MBE Advance Access published October 13, 2012

Evolution of the ARF Gene Family in Land Plants: Old Domains, New Tricks

Cédric Finet,*[‡,1] Annick Berne-Dedieu,[1] Charles P. Scutt,[1] and Ferdinand Marlétaz[2]

[1]Université de Lyon, Laboratoire de Reproduction et Développement des Plantes, UMR 5667, CNRS, INRA, Université Claude Bernard Lyon I, Ecole Normale Supérieure de Lyon, Lyon, France

[2]Department of Zoology, University of Oxford, Oxford, United Kingdom

[‡]Present address: Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, Madison, WI 53706

***Corresponding author:** E-mail: cedric.finet@ens-lyon.org.

Associate editor: Douglas Crawford

Abstract

Auxin response factors (ARF) are key players in plant development. They mediate the cellular response to the plant hormone auxin by activating or repressing the expression of downstream developmental genes. The pivotal activation function of ARF proteins is enabled by their four-domain architecture, which includes both DNA-binding and protein dimerization motifs. To determine the evolutionary origin of this characteristic architecture, we built a comprehensive data set of 224 ARF-related protein sequences that represents all major living divisions of land plants, except hornworts. We found that ARFs are split into three subfamilies that could be traced back to the origin of the land plants. We also show that repeated events of extensive gene duplication contributed to the expansion of those three original subfamilies. Further examination of our data set uncovered a broad diversity in the structure of ARF transcripts and allowed us to identify an additional conserved motif in ARF proteins. We found that additional structural diversity in ARF proteins is mainly generated by two mechanisms: genomic truncation and alternative splicing. We propose that the loss of domains from the canonical, four-domain ARF structure has promoted functional shifts within the ARF family by disrupting either dimerization or DNA-binding capabilities. For instance, the loss of dimerization domains in some ARFs from moss and spikemoss genomes leads to proteins that are reminiscent of Aux/IAA proteins, possibly providing a clue on the evolution of these modulators of ARF function. We also assessed the functional impact of alternative splicing in the case of *ARF4*, for which we have identified a novel isoform in *Arabidopsis thaliana*. Genetic analysis showed that these two transcripts exhibit markedly different developmental roles in *A. thaliana*. Gene duplications, domain rearrangement, and post-transcriptional regulation have thus enabled a subtle control of auxin signaling through ARF proteins that may have contributed to the critical importance of these regulators in plant development and evolution.

Key words: land plants, auxin pathway, truncation, alternative splicing, repeated evolution.

Introduction

The plant hormone auxin (indole-3-acetic acid) controls many physiological and developmental processes in land plants, such as organogenesis, vascular tissue differentiation, cell elongation, apical dominance, gravitropism, and embryo and root patterning (for review, see Kieffer et al. 2010). When applied at early stages, both auxin and chemical inhibitors of auxin transport severely affect the architecture of the embryo (Liu et al. 1993; Fischer and Neuhaus 1996). As the characteristic body plans of the different divisions of land plants are established during embryonic and early postembryonic development, auxin may act as a critical developmental signal in this process. Changes in auxin perception and signaling could therefore have generated the diversification of body plans that occurred during the evolution of the land plants (Cooke et al. 2002; Finet and Jaillais 2012).

To explore the interplay between auxin signaling and the evolution of land plants, we have focused on the evolution of the auxin response factor (ARF) proteins, which represent the core of auxin signaling (Chapman and Estelle 2009). According to transient assays and sequence analysis, ARF proteins were divided into transcriptional activators and repressors (Guilfoyle and Hagen 2007). ARF activators are transcription factors that mediate auxin-dependent transcriptional regulation by binding to auxin-response elements in the promoters of auxin-inducible genes in a dose-dependent manner (Ulmasov et al. 1997). In the absence of auxin, Aux/IAA proteins prevent ARF-mediated transcription by forming heterodimers with ARF activators (Ulmasov et al. 1999; Vernoux et al. 2011) and by recruiting corepressors of the TOPLESS family (Szemenyei et al. 2008). In the presence of auxin, Aux/IAA proteins are targeted to the proteasome by an SCF E3 ubiquitin ligase complex (Chapman and Estelle 2009), which can be hypothesized to release interacting ARF activators from inhibition. Contrary to ARF activators, ARF repressors have very limited interactions with other ARF and Aux/IAA proteins (Vernoux et al. 2011).

Though some auxin responses occur throughout the plant and are conferred by multiple members of the ARF and Aux/IAA families, others depend on the precise developmental context and involve the action of specific pairs of ARF and

Aux/IAA proteins (Weijers et al. 2005; Muto et al. 2007; De Rybel et al. 2010; De Smet et al. 2010), thus conferring the tissue-specific response to auxin. The domain architecture of ARF proteins plays a pivotal role in controlling their interaction network. In *Arabidopsis thaliana*, most ARF proteins consist of an amino-terminal DNA-binding region that includes both a B3 domain and an ARF domain, a variable middle region that confers activator or repressor activity, and a carboxy-terminal dimerization region (domains III and IV) involved in homo- and heterodimerization (Ulmasov et al. 1999; Tiwari et al. 2003) (fig. 2). These latter two domains are also essential for the interaction of ARFs with Aux/IAA proteins (Kim et al. 1997; Hardtke et al. 2004).

Given the key role that ARFs play in auxin signaling, as well as in activating tissue-specific developmental genes, we hypothesized that structural changes to ARF proteins could have played an essential role in the evolutionary diversification of land plants. Limited genome surveys of the ARF family have previously been performed in *Arabidopsis* (Remington et al. 2004), rice (Wang et al. 2007), poplar (Kalluri et al. 2007), tomato (Wu et al. 2011), and maize (Wang et al. 2012), but those studies failed to fully emphasize the molecular mechanisms underlying the structural diversity of ARF proteins. To address this neglected question, we therefore compiled a unique data set of 224 ARF proteins across all land plants from both available molecular resources and experimental screening of cDNA libraries. From this data set, we reconstructed the evolutionary history of the main ARF lineages and identified a large number of ARF proteins whose structure diverged from the canonical scheme by lacking one or several domains. Our findings suggest that not only changes in coding sequence but also rearrangements in domain organization made a substantial contribution to the diversity of ARF functions in the land plants.

Materials and Methods

Plant Material

Material of *Amborella trichopoda* was field collected from locations near Col d'Amieu, New Caledonia. Material of *Cabomba aquatica* was obtained from Anthias S.A., Les Chères 69, France. Material of *Illicium parviflorum*, *Ephedra distachya*, *Ginkgo biloba*, and *Saruma henryi* was collected from plants cultivated in the Botanical Garden "Parc de la Tête d'Or," Lyon, France. The *A. thaliana* ecotype Wassilewskija-2 (Ws-2) was used as the wild-type reference. Seeds of the *arf3-1* mutant in the Ws-2 genetic background were obtained from the Nottingham *Arabidopsis* Stock Centre (UK). Plants were grown to maturity in peat-based compost in a growth chamber at 20°C under 16 h light/8 h dark cycles.

Transgenic Plants

The full *ARF3* and *ARF4* coding sequences from *A. thaliana* and an isoform of ARF4 containing a deletion due to alternative splicing, *isoform*Δ(98–169)*ARF4* or *ΔARF4*, were amplified by high-fidelity polymerase chain reaction (PCR) from a wild-type young inflorescence cDNA sample. Amplified products were inserted by recombination into a derivative of the pCAMBIA3300 binary vector (Finet et al. 2010) between a DNA fragment of 3,703 bp representing the *ARF3* promoter and the nopaline synthase terminator. Constructs were introduced into *Agrobacterium* strain C58pmp90 by electroporation and the resulting strains used to transform heterozygous *Arabidopsis* plants carrying the *arf3-1* mutation by floral dipping (Clough and Bent 1998).

Selection of Transformants

Homozygous mutant *arf3-1* plants were distinguished from wild-type and heterozygous *arf3-1*/+ plants by PCR on genomic DNA, which generated 1,066-bp fragment for wild-type or heterozygous plants, and no amplified band, due to the insertion of a T-DNA, for *arf3-1* plants. We selected BASTA-resistant T_1 arf3-1 homozygotes, which were thus hemizygous for inserted transgenes, and used these for analysis in the T_2 generation during which genotypes were confirmed by monitoring Mendelian ratios (Finet et al. 2010).

Data Collection

ARF genes were identified in several complete genomes by TBLASTN using a set of selected *A. thaliana* ARFs as a probe. *Arabidopsis thaliana*, *Brachypodium distachyon*, *Carica papaya*, *Citrus clementina*, *Cucumis sativus*, *Eucalyptus grandis*, *Glycine max*, *Manihot esculenta*, *Medicago truncatula*, *Mimulus guttatus*, *Oryza sativa*, *Populus trichocarpa*, *Physcomitrella patens*, *Prunus persica*, *Ricinus communis*, *Selaginella moellendorffii*, *Setaria italica*, *Sorghum bicolor*, *Vitis vinifera*, and *Zea mays* genomes were retrieved from JGI website (http://genome.jgi-psf.org, last accessed 2012 September 25). We also included relevant ARF sequences from expressed sequence tag (EST) databases such as NCBI and TIGR (supplementary table S1, Supplementary Material online). Species of origin are indicated by the following prefixes: Aca, *Adiantum capillis-veneris*; Ace, *Allium cepa*; Afo, *Aquilegia formosa* × *pubescens*; Atr, *Amborella trichopoda*; Ath, *Arabidopsis thaliana*; Afi, *Aristolochia fimbriata*; Bna, *Brassica napus*; Caq, *Cabomba aquatica*; Cri, *Ceratopteris richardii*; Csi, *Citrus sinensis*; Cor, *Coleochaete orbicularis*; Csa, *Cucumis sativus*; Cru, *Cycas rumphii*; Edi, *Ephedra distachya*; Gbi, *Ginkgo biloba*; Ggn, *Gnetum gnemon*; Gar, *Gossypium arboreum*; Gba, *Gossypium barbadense*; Ghi, *Gossypium hirsutum*; Gra, *Gossypium raimondii*; Ipa, *Illicium parviflorum*; Lsa, *Lactuca sativa*; Ltu, *Liriodendron tulipifera*; Min, *Mangifera indica*; Mpo, *Marchantia polymorpha*; Mtr, *Medicago truncatula*; Mac, *Musa acuminata*; Nbe, *Nicotiana benthamiana*; Nad, *Nuphar advena*; Osa, *Oryza sativa*; Pam, *Persea americana*; Ppr, *Phyllostachys praecox*; Ppa, *Physcomitrella patens*; Psi, *Picea sitchensis*; Ppi, *Pinus pinaster*; Pta, *Pinus taeda*; Pju, *Polytrichum juniperinum*; Ptr, *Populus trichocarpa*; Ppe, *Prunus persica*; Smo, *Selaginella moellendorffii*; Ses, *Solanum esculentum*; Sly, *Solanum lycopersicum*; Stu, *Solanum tuberosum*; Tae, *Triticum aestivum*; Vvi, *Vitis vinifera*; Yfi, *Yucca filamentosa*; Zfu, *Zamia furfuracea*; and Zma, *Zea mays*.

Molecular cloning of ARF cDNAs was performed in clades for which few genomic data were available by using flower

cDNA libraries of *A. trichopoda* and *C. aquatica* that were previously made in the laboratory (Fourquin et al. 2005), a similarly prepared library from flowers of *I. parviflorum* and reproductive tissue libraries from *E. distachya* and *G. biloba* (Finet et al. 2010). Library screening was performed as described by Fourquin et al. (2005). The sequences reported in this article have been deposited in the EMBL database with the following accession numbers: FN433169 (*A. trichopoda ARF2*), FN433170 (*A. trichopoda ARF6*), FN433171 (*A. trichopoda ARF8*), FN433172 (*C. aquatica ARF2*), FN433173 (*C. aquatica ARF6*), FN433174 (*C. aquatica ARF8*), FN433175 (*I. parviflorum ARF1*), FN433176 (*I. parviflorum ARF3*), FN433177 (*I. parviflorum ARF8*), FN433178 (*E. distachya ARF6/8*), FN433179 (*G. biloba ARF6/8*), FN433180 (*Gne. gnemon ARF5/7*), FN433181 (*Cyc. rumphii ARF3/4*), FN433182 (*Cyc. rumphii ARF2/1/9*), FN433183 (*Cyc. rumphii ARF10/16/17*), and FN433184 (*P. pinaster ARF3/4*).

Prediction of Secondary Structure

The secondary structure of the novel motif identified in the middle region of ARFs was predicted using the MLRC method (Guermeur et al. 1999).

Detection of Putative uORFs

The detection of putative upstream open reading frames (uORFs) required the experimental characterization of the 5′-UTR of some ARF genes in *A. thaliana* and in the moss *Phy. patens*. RNA was extracted from tissues of *Phy. patens* using a Qiagen RNeasy Plant Extraction Kit and from *A. thaliana* tissues by using TRIzol reagent (Invitrogen). We performed 5′ RACE by using a MARATHON cDNA amplification kit (Clontech, Palo Alto, CA, USA) with the following specific primers: *AthARF10* 5′-TCGTCCCAAGCTACCTGAAGGAGA, *AthARF16* 5′-GAACCATACCACCAGCACATGCAT, *AthARF17* 5′-ATCCCATGTGATCTGAAGCTGCTT, *PpaARF10/16/17.1* 5′-CTTGGGGTCCCCCTGTATATGTGC, and *PpaARF10/16/17.2* 5′-AGTCCCCTTGAAAAGCCTCCACCGT. The RACE products were finally purified (Montage DNA Gel Extraction Kit, Millipore) and sequenced with traditional Sanger methodology.

Phylogenetic Analysis

Amino acid sequences were aligned with MUSCLE (Edgar 2004), manually adjusted, and selected blocks were used for phylogenetic reconstruction. Maximum-likelihood (ML) searches were performed using RAxML 7.2.8 (Stamatakis 2006) under the LG matrix with CAT optimization of site substitution rates and final likelihood evaluation using a gamma distribution. A total of 1,000 bootstrap replicates were conducted for support estimation. To detect potential impact of long-branch attraction, we also employed a site-heterogeneous model (mixture of profile C20) implemented in PhyloBayes 3.2 (Lartillot et al. 2009). We ran two independent chains for at least 21,000 cycles, discarded the first 500 cycles as burn-in, and recovered no major topological changes with respect to ML tree (supplementary fig. S1, Supplementary Material online). The different protein sequence alignments and tree files are downloadable from Dryad (http://dx.doi.org/10.5061/dryad.c8335, last accessed 2012 September 25).

Statistical Analysis

The number of sepals (Se), petals (P), and stamens (St) was counted in *arf3-1*, p*ARF3*::*ARF4*, and p*ARF3*::*ΔARF4* lines of *A. thaliana*. One hundred flowers per line were collected from several plants. Within each line, we tested the plant effect using a Kruskal–Wallis test, as the number of organs did not follow a normal distribution. The plant effect was not statistically significant, and thus the data were pooled within each line for next analyses. The effect of genetic background on the number of Se, P, or St was estimated by a Kruskal–Wallis test, followed by post hoc tests (pairwise Wilcoxon tests) with Bonferroni correction. Statistical tests and graphics were performed using R statistics package version 2.14.1 (the R Project for Statistical Computing, www.r-project.org, last accessed 2012 September 25).

Results and Discussion

Phylogenetic Analysis Reveals the Ancestral ARF Complement

The analysis of ARF family evolution is clearly essential to understanding the contribution that auxin signaling has made to land plant evolution. However, the large size of ARF family and the functional redundancy between its members have made it difficult to understand its evolutionary origin. To tackle this question, we employed a large-scale phylogenomic approach relying on an unprecedented sampling of ARF sequences from most relevant taxonomic lineages. We assembled this data set using three complementary approaches. First, we applied an exhaustive BLAST similarity search to a set of available plant genomes, which recovered ARF sets consistent with previous genome annotations (table 1 and supplementary fig. S2, Supplementary Material online). We then enhanced our sampling by incorporating sequences assembled from several sets of transcriptomic data from lineages with no sequenced genomes (e.g., early diverging angiosperms such as magnoliids). Finally, we used degenerate PCR to screen for ARF sequences in pivotal taxa with scarce genomic data including two gymnosperms (*E. distachya* and *G. biloba*) and three basal angiosperms belonging to the ANA grade (*A. trichopoda*, *C. aquatica*, and *I. parviflorum*).

We performed ML and Bayesian phylogenetic reconstruction using an alignment composed of the four conserved domains (B3, ARF, III, and IV) on our 224-ARF data set. The tree obtained clarifies the number of main ARF lineages and the relationships between these (fig. 1 and supplementary fig. S1, Supplementary Material online). In this tree, ARF sequences split into three main clades: clade A (including ARF5, ARF6, ARF7, and ARF8), clade B (including ARF1, ARF2, ARF3, ARF4, and ARF9), and clade C (including ARF10, ARF16, and ARF17) (fig. 1 and supplementary fig. S1, Supplementary Material online). Because they include bryophyte sequences, clades A and C can probably be

Table 1. Summary of ARF Gene Content in Complete Angiosperm Genomes.

Species	ARF Content	Genome	References
Aquilegia coerulea	12	Early	This study
Arabidopsis thaliana	23	v. 9.0	Hagen and Guilfoyle (2002)
Brachypodium distachyon	24	v. 1.0	This study
Carica papaya	15	ND	This study
Citrus clementina	17	v. 0.9	This study
Cucumis sativus	16	ND	This study
Eucalyptus grandis	15	v. 1.0	This study
Glycine max	23	v. 1.0	This study
Manihot esculenta	18	v. 4.1	This study
Medicago truncatula	15	v. 3.0	This study
Mimulus guttatus	19	v. 1.0	This study
Oryza sativa	25	v. 6.0	Wang et al. (2007)
Physcomitrella patens	13	v. 1.6	Rensing et al. (2008)
Populus trichocarpa	39	v. 2.0	Kalluri et al. (2007)
Prunus persica	17	v. 1.0	This study
Ricinus communis	17	v. 0.1	This study
Selaginella moellendorffii	10	v. 1.0	Banks et al. (2011)
Setaria italica	23	v. 2.0	This study
Sorghum bicolor	22	v. 1.0	This study
Vitis vinifera	19	ND	This study
Zea mays	22	v. 2.0	Wang et al. (2012)

traced back to the origin of land plants, whereas clade B is found only in lycophytes and subsequently occurring plant lineages. Furthermore, we detected a set of four *Phy. patens* proteins that are loosely associated with clade A. These sequences could constitute an alternative ARF lineage, specific to mosses, that remains from an early ARF diversification or, alternatively, result from the misplacement of a subset of fast evolving genes. In particular, we checked that phylogenetic trees reconstructed using each ARF domain separately were congruent with the combined topology (supplementary fig. S3, Supplementary Material online). We were unable to detect the ARF domain in several charophytes, the closest aquatic relatives of land plant, despite the occurrence of domains B3, III, and IV (De Smet et al. 2011), which very strongly suggests that the ARF domain is a land plant innovation. Furthermore, we may conclude with high confidence that the last common ancestor of the extant land plants possessed at least two ARF genes, corresponding to the precursors of the A and C clades, though the status of the B clade is less clear. The deep origin of many ARF subfamilies thus indicates an early diversification of ARF genes, although it is still difficult to estimate how secondary gene loss has affected the observed ARF diversity.

Additionally, it is noteworthy that ARF activators (ARF5, ARF6, ARF7, ARF8, and ARF19 in *Arabidopsis*) are all clustered in the A clade, whereas ARF repressors are split into the B and C clades. It has been previously reported that the middle regions are enriched for glutamine residues in *A. thaliana* ARF activators (Ulmasov et al. 1999). We comprehensively searched for Q-rich sequences in middle regions in all ARF sequences, and we found that glutamine enrichment seems to be a distinctive feature of ARF activators in all plant divisions (fig. 1 and supplementary table S2, Supplementary Material online). Taken together, these results suggest that the ancestor of the extant land plants possessed at least one ARF activator and one ARF repressor.

Angiosperm ARF Repertory Originated through Gene Duplications

Although the main three A, B, and C ARF lineages have an ancient origin, their respective content has been shaped by many lineage-specific duplication events. Early duplications took place in clades B and C of euphyllophytes (monilophytes and seed plants) to establish the ARF1/2/9, ARF3/4, ARF5/7, and ARF6/8 lineages (fig. 1). Functional characterization of ARF genes in lycophytes and monilophytes could provide insights into any new functions gained following these gene duplication events.

The structure of the ARF family in seed plants has been mostly shaped by duplications that predated the angiosperm radiation. Indeed, our phylogenetic tree indicates that 9 in 13 present-day subfamilies appeared after the divergence of gymnosperms. In particular, the ARF sequences from ANA grade species suggest that these subfamilies were established before the initial radiation of the angiosperms. For instance, the ARF6 and ARF8 subfamilies each include sequences from ANA grade and other angiosperms (fig. 1 and supplementary fig. S1, Supplementary Material online), but exclude their gymnosperm genes. Gymnosperm genes grouping closely to both the angiosperm ARF6 and ARF8 clades are probable pro-orthologs of these duplicated angiosperm genes.

Interestingly, the duplication pattern observed in the ARF family coincides remarkably well with the occurrence of whole-genome duplications during the evolution of seed

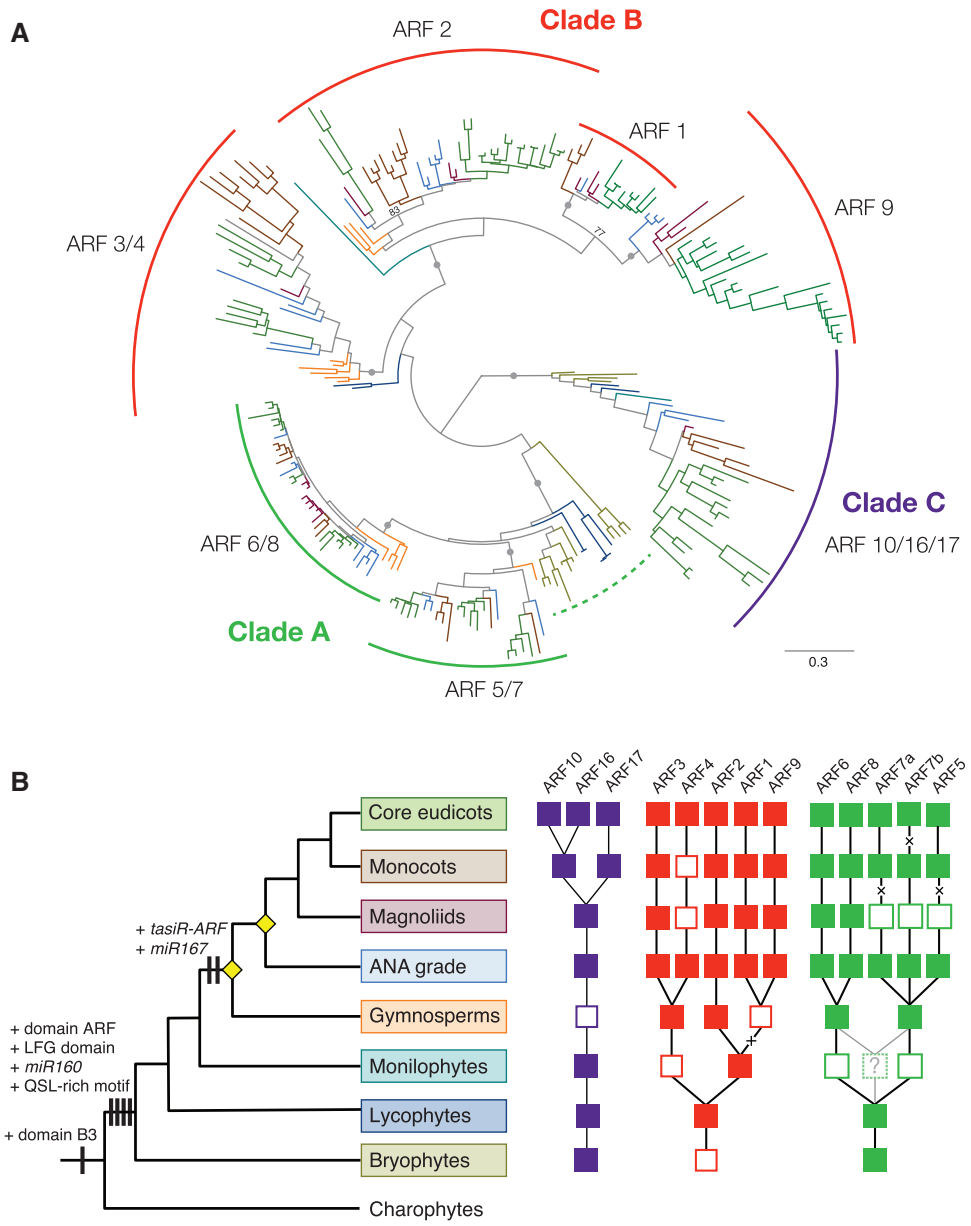

FIG. 1. Phylogeny and structure of ARF genes in land plants. (*A*) Phylogram of the 224-taxon analyses obtained through maximum-likelihood and Bayesian analyses were conducted using LG+Γ_4 and C20 mixture, respectively. The selected blocks used for phylogenetic reconstruction encompass domain B3, domain ARF, and domains III and IV. Support values are shown for selected nodes (bootstrap replicates BP, all support values in supplementary fig. S1, Supplementary Material online). A dot indicates support values of BP > 95. Scale bar indicates number of changes per site. (*B*) Evolutionary origin of the main structural and regulatory ARF domains. Filled squares indicate the presence of genomic data, open squares indicate lack of data, dotted lines indicate alternative hypothesis, and yellow diamonds indicate a whole-genome duplication event (according to Jiao et al. 2011).

plants (Jiao et al. 2011). Subsequent extensive genome rearrangements during angiosperm evolution prevent us from obtaining definite evidence that these ARF duplications specifically relate to given single whole-genome duplication events. However, the parallel duplications that occurred in several different ARF subfamilies before the radiation of the extant angiosperms make it likely that the preangiosperm duplication termed ε played a major role in establishing the diversity of angiosperm ARF genes (fig. 1).

It is tempting to consider the functional diversification of ARF subfamilies among angiosperms in relation to the evolutionary innovations characteristic of this division. First, both *arf6* and *arf8* single mutants show delayed flower maturation and subtly reduced fertility, whereas flower development in *arf6*/*arf8* double mutants is arrested before maturity, resulting in complete sterility (Nagpal et al. 2005). In addition to its redundant role with ARF6 in flower development, ARF8 plays an essential role in coupling fertilization and fruit development (Goetz et al. 2006). Although ARF6 and ARF8 act largely redundantly, these two genes may have evolved new functions related to the advent of floral organs.

Similarly, the clades ARF3 and ARF4 originated from a preangiosperm gene duplication (Finet et al. 2010), and both duplicates underwent functional divergence. In

Arabidopsis, single loss-of-function *arf3/ett* mutants show severe defects in the establishment of the polarity in the gynoecium (Sessions and Zambryski 1995), whereas no mutant phenotype has been observed in single *arf4* mutants. The *arf3/ett* phenotype is enhanced in *arf3 arf4* double mutants, which exhibit reduced abaxial identity in all lateral organs, including leaves (Pekker et al. 2005). These results indicate that the paralogs ARF3 and ARF4 are involved in the same developmental pathway. Last, both ARF5/MP and ARF7/NPH4 contribute to cotyledon development, whereas only MP is required for embryonic root initiation (Hardtke et al. 2004). By considering phylogenetic relationships together with available genetic data, these two latter examples illustrate how lineage-specific duplications may have played a part in triggering the functional diversification of ARF genes within land plants, with some duplicates involved in the development of novel organs such as the carpel.

Genomic Truncations and Alternative Splicing Enhanced the Functional Repertoire of ARF Proteins

We have found that gene duplication has played a major role in the functional diversification of the ARF family. However, we have also observed marked differences in domain organization when considering the protein sequence alignment of the ARFs collated for this study. We decided to check whether these variations, particularly including losses of domains III and IV, correspond to incomplete transcript recovery or to the faulty annotation of coding sequences. We aligned available transcripts to the reference genomic sequence where available, which demonstrated that alternative domain organization does indeed reflect the genomic sequence in most cases. Accordingly, four members of the ARF family—ARF3, ARF13, ARF17, and ARF23—are truncated at the genomic level in *A. thaliana* and as a result do not encode domains III and IV. In spite of these truncations, it has been shown that these proteins are all functional, except for ARF23, which appears to be a pseudogene (Guilfoyle and Hagen 2001).

However, we also found sets of transcripts matching the same genomic sequence but presenting multiple deletion patterns. We searched for canonical splice sites recognition GT/AG at the ends of the deletions, and we checked whether each deletion corresponded precisely to an intron/exon junction in *A. thaliana*, assuming the conservation of intron positions between orthologs. Using this method, we identified that alternative splicing, a mechanism previously found to operate on the ortholog of ARF4 in *A. trichopoda* (Finet et al. 2010), operates more widely in the ARF family.

To understand the potential evolutionary significance of both genomic truncations and alternative splicing, we comprehensively inspected our data set for the occurrence of these phenomena (fig. 2). Genomic truncations are restricted to clades B and C but occur independently in diverse taxa and diverse ARF subfamilies belonging to these two ARF lineages (fig. 2). For example, three sporadic occurrences of truncation events are apparent in clade C: one in a gene from *Phy. patens* (*Ppa ARF10/16/17.2*), one in a gene from *Pop. trichocarpa* (*Ptr ARF16.5*), and the other one in all *ARF17* orthologs in flowering plants. On the contrary, alternative splicing is of widespread occurrence in all ARF subfamilies, and the distribution and nature of the different spliced variants present are not explained by the phylogeny (fig. 2). The multiple exon loss patterns reported are also consistent independent origins of spliced variants (fig. 2). For instance, this mechanism can generate some transcription factors without the domains III and IV involved in dimerization (*A. trichopoda* ARF4, *Pop. trichocarpa* ARF2.4, and *Pop. trichocarpa* ARF9.3), without the repressor domain (*S. lycopersicum* ARF2), without a serine-rich region in the N-terminal end (*A. thaliana ARF4*), or proteins for which the ARF domain is not entire (*G. biloba* ARF6/8).

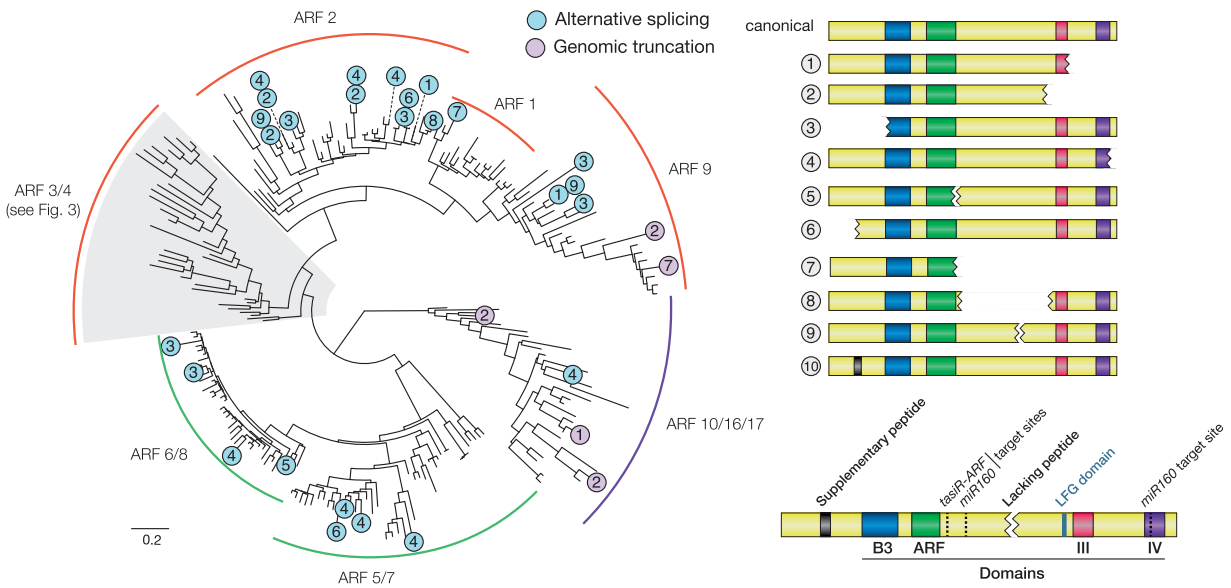

Fig. 2. Structural diversity of ARF proteins in land plants. (Left panel) Identified cases of alternative splicing (green) and genomic truncation (violet) are mapped onto the ARF phylogenetic tree obtained in figure 1. (Right panel) Variants of ARF protein structures are predicted from main variants of ARF transcripts.

We have thus uncovered a high diversity of domain architecture in ARF transcripts that is generated by genomic truncation and alternative splicing. In the next paragraphs, we will investigate the functional implications of domain losses through selected examples.

Serial Modifications of Domain Architecture: The Cases of ARF 3/4 and AuxIAAs

To better evaluate the functional and evolutionary importance of domain rearrangement, we focused on ARF3 or ARF4 subfamilies. In particular, ARF3 and ARF4 have become modified to encode truncated ARF proteins in two representatives of the early-diverging ANA grade (*A. trichopoda* and *C. aquatica*) and in the remaining angiosperm clade (Finet et al. 2010). To clarify the origin of these truncations, we performed a phylogenetic reconstruction of the clades ARF3 and ARF4 based on an improved taxonomic sampling (fig. 3). The cloned cDNA of *I. parviflorum* ARF3 encodes an ARF protein containing the domains III and IV, similar to ARF3 orthologs in *A. trichopoda* and *C. aquatica* (Finet et al. 2010). This finding reinforces the idea that the protein ARF3 had probably the domains III and IV in the last common ancestor of extant angiosperms. In magnoliids, the structure of the different ARF3 orthologs is highly variable, ranging from the presence of the two domains III and IV (*L. tulipifera*) to their complete absence (*Persea americana*), as well as intermediate forms that only have the domain III (*Sar. henryi*). The latter sequence, originally retrieved from an EST database, was confirmed by reverse transcription-PCR amplification from an independent RNA extraction.

The clade ARF3 underwent a complex evolutionary path in monocots. Poales contain two distinct ARF3 subclades: one of these comprises ARF3 proteins that lack the domains III and IV, whereas the other contains ARF3 proteins that lack domain IV (but which contain domain III). This latter intermediate form of the ARF3 C-terminal region is similar to the partially truncated transcript found in the magnoliid *Sar. henryi*. Interestingly, the transcript structure of the first group (lacking both domains III and IV) is also achieved through alternative splicing of genes from the second group in *O. sativa* and *Z. mays*. The most parsimonious evolutionary scenarios would be to consider that both the domains III and IV (or at least the domain III) were already present in the last common ancestor of extant monocots, domain III being conserved only in one ARF3 subclade within the Poales (fig. 3). At least two independent events of truncation led to the loss

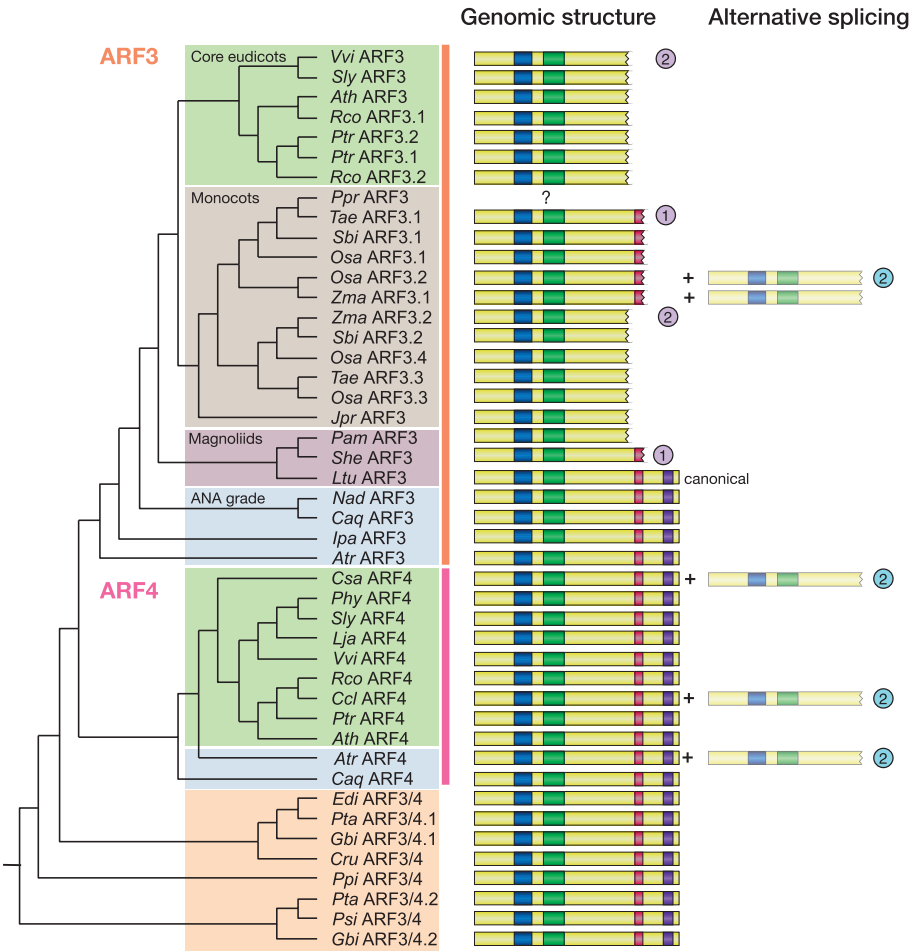

Fig. 3. Evolutionary history of truncations in the clade ARF3/4. Truncated ARF3 and ARF4 proteins are mapped onto a phylogenetic tree of the clade ARF3/4. The ML analysis was conducted under the WAG+Γ+I_4 model, and support values were obtained after 100 bootstrap replicates. The selected blocks used for phylogenetic reconstruction encompass domain B3, domain ARF, and domains III and IV.

of the domain IV of ARF3 during the evolution of flowering plants, both within the magnoliids and in the common lineage leading to monocots and core eudicots.

Interestingly, the latter example might help us to understand the origin of Aux/IAAs. These proteins are the main repressors of ARF activity in the nucleus. They possess domains III and IV that interact with the domains III and IV of ARF proteins in the absence of auxin, resulting in the repression of ARF activity (Ulmasov et al. 1997). The domains III and IV of Aux/IAA and ARF proteins present a degree of sequence conservation indicative of a shared evolutionary origin. However, it remains unresolved whether the Aux/IAA family arose by the loss of the DNA-binding domain in an ancestral ARF protein or the ARF family arose by the addition of a DNA-binding domain to domains III and IV of an ancestral Aux/IAA protein. We discovered a novel class of genes in *Phy. patens* and *Sel. moellendorffii*, which lack B3 and ARF domains but which possess domains III and IV that are clearly more related to those of ARF proteins than to Aux/IAA ones (supplementary fig. S4, Supplementary Material online). Moreover, canonical signatures of Aux/IAA domains I and II were not detected in this new class of genes. These examples of domain losses in genes from *Phy. patens* and *Sel. moellendorffii* suggest a possible mechanism for the origin of Aux/IAA regulators.

The numerous and independent losses of domains III and IV that have occurred during land plant evolution raise the question of the biological consequences of such changes. Truncated ARFs, which lack domains of interaction with Aux/IAAs, should consequently be insensitive to auxin (Finet et al. 2010). Although relevant for ARF activators, this hypothesis seems unlikely to be relevant for ARF repressors, which have limited interactions with Aux/IAA proteins (Vernoux et al. 2011). Loss of domains III and IV could also have consequences on the interaction of ARFs with other transcription factors. For instance, it has been recently shown that ARF7 and ARF8 can interact, by their domains III and IV, with MYB77 (Shin et al. 2007) and the bHLH factor BIGPETALp (Varaud et al. 2011), respectively. Noteworthy, loss of the functional domain I motif (LxLxL) that confers transcriptional repressor function of Aux/IAA occurred independently several times during evolution of land plants (Paponov et al. 2009).

Distinct Functions for Two Splicing Variants of *ARF4* in *A. thaliana*

Identification of numerous alternative splice variants among ARF transcripts suggests that the functional diversity of ARF proteins could have been underestimated. In other words, our findings raise the question as to whether noncanonical ARF transcripts code for proteins with previously undiscovered biological functions. Here, we test the biological relevance of the existence of two isoforms of the *ARF4* mRNA, *ARF4* and *ΔARF4* (fig. 4). First, the *ΔARF4* variant quantitatively represents one-third of total *ARF4* transcripts in *Arabidopsis* inflorescences (data not shown), challenging the idea that the *ΔARF4* variant could be an aberrant transcript. Second, this alternative splicing leads to the loss of an S-rich region, which is widely conserved between ARF4 orthologs, suggesting a potential divergent role for the *ΔARF4* transcript form.

We designed an experiment to indirectly distinguish the functions of the two isoforms *ARF4* and *ΔARF4*. Mutants of *arf4* do not exhibit visible phenotypes in *Arabidopsis* (Pekker et al. 2005), rendering direct complementation assays of *arf4* mutants infeasible. Consequently, we took advantage of the fact that a *pARF3::ARF4* transgene was found to partially restore the wild-type phenotype of transformed *arf3-1* mutants in a T_2 population of 40 plants (Finet et al. 2010). Carpels and siliques of these transformants were completely closed, but they harbored enlarged replum and style tissues (Finet et al. 2010). In this study, we performed a similar experiment with the variant form *ΔARF4*. We found that a *pARF3::ΔARF4* transgene dramatically enhanced the *arf3-1* mutant phenotype in that defects in carpel polarity and fusion were accompanied by a variation in the number of floral organs (fig. 4*A*). There is a general significant effect toward expansion of sepal (Kruskal–Wallis: $K = 98.4$, $P < 2.2e-16$; Wilcoxon, *ΔARF4*/*ARF4*: $P < 2.2e-16$, *ΔARF4*/*arf3-1*: $P < 4e-14$) and petal number (Kruskal–Wallis: $K = 22.5$, $P < 1.3e-5$; Wilcoxon, *ΔARF4*/*ARF4*: $P < 6.5e-4$, *ΔARF4*/*arf3-1*: $P < 4e-4$), and reduction of stamen number (Kruskal–Wallis: $K = 91.1$, $P < 2.2e-16$; Wilcoxon, *ΔARF4*/*ARF4*: $P < 2e-16$, *ΔARF4*/*arf3-1*: $P < 1.9e-7$). Moreover, sepals and petals were often narrower than in the wild type, as has been previously described for NPA treatment of wild-type flowers in *A. thaliana* (Nemhauser et al. 2000).

This experiment suggests different functional roles for two isoforms of the *ARF4* mRNA, *ARF4* and *ΔARF4*, during carpel development in *A. thaliana*. Further experiments will be necessary to directly tackle the biological function of the variant *ΔARF4*.

ARF Regulatory Processes Can Be Traced to Lower Embryophytes

We have extensively explored the diversification of ARF genes in terms of gene duplication and domain architecture. However, protein function is known to not always play a major part in functional diversification during evolution, and regulation should be thoroughly considered (Alonso and Wilkins 2005; Carroll 2008). ARF gene activity is known to be finely tuned by several regulatory processes during plant development, but very little is known about the evolutionary origin of these regulatory mechanisms. We have therefore examined the molecular clues of the presence of such regulatory sites in our 224-protein data set (supplementary table S2, Supplementary Material online).

Trans-Acting Short-Interfering RNAs

ARF2, *ARF3/ETT*, and *ARF4* transcripts have been identified as targets of an endogenous trans-acting short-interfering RNAs (*tasiR-ARFs*), which guides their cleavage in a similar fashion to microRNAs (Williams et al. 2005). The conservation of *tasiR-ARFs* and their target genes between *Arabidopsis*, rice, and maize has prompted the hypothesis that this

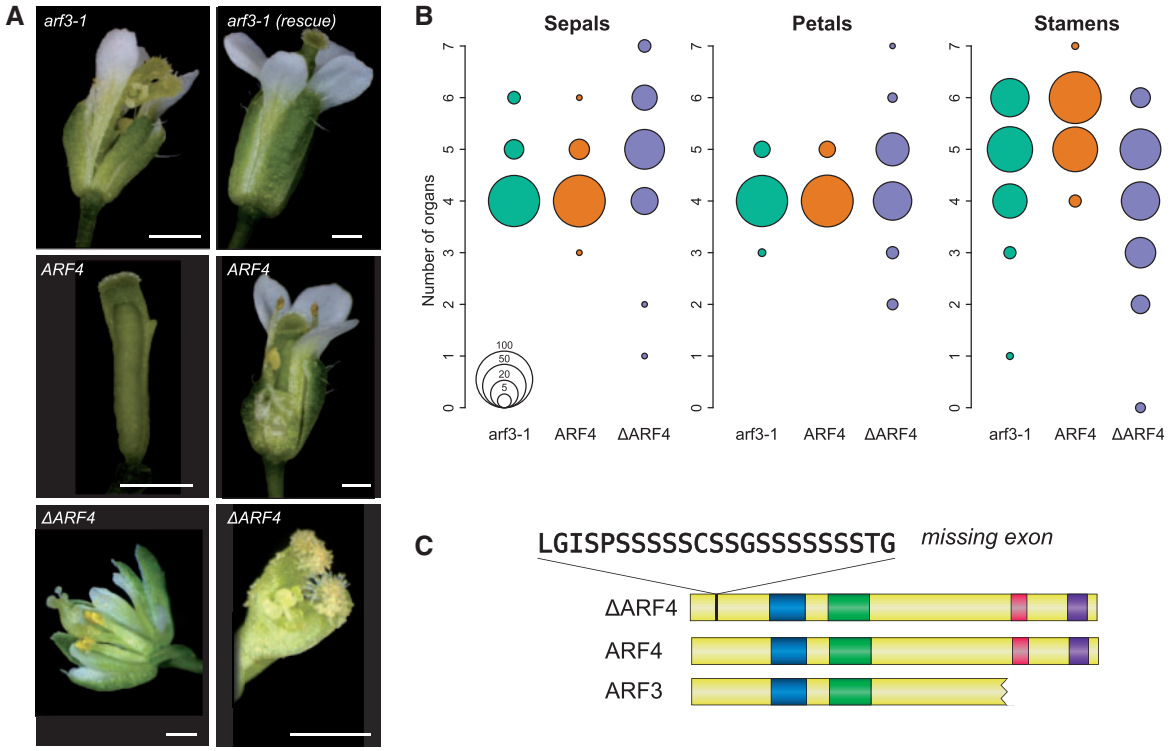

Fig. 4. Complementation of the *arf3-1* mutation by transformation with two different isoforms of *AtARF4* under the control of ARF3 promoter. (*A*) Variation in the number of sepals, stamens, and petals in *arf3-1* mutants and *arf3-1* mutants transformed with the coding sequence of *ARF4* or *isoform Δ(98-169)ARF4*. (*B*) Schematic representation of the two ARF4 transcripts generated by alternative splicing in *A. thaliana*. (*C*) Floral phenotypes of *arf3-1* mutants obtained before or after genetic transformation with the two isoforms of *ARF4*.

post-transcriptional regulation arose before the separation of the monocot and eudicot lineages (Williams et al. 2005). We identified *tasiR-ARF* target sites in orthologs of *ARF2*, *ARF3*, and *ARF4* genes in the major groups of angiosperms, including the ANA grade, the monocot, and eudicot lineages. In addition, *tasiR-ARF* target sites are also found in the *ARF3/4* and *ARF2/1/9* genes of several gymnosperm species, whereas no conserved sites have been identified in B clade homologs in mosses and spikemosses. This pattern indicates that regulation by *tasiR-ARF* likely arose in the lineage leading to the seed plants. Strikingly, *tasiR-ARF* target sites are absent from the ARF1/9 clade in angiosperms. Given that ARF1/9 and ARF2 are sister groups, *tasiR-ARF* target sites were probably lost from the ARF1/9 clade in a common ancestor of the extant flowering plants (fig. 2).

microRNAs

In *Arabidopsis*, *ARF6* and *ARF8* are targets for *miR167*, whereas *ARF10*, *ARF16*, and *ARF17* are targets for *miR160* (Rhoades et al. 2002). We have identified *miR160* target sites in all ARF genes of the clade C (supplementary table S2, Supplementary Material online). Moreover, our analysis reveals the presence of predicted *MIR160* genes in *Physcomitrella* and *Selaginella* genomes, corroborating the hypothesis that C-clade genes were regulated by *miR160* in the last common ancestor of the extant land plants (fig. 2). On the contrary, *miR167* regulation seems to be restricted to seed plants. We identified *miR167* target sites in the *ARF6/8* orthologs among a wide sampling of seed plants, but found no evidence of its presence in those of bryophytes and lycophytes. In agreement with this observation, the *MIR167* gene has only been found in seed plants and is not present in mosses, lycopods, or ferns (Axtell and Bartel 2005). Moreover, we found no miR167 target sites in genes of the ARF5/7 clade, the sister clade to the ARF6/8 clade, nor in putative ARF orthologs in mosses, indicating that the regulation by *miR167* may have specifically appeared in the ARF6/8 lineage. Interestingly, angiosperms and gymnosperms have evolved sporophytic structures surrounding gametophytes that consist of integuments (future seed integuments) in females and sterile cells in males. A recent study has shown that *miR167* is involved in the proper development of these outer structures because mutations in the *miR167* target sites of *ARF6* or *ARF8* in *A. thaliana* cause arrested growth of ovule integuments and anther sterile tissues (Wu et al. 2006). Thus, the evolutionary appearance of *miR167* regulation correlates with the evolutionary origin of the seed.

uORFs

uORFs are small ORFs in the 5′-leader sequence of a mature mRNA, which can mediate translational regulation of the major ORF. If uORFs are recognized by a ribosome scanning the mRNA, translation will be terminated at the stop codon of the uORF, and translation of the downstream ORF will require the reinitiation of translation (Kozak 1992). In *A. thaliana*, some uORFs have been predicted in the 5′-leader

sequence of numerous ARF transcripts, and the occurrence of peptides translated from the uORFs was experimentally confirmed (Nishimura et al. 2004). In particular, *ARF3* and *ARF5* have uORFs in their 5′-leader sequence that negatively regulate the translation of the main ORF (Nishimura et al. 2005). We identified uORFs in most of the ARF clades, which suggests the relevance of this regulatory mechanism on a macroevolutionary scale. Unfortunately, it is difficult to determine the presence of uORFs in clade C due to a lack of data, and this is the case even for *Arabidopsis ARF10*, *ARF16*, and *ARF17*, in which the start of transcription has not yet been determined. We have therefore characterized the 5′-UTR regions of *ARF10*, *ARF16*, and *ARF17* in *Arabidopsis* and *PpaARF10/16/17.1* and *PpaARF10/16/17.2* in *Physcomitrella* by performing 5′ RACE-PCR. We found evidence for uORFs in these genes, suggesting that this regulatory mechanism was probably already present in the last common ancestor of extant land plants (fig. 1).

Remaining Mysteries: Novel Conserved Domain and Duplicate Loss in ARF5

The canonical structure of ARF proteins consists of a DNA-binding region (a B3 domain and an ARF domain), a variable middle region that confers activator or repressor transcriptional activity, and a dimerization region (domains III and IV) (fig. 2). Although originally considered as a nonfunctional region, the middle region includes numerous regulatory sites. Particularly, recent studies revealed the presence of target sites for *miR160* in *ARF10, ARF16, and ARF17* (Rhoades et al. 2002), for *miR167* in *ARF6 and ARF8* (Rhoades et al. 2002), and for ta-siRNAs in *ARF2*, *ARF3*, and *ARF4* (Williams et al. 2005). Here, we report the existence of a previously uncharacterized domain located in the middle region of ARF proteins, upstream of domain III. This latter domain includes a highly conserved signature motif LFG (L for leucine, F for phenylalanine, and G for glycine) and may form a β-sheet secondary structure, though it has not yet been possible to more precisely predict its secondary structure. Nevertheless, a similar LFG motif, which is present in the N-terminus of cyclin-dependent kinase inhibitory proteins, has been shown to mediate the binding of those kinases to cyclins (Zhu et al. 1995; Russo et al. 1996). Given that the ARF-mediated response to auxin interacts with the regulation of the cell cycle (see review by De Veylder et al. 2007), the putative binding of ARF proteins to cyclins represents an interesting hypothesis that might be worthy of testing.

The LFG motif is present in all ARF clades but is absent from genes of the ARF5/7 clade in certain species (supplementary table S2, Supplementary Material online). The most parsimonious scenario to explain this pattern of occurrence is to postulate three independent domain losses in angiosperms: one in the ARF5 clade after the initial radiation of the living angiosperms, one in the ARF7a clade, and one in the ARF7b clade, specifically in the lineage leading to the eudicots.

Interestingly, the ARF5 clade is also characterized by a repeated loss of duplicates after the successive rounds of genome duplication that occurred in flowering plants (Jiao et al. 2011). As a result, the ARF5 complement is often limited to a single copy in available sequenced genomes of flowering plants (supplementary fig. S2, Supplementary Material online), except in *Pop. trichocarpa* where two copies are present (Kalluri et al. 2007). This observation from the plant kingdom is reminiscent of the finding that genes expressed early in zebrafish development are less retained in duplicate after whole genome duplication, relative to lately expressed genes (Roux and Robinson-Rechavi 2008). *ARF5/MP* plays a role in the formation of vascular strands and in the initiation of the primary root in the early *A. thaliana* embryo (Hardtke and Berleth 1998). These latter anatomical structures are key innovations during the evolution of land plants. Roots are thought to have arisen independently in lycophytes and seed plants (Raven and Edwards 2001). Interestingly, this study shows that the clade ARF5/7 is restricted to seed plants. In this context, we venture the hypothesis that the clade ARF5/7 could have played a role in root evolution in seed plants but not in lycophytes. If corroborated by future functional studies, our findings suggest that the development of roots in lycophytes and seed plants rely on at least partly different molecular mechanisms. Similarly, members of the ARF5/7 clade could have been determinant in the generation of the vascular cambium present in seed plants.

Conclusion

Extensive gene duplication and domain rearrangement, especially through alternative splicing, have generally been viewed as opposite trends in gene family evolution (Kopelman et al. 2005; Su et al. 2006). We provide here a clear example of a gene family in which both these processes played a significant role in functional diversification. Particularly, multiple independent losses of domains III and IV suggest that changes in protein–protein interaction ability may have driven the evolution of the ARF protein interactome. As the full interactome of ARF and Aux/IAA proteins has recently been reconstructed in *Arabidopsis* (Vernoux et al. 2011), our analysis paves the way for future studies of the evolution of auxin response in terms of ARF-Aux/IAA network connectivity. More generally, we formulate in this study numerous evolutionary hypotheses based on the correlation between morphological innovations and molecular changes within the ARF family during the course of land plants evolution. These changes range from gene duplication to gain or loss of protein domains or regulatory mechanisms. Future functional studies carried out in emerging plant models, such as the moss *Physcomitrella*, will help to assess the full potential of these hypotheses. Nevertheless, these correlations emphasize the importance of careful molecular evolutionary analyses and multigenic family surveys as a source of hypotheses for functional studies.

Supplementary Material

Supplementary figures S1–S4 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Prof. Jody Banks for access to the *S. moellendorfii* genome before public release, Dr Frédéric Pautz for access to the living collection of Lyon Botanic Garden, Dr Bruno Fogliani for help with field collection of *A. trichopoda*, and Dr F. Nogué for providing *P. patens* material. C.F. received doctoral thesis funding from the French Ministry of National Education and Research. This work was supported by the RDP laboratory's parent organization: CNRS, INRA, ENS-Lyon, and UCBL.

References

Alonso CR, Wilkins AS. 2005. The molecular elements that underlie developmental evolution. *Nat Rev Genet.* 6:709–715.

Axtell MJ, Bartel DP. 2005. Antiquity of microRNAs and their targets in land plants. *Plant Cell* 17:1658–1673.

Banks JA, Nishiyama T, Hasebe M, et al. (103 co-authors). 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332:960–963.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25–36.

Chapman EJ, Estelle M. 2009. Mechanism of auxin-regulated gene expression in plants. *Annu Rev Genet.* 43:265–285.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16:735–743.

Cooke TJ, Poli D, Sztein AE, Cohen JD. 2002. Evolutionary patterns in auxin action. *Plant Mol Biol.* 49:319–338.

De Rybel B, Vassileva V, Parizot B, et al. (22 co-authors). 2010. A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. *Curr Biol.* 20:1697–1706.

De Smet I, Lau S, Voss U, et al. (21 co-authors). 2010. Bimodular auxin response controls organogenesis in Arabidopsis. *Proc Natl Acad Sci U S A.* 107:2705–2710.

De Smet I, Voss U, Lau S, et al. (13 co-authors). 2011. Unraveling the evolution of auxin signaling. *Plant Physiol.* 155:209–221.

De Veylder L, Beeckman T, Inze D. 2007. The ins and outs of the plant cell cycle. *Nat Rev Mol Cell Biol.* 8:655–665.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Finet C, Fourquin C, Vinauger M, Berne-Dedieu A, Chambrier P, Paindavoine S, Scutt CP. 2010. Parallel structural evolution of auxin response factors in the angiosperms. *Plant J.* 63:952–959.

Finet C, Jaillais Y. 2012. AUXOLOGY: when auxin meets plant evo-devo. *Dev Biol.* 369:19–31.

Fischer C, Neuhaus G. 1996. Influence of auxin on the establisment of bilateral symmetry in monocots. *Plant J.* 9:659–669.

Fourquin C, Vinauger-Douard M, Fogliani B, Dumas C, Scutt CP. 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 U S A.* 102:4649–4654.

Goetz M, Vivian-Smith A, Johnson SD, Koltunow AM. 2006. AUXIN RESPONSE FACTOR8 is a negative regulator of fruit initiation in *Arabidopsis*. *Plant Cell* 18:1873–1886.

Guermeur Y, Geourjon C, Gallinari P, Deleage G. 1999. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* 15:413–421.

Guilfoyle T, Hagen G. 2001. Auxin response factors. *J Plant Growth Regul.* 10:281–291.

Guilfoyle TJ, Hagen G. 2007. Auxin response factors. *Curr Opin Plant Biol.* 10:453–460.

Hagen G, Guilfoyle T. 2002. Auxin-responsive gene expression: genes, promoters, and regulatory factors. *Plant Mol Biol.* 49:373–385.

Hardtke CS, Berleth T. 1998. The *Arabidopsis* gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J.* 17:1405–1411.

Hardtke CS, Ckurshumova W, Vidaurre DP, Singh SA, Stamatiou G, Tiwari SB, Hagen G, Guilfoyle TJ, Berleth T. 2004. Overlapping and non-redundant functions of the *Arabidopsis* auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4. *Development* 131:1089–1100.

Jiao Y, Wickett NJ, Ayyampalayam S, et al. (17 co-authors). 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Kalluri UC, Difazio SP, Brunner AM, Tuskan GA. 2007. Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. *BMC Plant Biol.* 7:59.

Kieffer M, Neve J, Kepinski S. 2010. Defining auxin response contexts in plant development. *Curr Opin Plant Biol.* 13:12–20.

Kim J, Harter K, Theologis A. 1997. Protein-protein interactions among the Aux/IAA proteins. *Proc Natl Acad Sci U S A.* 94:11786–11791.

Kopelman NM, Lancet D, Yanai I. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat Genet.* 37:588–589.

Kozak M. 1992. Regulation of translation in eukaryotic systems. *Annu Rev Cell Biol.* 8:197–225.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.

Liu C, Xu Z, Chua NH. 1993. Auxin polar transport is essential for the establishment of bilateral symmetry during early plant embryogenesis. *Plant Cell* 5:621–630.

Muto H, Watahiki MK, Nakamoto D, Kinjo M, Yamamoto KT. 2007. Specificity and similarity of functions of the Aux/IAA genes in auxin signaling of *Arabidopsis* revealed by promoter-exchange experiments among MSG2/IAA19, AXR2/IAA7, and SLR/IAA14. *Plant Physiol.* 144:187–196.

Nagpal P, Ellis CM, Weber H, et al. (12 co-authors). 2005. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development* 132:4107–4118.

Nemhauser JL, Feldman LJ, Zambryski PC. 2000. Auxin and ETTIN in *Arabidopsis gynoecium* morphogenesis. *Development* 127: 3877–3888.

Nishimura T, Wada T, Okada K. 2004. 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, Yamamoto KT, Okada K. 2005. The *Arabidopsis* STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. *Plant Cell* 17:2940–2953.

Paponov IA, Teale W, Lang D, Paponov M, Reski R, Rensing SA, Palme K. 2009. The evolution of nuclear auxin signalling. *BMC Evol Biol.* 9:126.

Pekker I, Alvarez JP, Eshed Y. 2005. Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *Plant Cell* 17:2899–2910.

Raven JA, Edwards D. 2001. Roots: evolutionary origins and biogeochemical significance. *J Exp Bot.* 52:381–401.

Remington DL, Vision TJ, Guilfoyle TJ, Reed JW. 2004. Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiol.* 135:1738–1752.

Rensing SA, Lang D, Zimmer AD, et al. (70 co-authors). 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319:64–69.

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP. 2002. Prediction of plant microRNA targets. *Cell* 110:513–520.

Roux J, Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* 4:e1000311.

Russo AA, Jeffrey PD, Patten AK, Massague J, Pavletich NP. 1996. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* 382:325–331.

Sessions RA, Zambryski PC. 1995. *Arabidopsis gynoecium* structure in the wild and in ettin mutants. *Development* 121:1519–1532.

Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman DP. 2007. The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. *Plant Cell* 19: 2440–2453.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Su Z, Wang J, Yu J, Huang X, Gu X. 2006. Evolution of alternative splicing after gene duplication. *Genome Res.* 16:182–189.

Szemenyei H, Hannon M, Long JA. 2008. TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science* 319:1384–1386.

Tiwari SB, Hagen G, Guilfoyle T. 2003. The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell* 15: 533–543.

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

Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ. 1997. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* 9: 1963–1971.

Varaud E, Brioudes F, Szecsi J, Leroux J, Brown S, Perrot-Rechenmann C, Bendahmane M. 2011. AUXIN RESPONSE FACTOR8 regulates *Arabidopsis* petal growth by interacting with the bHLH transcription factor BIGPETALp. *Plant Cell* 23:973–983.

Vernoux T, Brunoud G, Farcot E, et al. (24 co-authors). 2011. The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol Syst Biol.* 7:508.

Wang D, Pei K, Fu Y, Sun Z, Li S, Liu H, Tang K, Han B, Tao Y. 2007. Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* 394:13–24.

Wang Y, Deng D, Shi Y, Miao N, Bian Y, Yin Z. 2012. Diversification, phylogeny and evolution of auxin response factor (ARF) family: insights gained from analyzing maize ARF genes. *Mol Biol Rep.* 39: 2401–2415.

Weijers D, Benkova E, Jager KE, Schlereth A, Hamann T, Kientz M, Wilmoth JC, Reed JW, Jurgens G. 2005. Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. *EMBO J.* 24:1874–1885.

Williams L, Carles CC, Osmont KS, Fletcher JC. 2005. A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis* ARF2, ARF3, and ARF4 genes. *Proc Natl Acad Sci U S A.* 102:9703–9708.

Wu J, Wang F, Cheng L, Kong F, Peng Z, Liu S, Yu X, Lu G. 2011. Identification, isolation and expression analysis of auxin response factor (ARF) genes in *Solanum lycopersicum*. *Plant Cell Rep.* 30: 2059–2073.

Wu MF, Tian Q, Reed JW. 2006. *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* 133:4211–4218.

Zhu L, Harlow E, Dynlacht BD. 1995. p107 uses a p21CIP1-related domain to bind cyclin/cdk2 and regulate interactions with E2F. *Genes Dev.* 9:1740–1752.