novel clades include the C- and D-function clades (Kramer *et al.*, 2004), and the *SEPALLATA3* and *SEPALLATA1/2/4* MADS-box clades (Zahn *et al.*, 2005).

Our evolutionary reconstruction indicates that, subsequent to the radiation of the angiosperms, genetic truncation events occurred that removed sequences encoding C-terminal dimerization domains from the *ETT* orthologs of monocots and eudicots (T1 in Figure 4), and from an *ARF4* ortholog of *C. aquatica* in Nymphaeales (T2 in Figure 4). The *ARF4* lineage appears to have been lost from the Poaceae monocot *O. sativa*, subsequent to the divergence of the monocot and eudicot lineages (Figure 4). A wider sampling of monocots would help to place this gene loss with greater precision along the monocot lineage. In addition, duplications of the *ETT* gene have occurred in several angiosperm lineages, including that of *O. sativa* (Figure 4).

We also found truncated transcripts of *AtrARF4* in *A. trichopoda*, the probable sister species to all other flowering plants, which are generated by an alternative splicing mechanism that removes exon 11, along with its adjacent introns (Figures 2 and S5). In animal systems, such alternative splicing mechanisms are typically associated with the presence of exonic splicing silencer (ESS) or intronic splicing silencer (ISS) motifs in the affected gene (Dredge *et al.*, 2005; Wang *et al.*, 2006), and the general conservation of splicing phenomena between plants and animals strongly suggests similar mechanisms to exist in plants (Irimia *et al.*, 2007). As the alternative splicing mechanism revealed in our study has so far only been found to operate on *AtrARF4*, the most parsimonious position for the evolutionary origin of this mechanism is on the *AtrARF4* lineage, after the radiation of the extant angiosperms (AS1 in Figure 4). According to this interpretation, truncated *ETT* or *ARF4* transcripts would have arisen in parallel through three different mechanisms in Amborellales, Nymphaeales and the remaining angiosperms, which might imply some common evolutionary constraint operating on these plant lineages, or alternatively suggest the presence of relaxed selection that allowed the truncation of one or other member of the *ETT/ARF4* pair. However, we cannot exclude the possibility that the alternative splicing mechanism described here may be of more ancient origin, predating the radiation of the extant angiosperms (AS2 in Figure 4), or even the separation of the *ETT* and *ARF4* lineages (AS3 in Figure 4). In either of these latter two cases, this alternative splicing mechanism would initially have been present in all angiosperm lineages, but would subsequently have been lost from some of these, perhaps following permanent genetic truncations to their *ETT* or *ARF4* orthologs (e.g. T1 and T2 in Figure 4). Such losses of alternative splicing might have occurred through mutations to ESS or ISS motifs in the alternatively spliced genes. A wider sampling of ANA-grade angiosperms, in which further cases of alternative splicing of *ARF4* and/or *ETT* might be found, could provide evidence to choose between these various hypotheses (AS1–AS3 in Figure 4).

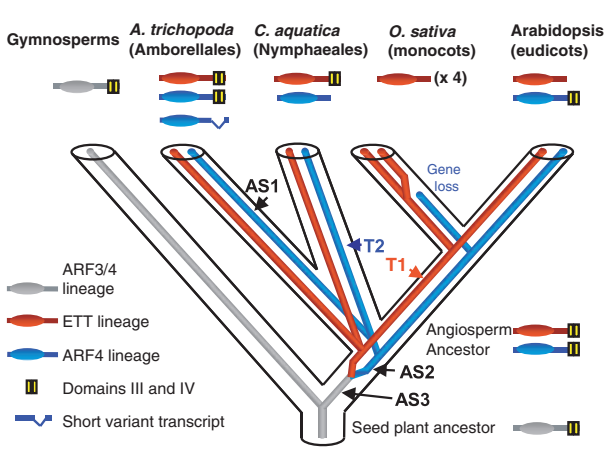

Figure 4. A partial evolutionary reconstruction of *ETT* and *ARF4* in seed plants.
Protein truncations by the insertion of premature stop codons (T1 and T2), and three possible origins for an alternative splicing mechanism operating in *A. trichopoda* (AS1–AS3), are indicated.

The functional significance of truncations to *ETT* and *ARF4*

The presence of truncated *ETT* or *ARF4* transcripts in three plant groups that diverged very early in angiosperm evolution (Figure 4) led us to test the functional significance of such truncations in Arabidopsis. This analysis (Figure 3; Table S3) indicated that the truncation of *ETT* contributed to its function in carpel development, and also revealed an underlying difference in the capacity of the N-terminal (DNA-binding) domains of ETT and ARF4 to function in the absence of C-terminal dimerization domains. This latter observation raises the question of whether, in ANA-grade angiosperms, it might be the non-truncated *ETT* orthologs, rather than the truncated *ARF4* orthologs, which require C-terminal domains for their biological activity. A full answer to this question will require the development of functional genetic approaches in ANA-grade angiosperms, which should now become a high priority for plant evo-devo studies. The truncations of *ETT* and *ARF4*, revealed by the present work to have occurred in multiple early diverging angiosperm lineages, combined with the potential functional significance of these truncations, suggests the importance of such structural modifications to ARF factors in the early evolution of the angiosperms.

EXPERIMENTAL PROCEDURES

Plant material

Tissues of *A. trichopoda* were field-collected at Col d'Amieu, New Caledonia; those of *E. distachya* and *G. biloba* were obtained from Lyon Botanic Garden, France. Plants of *C. aquatica* were obtained from Anthias S.A. (Les Chères, France) and were grown in a small aquarium. Seed of the Columbia-0 (Col-0) and Wassilewskija-2 (Ws-2) ecotypes of Arabidopsis [*Arabidopsis thaliana* (L.) Heynh.] and of the *ett-1* mutant, in a Ws-2 background, were obtained from the Nottingham Arabidopsis Stock Centre (NASC, http://arabidopsis.info). Arabidopsis plants, potted in peat-based compost, were grown in a growth chamber at 20°C, initially for 4 weeks under 8-h light/16-h dark cycles, and were then grown to flowering under 18-h light/6-h dark cycles.

Gene identification and expression analyses

cDNAs homologous to *ETT* and *ARF4* were obtained by screening previously described cDNA libraries constructed from *A. trichopoda* and *C. aquatica* flower RNA (Fourquin *et al.*, 2005), and similarly constructed cDNA libraries from mixed developmental stages of male and female reproductive structures of *E. distachya*, and from mixed developmental stages of ovules from *G. biloba*. A partial cDNA corresponding to *AtrARF4* was found by BLAST searching of an *A. trichopoda* flower expressed sequence tag database (http://pgn.cornell.edu/blast/blast_search.pl), and its sequence was then completed by RACE PCR from *A. trichopoda* flower mRNA. Genomic loci were amplified by PCR using primers derived from the corresponding full-length cDNA sequences. Northern blot hybridizations using radiolabeled probes, and *in situ* hybridizations using digoxygenin-labeled probes, were performed as previously described (Fourquin *et al.*, 2005).

Phylogenetic analyses

Amino acid and nucleotide sequences were aligned using MUSCLE (Edgar, 2004), and unambiguously aligned sites were then manually selected using SEAVIEW (Galtier *et al.*, 1996) for use in phylogenetic reconstructions. These reconstructions were made using data sets both from the highly conserved 5′ and N-terminal extremities of the sequences analyzed, and from a more extensive data set covering all well-aligned sites (Figure S1). Maximum likelihood and Bayesian inference analyses were performed using PhyML (Guindon and Gascuel, 2003) and MrBayes (Ronquist and Huelsenbeck, 2003), respectively. Analyses of amino acid sequences were performed using a WAG + I + Γ4 substitution model, whereas those of DNA sequences were performed using a JTT model. In all cases, the robustness of nodes was estimated using bootstrap tests of 1000 replicates.

Analysis of transgenic plants

ETT and *ARF4* coding sequences were amplified by RT-PCR from Arabidopsis Col-0 flower RNA, and were then re-amplified to incorporate Gateway (Invitrogen, http://www.invitrogen.com) *attB1* and *attB2* recombination sites upstream and downstream of their initiation and stop codons, respectively. Chimeric coding sequences, recombined between the DNA-binding and dimerization domains of *ETT* and *ARF4*, were produced by separately amplifying the 5′ and 3′ extremities of the *ETT* and *ARF4* coding sequences using chimeric internal primers, and combining the resulting PCR products by further rounds of PCR amplification using the appropriate terminal primers. A truncated coding sequence of *ARF4* was generated by re-amplification using an *attB2*-containing primer that also contained a stop mutation. The resulting wild-type, chimeric and truncated coding sequences were incorporated, using Gateway BP (Invitrogen) recombination reactions, into *pDONR207* (Invitrogen), and were then fully sequenced. In parallel, a fragment carrying the presumptive promoter and 5′ leader of *ETT* (*proETT*) was amplified from Arabidopsis Col-0 genomic DNA. The PCR primers used in this amplification incorporated an *Xba*I site immediately upstream of the ATG start codon of the main ORF in *ETT*, and straddled a naturally occurring *Xba*I site at –3736 bp relative to the *ETT* start codon. The resulting promoter DNA was fully sequenced in a plasmid cloning vector, released by treatment with *Xba*I, and inserted into the unique *Xba*I site of a plant transformation vector carrying a Gateway (Invitrogen) recombination cassette (Fourquin *et al.*, 2007), upstream of the nopaline synthase terminator. Wild-type, chimeric and truncated *ETT* and *ARF4* coding sequences from *pDONR207*-derived plasmids were then transferred to the resulting *proETT*-containing plant transformation vector by Gateway LR (Invitrogen) recombination reactions. The resulting plant transformation vectors were introduced into *Agrobacterium tumefaciens* strain *C58pmp90* and used to transform segregating populations of *ett-1* mutants by the floral-dip method (Clough and Bent, 1998). T_1 seed, harvested from dipped (T_0) plants, were selected on MS medium containing 8 mg l^{-1} of ammonium glufosinate (BASTA). The *ETT* locus of the BASTA-resistant T_1 plants selected was characterized by genomic PCR using the *ETT* coding sequence primers 5′-CTCGATGTTAAGCTTCACG-3′ and 5′-GCACTCCACCCGGTAGTGAGC-3′ and a T-DNA primer. Native *ETT* alleles and *ETT* transgenes were distinguished in this procedure by PCR product size, as the latter lacked introns.

ACKNOWLEDGEMENTS

We thank the technical staff of the RDP Laboratory, Dr Frédéric Pautz for access to the living collection of Lyon Botanic Garden, and Drs Jérome Munzinger, Marc Ducousso and Gildas Gateblé for help with the field collection of *A. trichopoda*. C. Finet and C. Fourquin received doctoral thesis funding from the French Ministry of Sciences. Our research group is supported by grant ANR-07-BLAN-0211-01 from the French National Research Agency.

SUPPORTING INFORMATION

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

Figure S1. Auxin response factor (ARF) nucleotide (a) and amino acid (b) sequence alignments used in the phylogenetic analyses shown in Figures 1 and S2.

Figure S2. Phylogenetic analyses of *ETT*- and *ARF4*-like genes in seed plants.

Figure S3. Northern blot hybridizations of *ETT* and *ARF4* orthologs from *Amborella trichopoda*, *Cabomba aquatica* and *Ephedra distachya*, showing comparable expression levels in vegetative and reproductive tissues.

Figure S4. *In situ* hybridization of *ETT* and *ARF4* orthologs.

Figure S5. Alternative splicing of exon 11 of *AtrARF4* in *Amborella trichopoda*.

Table S1. Accession numbers of *ETT* and *ARF4*-like sequences analyzed.

Table S2. Motifs associated with the post-transcriptional and translational regulation of *ETT* and *ARF4* orthologs.

Table S3. Numbers of Arabidopsis *ett-1* lines showing full, partial or no complementation following transformation with the wild-type, chimeric and truncated *ETT* and *ARF4* transgenes shown in Figure 3a.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such

materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Bremer, B., Bremer, K., Chase, M.W. *et al.* (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743.

De Bodt, S., Maere, S. and Van de Peer, Y. (2005) Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**, 591–597.

Dharmasiri, N., Dharmasiri, S. and Estelle, M. (2005) The F-box protein TIR1 is an auxin receptor. *Nature*, **435**, 441–445.

Dredge, B.K., Stefani, G., Engelhard, C.C. and Darnell, R.B. (2005) Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* **24**, 1608–1620.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 1–19.

Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C. and Scutt, C.P. (2005) Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proc. Natl Acad. Sci. USA*, **102**, 4649–4654.

Fourquin, C., Vinauger-Douard, M., Chambrier, P., Berne-Dedieu, A. and Scutt, C.P. (2007) Functional conservation between CRABS CLAW orthologues from widely diverged Angiosperms. *Ann. Bot.* **100**, 651–657.

Galtier, N., Gouy, M. and Gautier, C. (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**, 543–548.

Garcia, D., Collier, S.A., Byrne, M.E. and Martienssen, R.A. (2006) Specification of leaf polarity in Arabidopsis via the trans-acting siRNA pathway. *Curr. Biol.* **16**, 933–938.

Goff, S.A., Ricke, D., Lan, T.H. *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp japonica). *Science*, **296**, 92–100.

Guilfoyle, T.J. and Hagen, G. (2007) Auxin response factors. *Curr. Opin. Plant Biol.* **10**, 453–460.

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

Irimia, M., Rukov, J.L., Penny, D. and Roy, S.W. (2007) Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol. Biol.* **7**, 188–200.

Kepinski, S. and Leyser, O. (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature*, **435**, 446–451.

Kramer, E.M., Jaramillo, M.A. and Di Stilio, V.S. (2004) Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics*, **166**, 1011–1023.

Nishimura, T., Wada, T. and Okada, K. (2004a) A key factor of translation reinitiation, ribosomal protein L24, is involved in gynoecium development in Arabidopsis. *Biochem. Soc. Trans.* **32**, 611–613.

Nishimura, T., Wada, T. and Okada, K. (2004b) Upstream ORFs of ETTIN (ETT) and MONOPTEROS (MP) repress the expression of downstream ORFs. *Plant Cell Physiol.* **45**, S165–S165.

Nishimura, T., Wada, T., Yamamoto, K.T. and Okada, K. (2005) The Arabidopsis STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. *Plant Cell*, **17**, 2940–2953.

Pekker, I., Alvarez, J.P. and Eshed, Y. (2005) Auxin response factors mediate Arabidopsis organ asymmetry via modulation of KANADI activity. *Plant Cell*, **17**, 2899–2910.

Qiu, Y.L., Lee, J., Whitlock, B.A., Bernasconi-Quadroni, F. and Dombrovska, O. (2001) Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Mol. Biol. Evol.* **18**, 1745–1753.

Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sato, Y., Nishimura, A., Ito, M., Ashikari, M., Hirano, H.Y. and Matsuoka, M. (2001) Auxin response factor family in rice. *Genes Genet. Syst.* **76**, 373–380.

Sessions, R.A. and Zambryski, P.C. (1995) Arabidopsis gynoecium structure in the wild-type and in effin mutants. *Development*, **121**, 1519–1532.

Sessions, A., Nemhauser, J.L., McColl, A., Roe, J.L., Feldmann, K.A. and Zambryski, P.C. (1997) ETTIN patterns the Arabidopsis floral meristem and reproductive organs. *Development*, **124**, 4481–4491.

Tiwari, S.B., Wang, X.J., Hagen, G. and Guilfoyle, T.J. (2001) AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell*, **13**, 2809–2822.

Ulmasov, T., Hagen, G. and Guilfoyle, T.J. (1997) ARF1, a transcription factor that binds to auxin response elements. *Science*, **276**, 1865–1868.

Ulmasov, T., Hagen, G. and Guilfoyle, T.J. (1999a) Activation and repression of transcription by auxin-response factors. *Proc. Natl Acad. Sci. USA*, **96**, 5844–5849.

Ulmasov, T., Hagen, G. and Guilfoyle, T.J. (1999b) Dimerization and DNA binding of auxin response factors. *Plant J.* **19**, 309–319.

Wang, Z.F., Xiao, X.S., Van Nostrand, E. and Burge, C.B. (2006) General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell*, **23**, 61–70.

Zahn, L.M., King, H.Z., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., dePamphilis, C.W. and Ma, H. (2005) The evolution of the SEPALLATA subfamily of MADS-Box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics*, **169**, 2209–2223.

Zanis, M.J., Soltis, D.E., Soltis, P.S., Mathews, S. and Donoghue, M.J. (2002) The root of the angiosperms revisited. *Proc. Natl Acad. Sci. USA*, **99**, 6848–6853.

