Successive Domain Rearrangements Underlie the Evolution of a Regulatory Module Controlled by a Small Interfering Peptide

João Raimundo,[†,‡,1] Rómulo Sobral,[†,1] Sara Laranjeira,[1] and Maria Manuela R. Costa*,[1]

[1]Biosystems and Integrative Sciences Institute (BioISI), Plant Functional Biology Center, University of Minho, Braga, Portugal

[†]These authors contributed equally to this work.

[‡]Present address: Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ

*Corresponding author: E-mail: manuela.costa@bio.uminho.pt.

Abstract

The establishment of new interactions between transcriptional regulators increases the regulatory diversity that drives phenotypic novelty. To understand how such interactions evolve, we have studied a regulatory module (DDR) composed by three MYB-like proteins: DIVARICATA (DIV), RADIALIS (RAD), and DIV-and-RAD-Interacting Factor (DRIF). The DIV and DRIF proteins form a transcriptional complex that is disrupted in the presence of RAD, a small interfering peptide, due to the formation of RAD–DRIF dimers. This dynamic interaction result in a molecular switch mechanism responsible for the control of distinct developmental processes in plants. Here, we have determined how the DDR regulatory module was established by analyzing the origin and evolution of the DIV, DRIF, and RAD protein families and the evolutionary history of their interactions. We show that duplications of a pre-existing MYB domain originated the DIV and DRIF protein families in the ancestral lineage of green algae, and, later, the RAD family in seed plants. Intraspecies interactions between the MYB domains of DIV and DRIF proteins are detected in green algae, whereas the earliest evidence of an interaction between DRIF and RAD proteins occurs in the gymnosperms, coincident with the establishment of the RAD family. Therefore, the DDR module evolved in a stepwise progression with the DIV–DRIF transcription complex evolving prior to the antagonistic RAD–DRIF interaction that established the molecular switch mechanism. Our results suggest that the successive rearrangement and divergence of a single protein domain can be an effective evolutionary mechanism driving new protein interactions and the establishment of novel regulatory modules.

Key words: MYB, RADIALIS, DIVARICATA, DRIF, protein evolution, protein–protein interaction, domain rearrangement, molecular antagonism, small interfering peptide, flower asymmetry, *Antirrhinum majus*.

Introduction

The developmental intricacy of multicellular organisms, and particularly the adaptive flexibility that plants exhibit, is often associated with a cumulative complexity in gene regulatory networks (Bartlett and Whipple 2013; Pires et al. 2013; Breuninger et al. 2016; Cho 2017; Serrano-Bueno et al. 2017).

The emergence of new biological functions and morphologies is coupled with the evolution of proteomes through duplication and recombination of a limited set of protein domains, which are independent folding units with particular subfunctions that have been proposed to represent the unit of modular evolution (Pawson 1995; Vogel et al. 2004; Jin et al. 2009). Proteins can evolve as a consequence of duplication and divergence of a domain or by rearranging pre-existing domains using various mechanisms of genetic recombination (Pasek et al. 2006; Schmidt and Davies 2007). Modular domain rearrangements were, in fact, the main mechanism behind the evolution of the bHLH family, one of the largest and most diverse transcription factor families in plants (Morgenstern and Atchley 1999). The association of highly conserved bHLH domains with other distinct functional domains strongly suggests that modular evolution must have had an important role in the emergence of transcription factor families in plants (Morgenstern and Atchley 1999; Brkljacic and Grotewold 2017).

Protein domains are often involved in interactions with proteins or other ligands such as DNA or RNA, thus the acquisition of a domain or the functional divergence of an already existing one can drive the new protein to establish new molecular connections within the cell. Most transcription factors act as homo or heterodimers to increase DNA-binding specificity or transcriptional activation specificities (Kosugi and Ohashi 2002; Amoutzias et al. 2008; van der Graaff et al. 2009). Therefore, the combinatorial assortment driven by the dynamic formation of homo and heterodimers constitutes a molecular mechanism that diversifies DNA-binding specificities and increases regulatory complexity (Amoutzias et al. 2008).

Structural analysis of molecular networks has been greatly advanced by the availability of large-scale protein–protein interaction studies that allows the identification of modular network structures in several organisms (Uetz et al. 2000;

Open Access

Article

Han et al. 2004; Proost and Mutwil 2016). There are, however, very few empirical studies demonstrating molecular evolutionary events that have functionally contributed to the formation of new regulatory modules. With the increased availability of sequenced transcriptomes, it is now possible to trace the origin of genetic regulators that control key aspects of plant development and to understand what molecular events have led to the establishment of regulatory modules. To better understand this question, we have investigated the molecular origin and evolution of the DIVARICATA (DIV), RADIALIS (RAD), and DIV-and-RAD-Interacting Factors (DRIF) protein families and studied the evolutionary molecular mechanisms driving the establishment of their interactions.

DIV, RAD, and DRIF proteins are part of a regulatory module (DDR) that controls diverse key aspects of plant development (Machemer et al. 2011; Raimundo et al. 2013). In *Antirrhinum majus*, the dorsoventral asymmetry of the flower depends on the activity of CYCLOIDEA (CYC) (Luo et al. 1996, 1999; Cubas et al. 1999) that controls the expression of *RAD*, a gene encoding a MYB-like protein, in the dorsal domain of the flower meristem (Corley et al. 2005; Costa et al. 2005). RAD antagonizes the function of DIV, another MYB-like protein required for specifying petal ventral identity (Almeida et al. 1997; Galego and Almeida 2002; Raimundo et al. 2013). In the ventral region of the meristem, DIV interacts with a DRIF protein, forming a complex that most likely controls the regulatory network that specifies ventral identity. In the dorsal region of the floral meristem, where *RAD* is expressed, RAD sequesters the DRIF proteins in the cytoplasm, preventing their shuttling to the nucleus and the formation of the DRIF–DIV complex, which results in the establishment of the dorsal petal identity of the *Antirrhinum* flower (Raimundo et al. 2013). The small size of the RAD proteins (∼99 amino acids [a.a.]) and their particular mode of action, as antagonistic agents in the establishment of the DIV–DRIF complex, classifies them as microproteins or small interfering peptide (siPEP) (Seo et al. 2011; Staudt and Wenkel 2011; Eguen et al. 2015). A similar subcellular competition involving a homologous DDR regulatory module seems to take place in *Solanum lycopersicum* (tomato), where a RAD homolog (FSM1) negatively regulates cell expansion of the tomato fruit pericarp by competing with a DIV homolog (MYBI) for the interaction with FSB1, a DRIF homolog. In the absence of FSM1, FSB1 and MYBI establish a regulatory complex that enhances cell expansion of the pericarp (Rose et al. 1999; Machemer et al. 2011).

Homologous genes to the ones of the DDR module have been also implicated in distinct developmental functions in several different species. These include the regulation of early photomorphogenesis and floral transition in *Arabidopsis thaliana* (Hamaguchi et al. 2008; Li et al. 2015), the development of petal shape and color in *Torenia fournieri* flowers (Shikata et al. 2011; Su et al. 2017) or the sugar and hormonal regulation of an α-amylase gene in *Oryza sativa* (Lu et al. 2002). It is still unknown whether these proteins participate in a similar subcellular molecular antagonism to the one observed in *Antirrhinum* and tomato. However, it has been shown that *Arabidopsis* DRIF proteins can interact with

DIV and RAD proteins, as well as with their respective homologs in tomato (Machemer et al. 2011), showing that the interactions between the proteins that constitute the DDR regulatory module are conserved among different species, and that this module has most likely been recruited during evolution to perform different roles across the angiosperm lineage.

The interactions between the DIV, DRIF, and RAD proteins are essential for the functioning of the DDR regulatory module and are established through the MYB domain of each of the three protein families (Machemer et al. 2011; Raimundo et al. 2013). The DIV proteins contain two MYB domains, the C-terminal domain is a SHAQKYF-type MYB (MYBII) (Wang 1997; Lu et al. 2002, 2009) that is known to bind DNA (Rose et al. 1999; Raimundo et al. 2013). The N-terminal MYB domain (MYBI) is responsible for the protein interactions that DIV proteins establish with DRIF proteins and is very similar to the RAD MYB domain (Almeida et al. 1997; Rose et al. 1999; Galego and Almeida 2002; Machemer et al. 2011). Members of the DRIF protein family have two conserved domains: a C-terminal domain of unknown function (DUF3755) and an N-terminal MYB domain that interacts with the MYBI of DIV proteins and with the single-MYB domain of RAD proteins. To understand the evolutionary history of the DDR regulatory module is, therefore, essential to know how each of the three MYB protein families has evolved and when the combinatorial interactions were first established.

In the present study, we demonstrate that the DIV and DRIF protein families have emerged in the lineage that originated the green algae, while the first RAD family members were identified only in gymnosperms. We also show that the MYB domain responsible for the combinatorial interaction between the three proteins has a common origin and has most probably evolved by successive domain rearrangements. The interaction between DRIF and DIV proteins is first detected in the green algae, and the antagonistic RAD-DRIF interaction is detected in the gymnosperms associated with the origin of the RAD family. Therefore, the DIV–DRIF interaction was established much earlier than the antagonistic interaction between RAD and DRIF proteins. We, therefore, propose that the successive rearrangement and divergence of a single protein domain can be an effective evolutionary mechanism driving the establishment of new protein interactions and regulatory diversity.

Results

DIV, DRIF, and RAD Contain Distinctive MYB Domains

The DIV, DRIF, and RAD proteins establish a regulatory module that controls different developmental processes across the angiosperms. The molecular dynamics that govern the mode of action of the DDR regulatory module are conserved in *Antirrhinum* and tomato, where it regulates very distinct developmental traits. In order to understand how this regulatory module has been established, it is, therefore, essential to analyze the domain topology of the protein families that
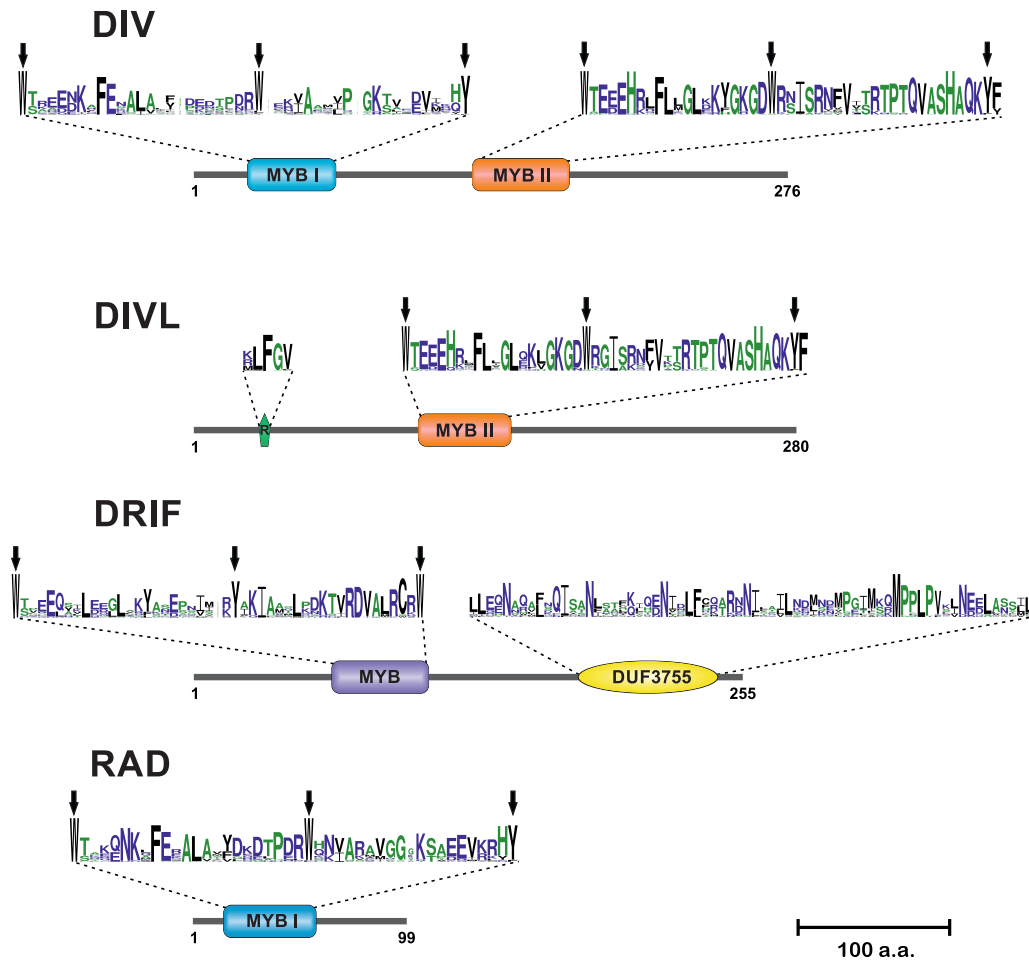
Fig. 1. Structure of the DIV, DIVL, DRIF, and RAD protein families. Schematic representation of the general structure of the DIV, DIVL, DRIF, and RAD proteins from five angiosperm species: *Amborella trichopoda*, *Oryza sativa*, *Solanum lycopersicum*, *Antirrhinum majus*, and *Arabidopsis thaliana*. The schematics are drawn to scale based on the average values of protein length, domain position, and domain length for all of the homologs found on the chosen angiosperm species. Domains with the same color share the same topology that is represented on the sequence logos that were generated based on the alignment depicted in supplementary figures 1–4, Supplementary Material online. The degree of certainty of each amino acid position is indicated by the height of the respective symbol. The conserved aromatic residues typical of the MYB domain topology are signaled with black arrows. The scale bar represents 100 a.a.

constitute the DDR module. To investigate the diversity of the DIV, DRIF, and RAD proteins, extensive BLAST searches were performed on several transcript and genome databases in order to identify all the homologs of the three protein families in an early diverging angiosperm, *Amborella trichopoda*, a monocot, *O. sativa*, and three dicots, *S. lycopersicum*, *A. majus*, and *Ar. thaliana*. The DIV, DRIF, and RAD homologous proteins were retrieved in order to generate the alignments that were used to calculate the sequence logos for all the protein domains (fig. 1 and supplementary figs. 1–4, Supplementary Material online).

DIV proteins have on average 276 a.a. and contain two MYB domains; the first (MYBI) is composed of approximately 44 a.a. and the second (MYBII) of 51 a.a. MYBI is an atypical MYB domain with the last of the three regularly spaced tryptophan residues, which characterize a MYB domain (Wang 1997), being replaced by a tyrosine ($-W-X_{23}-W-X_{20}-Y-$). The second conserved domain (MYBII) is commonly denominated as a SHAQKYF-type MYB due to the presence of the

characteristic SHAQKYF amino acid sequence that integrates the last tyrosine of the MYB motif ($-W-X_{19}-W-X_{22}-Y-$) (Rose et al. 1999; Lu et al. 2002, 2009). The alignment of the DIV protein family revealed that some of the retrieved homologs contained the C-terminal MYB domain (MYBII) but completely lack the N-terminal domain (MYBI) typical of the DIV proteins (compare supplementary fig. 1 with supplementary fig. 4, Supplementary Material online). These sequences were therefore denominated by DIV-Like (DIVL) proteins. The proteins of the DIVL family have on average 280 a.a. and contain a MYBII SHAQKYF-type domain almost identical to the one present in the DIV family. However, instead of the MYBI domain, DIVL proteins contain a very small motif (R/KLFGV), denominated by R motif, identified as a plantspecific active repression domain that occurs in at least 29 *Arabidopsis* transcription factors, including members of the ABI3/VP1, ARF, HSF, and RAV families (Ikeda and Ohme-Takagi 2009). DRIF proteins have approximately 255 a.a. and contain two conserved domains. The first is a MYB

domain with about 46 a.a. and characterized by an atypical tyrosine replacing the central characteristic tryptophan of the canonical MYB motif (-W-X$_{23}$-Y-X$_{20}$-W-). The second DRIF conserved domain covers ~66 a.a., has an unknown function, and has been annotated as DUF3755. The RAD family is constituted by very small proteins with ~99 a.a. in length. RAD proteins contain only one conserved domain composed by 44 a.a., almost identical to DIV MYBI domain, sharing a similar MYB topology (-W-X$_{23}$-W-X$_{20}$-Y-) (fig. 1).

To determine whether the conserved domains of the DDR proteins are present in other protein families, the isolated MYBI, MYBII, and DUF3755 conserved domains were used to perform BLAST searches on angiosperm protein databases. The BLAST searches for DIV MYBI have shown that this domain has indeed a topology unique to DIV and RAD proteins, thus lacking significant homology to any other MYB protein. The MYBII SHAQKYF domain is present not only in DIV and DIVL proteins but also in proteins involved in circadian clock control such as LATE ELONGATED HYPOCOTYL (LHY) and CIRCADIAN CLOCK ASSOCIATED 1 (CCA1). To distinguish the DIV and CCA1/LHY families, a phylogenetic tree was produced using the MYBII SHAQKYF of DIV proteins and single MYB domain of CCA1/LHY homologs in different species across the plant lineage. The phylogenetic tree suggests that the protein families are clearly differentiated as the SHAQKYF MYBs of DIV and CCA1/LHY homolog proteins group into two different monophyletic clades (supplementary fig. 5, Supplementary Material online). The DRIF MYB and DUF3755 are unique to the DRIF protein family as no other proteins were found that contain either of these domains.

These results show that the conserved protein domains of the DDR regulatory module are unique to each protein family and, therefore, can be traced throughout plant evolution in order to understand their molecular origin and evolution.

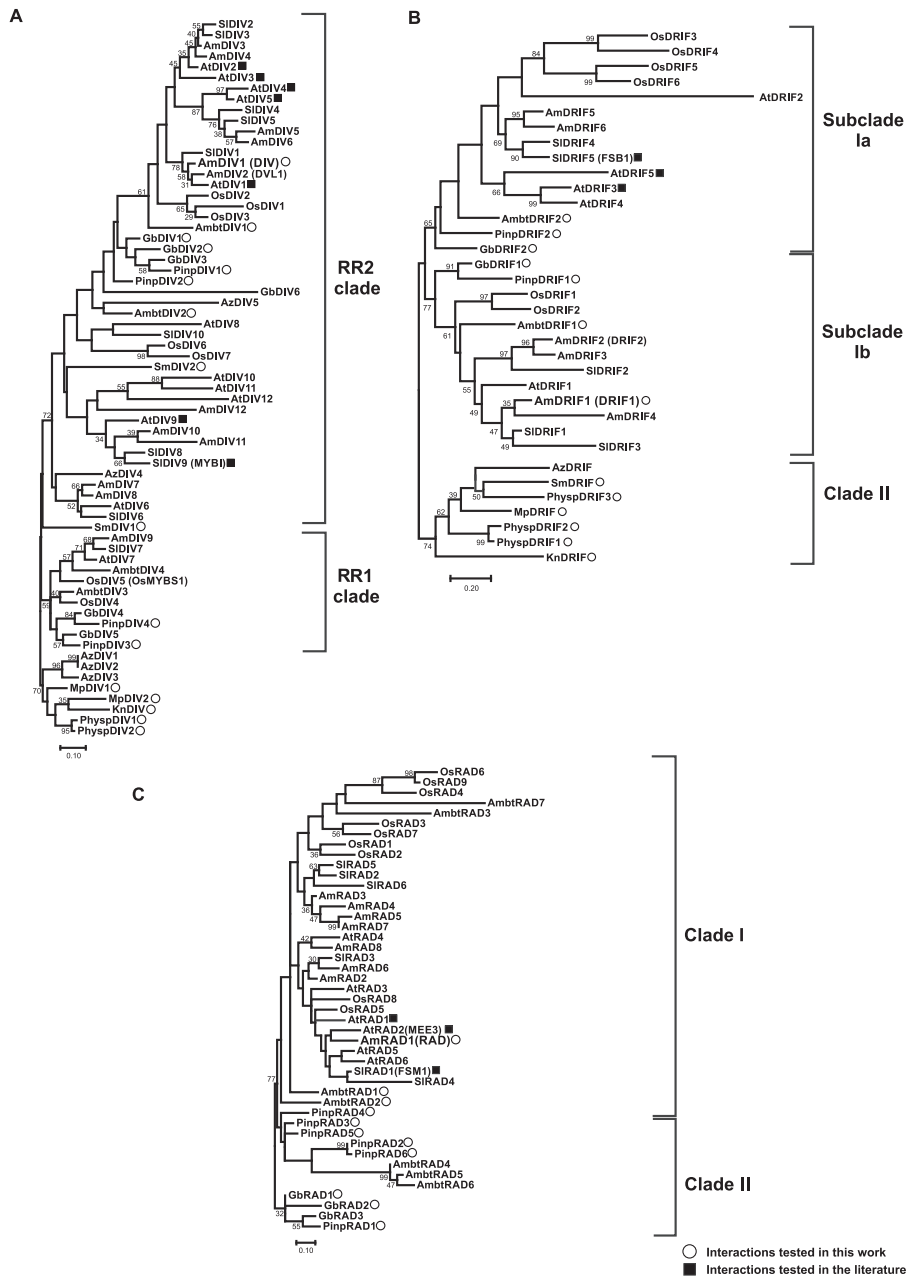## The MYB Domains from DIV, DRIF, and RAD Proteins Have a Common Evolutionary Origin

To understand the origin of the protein families that comprise the DDR regulatory module, a thorough analysis of the phylogeny of the RAD, DIV, and DRIF families was performed. Several authors have studied RAD, DIV, and DRIF phylogenies using diverse angiosperms taxa (Howarth and Donoghue 2009; Boyden et al. 2012; Raimundo et al. 2013; Gao et al. 2017). However, in order to determine the evolutionary origin of these protein families, a phylogenetic analysis that includes homologs from distinct species outside the flowering plants was required. Accordingly, homologous sequences to the DIV, DRIF, and RAD proteins were retrieved from different species on the algae/plant lineage (green lineage). The selected species used to retrieve the DDR homologs represent major groups that characterize the green lineage: *Pinus pinaster* (gymnosperm), *Azolla filiculoides* (fern), *Selaginella moellendorffii* (lycophyte), *Physcomitrella patens* (moss), *Marchantia polymorpha* (liverwort), and *Klebsormidium nitens* (multicellular green algae). The typical MYB domain architectures of the DIV, DIVL, and DRIF proteins remained unaltered

throughout the evolution of green algae, liverworts, mosses, lycophytes, ferns, gymnosperms, and angiosperms. The genes encoding DDR proteins have duplicated at a low rate up until the emergence of the gymnosperms that contain four *DIV*, four *DIVL*, and two *DRIF* genes. These gene families then experienced extensive duplication during the evolution of angiosperms (figs. 2 and 3).

Figure 2 shows the phylogenetic analysis performed on the three DDR protein families using homologs of *P. pinaster*, *Ginkgo biloba*, *Az. filiculoides*, *Se. moellendorffii*, *Ph. patens*, *M. polymorpha*, and *K. nitens* and the angiosperm species homologs. The alignments used to generate the phylogenetic trees were produced using the amino acid sequences of the MYB domains for each family (supplementary figs. 1–4, Supplementary Material online). Analysis of the DIV phylogeny revealed that this family could be subdivided into two clades, both with proteins from ferns, gymnosperms, and angiosperms (fig. 2a). This subdivision of the DIV tree is coincident to the one proposed by Howarth and Donoghue (2009) and Gao et al. (2017) that divided the phylogeny of the DIV family into two monophyletic clades, RR1 and RR2. Analysis of the DIV phylogeny also suggests that separation of the RR1 and RR2 clades might have occurred after the divergence between the mosses and the lycophytes as DIV proteins belonging to *Ph. patens*, *M. polymorpha*, and *K. nitens* lay outside these clades.

The DRIF phylogenetic tree is also subdivided into two main clades. One clade contains ancient plant DRIFs (clade II) while the other, which is subdivided into two subclades (subclades Ia and Ib), contains the seed plant DRIF homologs (fig. 2b). Subclades Ia and Ib show independent duplications that have occurred during the evolution of the angiosperms (fig. 2b).

The evolution of the RAD family is characterized by extensive duplication in both gymnosperms and angiosperms. *Pinus pinaster* homologs are clustered together in different clades indicating parallel duplications that occurred after the divergence between gymnosperms and angiosperms (fig. 2c). The phylogenetic profiling of the angiosperms RAD family is also coincident with published data that suggested the presence of three RAD paralog lineages that were originated in a common ancestor of the core eudicots (Boyden et al. 2012; Gao et al. 2017). Additionally, just five of the seven *Am. trichopoda* RAD proteins are clustered together, showing that in angiosperms, the duplication of the RAD genes has likely occurred after the split between basal angiosperms and the higher angiosperms (fig. 2c). No homologs were found outside the seed plants group suggesting that the RAD family likely arose during the lineage that gave rise to the seed plants (figs. 2c and 3). To assess the origin of the RAD proteins, all MYB domains of the MYB proteins from four gymnosperm species, *P. pinaster*, *Gnetum montanum*, *G. biloba*, and *Picea sitchensis*, were retrieved and aligned to produce a phylogenetic tree (supplementary fig. 6, Supplementary Material online). The phylogenetic analysis shows that the RAD family clusters very closely to the DIV family clade suggesting that the RAD family might have had its origins on a duplication of the MYBI domain from a *DIV* gene or, alternatively, on an

FIG. 2. Phylogenetic trees representing the evolutionary history of the DIV, DRIF, and RAD protein families. (*a*) Phylogenetic history of the DIV protein family constructed using an alignment of the MYBI and MYBII conserved domains of DIV homolog sequences from *Klebsormidium nitens* (Kn), *Marchantia polymorpha* (Mp), *Physcomitrella patens* (Physp), *Selaginella moellendorffii* (Sm), *Azolla filiculoides* (Az), *Ginkgo biloba* (Gb), *Pinus pinaster* (Pinp), *Amborella trichopoda* (Ambt), *Oryza sativa* (Os), *Solanum lycopersicum* (Sl), *Antirrhinum majus* (Am), and *Arabidopsis thaliana* (At). The root of the tree was placed on the ancestral node harboring the *K. nitens*, *Ph. patens*, and *M. polymorpha* clade. The tree is divided into two clades (RR1 and RR2) that represent the evolutionary history of the DIV family. (*b*) Phylogenetic history of the DRIF protein family constructed using the alignment of the MYBI and DUF3755 conserved domains of DRIF homolog sequences from the same species used to build the tree in (*a*). The root of the tree was placed on the ancestral node harboring the *K. nitens*, *Ph. patens*, *M. polymorpha*, *Az. filiculoides*, and *Se. moellendorffii* clade. The tree is divided into different clades (clade I and II). Clade I is subdivided into clades Ia and Ib. (*c*) Phylogenetic history of the RAD protein family constructed using the aligned sequences from *P. pinaster* (Pinp), *G. biloba* (Gb), *Am. trichopoda* (Ambt), *O. sativa* (Os), *S. lycopersicum* (Sl), *A. majus* (Am), and *Ar. thaliana* (At). The root of the tree was placed on the node harboring the *P. pinaster/G. biloba* clade. The trees were generated using the maximum likelihood method and are supported by bootstrap values generated from 1,000 replicates indicated next to each node for values >30%. The trees represent the most likely trees generated by the algorithm. The scale bar indicates the evolutionary distance between the groups.

entirely duplicated *DIV* gene that progressively lost the MYBII domain.

To determine the evolutionary origin of the DRIF and DIV protein families, a BLAST search was conducted in species

established before the emergence of the green lineage. The species chosen were the red algae *Galdieria sulphuraria*, *Chondrus crispus*, and *Porphyra umbilicalis*. The search for DDR homologs in the red algae revealed that no DRIFs
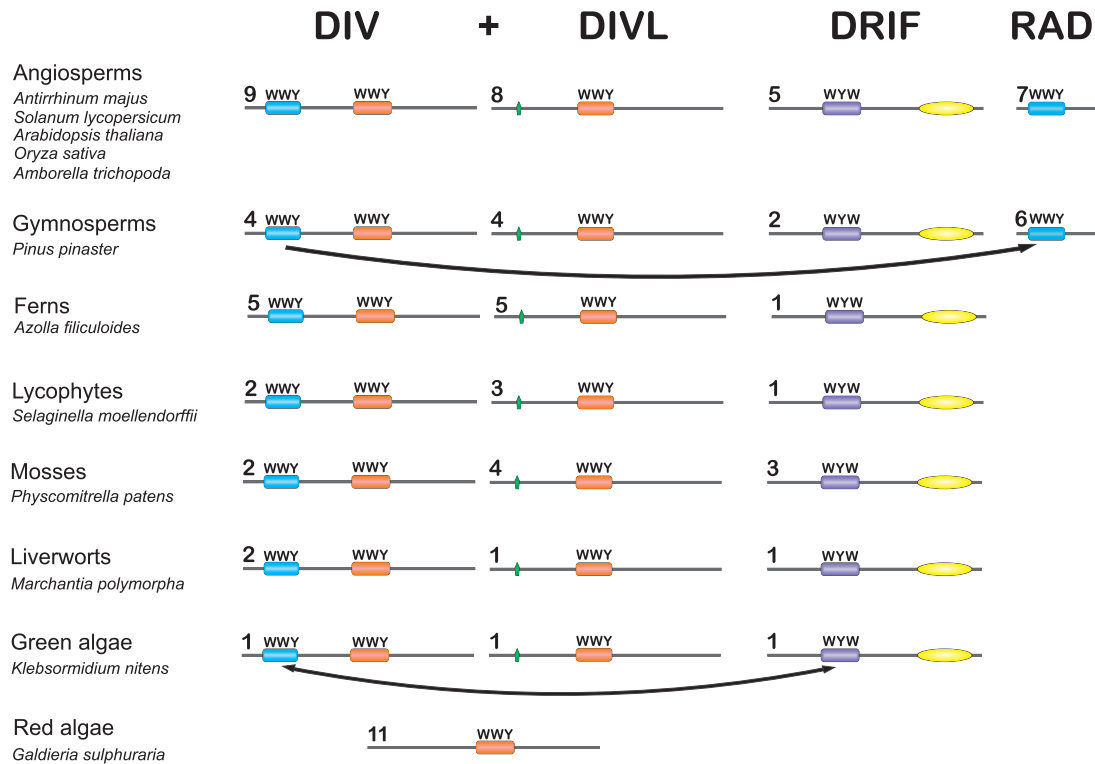
Fig. 3. Evolutionary history of the DRR protein conserved domains. Representation of the protein domain structure of the DIV, DIVL, DRIF, and RAD protein families at several key evolutionary points, from red algae to angiosperms (*Galdieria sulphuraria*, *Klebsormidium nitens*, *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Azolla filiculoides*, *Pinus pinaster*, *Amborella trichopoda*, *Oryza sativa*, *Solanum lycopersicum*, *Antirrhinum majus*, and *Arabidopsis thaliana*). The number of homologs from each protein family found on each species is shown next to each scheme (the angiosperm homolog number corresponds to an average between *Am. trichopoda*, *O. sativa*, *S. lycopersicum*, *A. majus*, and *Ar. thaliana*). Conserved protein domains are represented using the same color code as in figure 1, and their general MYB topology denoted by the three typical aromatic residues (-W-X-Y-X-W- or -W-X-W-X-Y-) is represented above the respective domains. The arrows point to domain duplication events.

(or proteins containing any of its conserved domains) are present in these organisms. This suggests that the DRIF family must have evolved after the divergence between the red algae and the green algae lineages. In all the three red algae species analyzed, several genes were found that encode for proteins containing SHAQKYF MYB domains. Although these SHAQKYF-type MYB genes contain a domain similar to the MYBII of DIV and DIVL proteins, none of them has a DIV MYBI domain or a DIVL R motif (supplementary fig. 7, Supplementary Material online, and fig. 3). These results suggest that the MYBI and MYB domains, respectively of DIV and DRIF, had their origin specifically in the green algae lineage.

In several green algae species belonging to the chlorophytes (*Chlamydomonas* sp, *Volvox carteri*, *Dunaliella salina*, Micromonas sp, *Ostreococcus tauri*, *Chlorella vulgaris*, *Gonium pectorale*, and *Bathycoccus prasinos*) and the charophytes (*K. nitens*), it was possible to identify at least one DIV homolog containing the two characteristic MYB domains, thus indicating that the N-terminal MYBI domain has evolved specifically in an ancestral of the green algae lineage. All the green algae SHAQKYF MYB domains show the typical angiosperm MYBII topology (-W-X-W-X-Y-) (supplementary fig. 8, Supplementary Material online).

However, only *K. nitens*, *B. prasinos*, and *O. tauri* DIV proteins have the canonical MYBI domain topology with the canonical aromatic residues (-W-X-W-X-Y-) (supplementary fig. 8, Supplementary Material online).

Homologs of DRIF proteins containing the two domains (MYB and DUF3755) were also identified in the transcriptomes of some green algae taxa (*C. vulgaris*, *Auxenochlorella protothecoides*, *Go. pectorale*, *Coccomyxa subellipsoidea*, *D. salina*, *V. carteri*, *Chlamydomonas* sp., *Ostreococcus* sp., and *K. nitens*) (supplementary fig. 9, Supplementary Material online). The conserved MYB topology (-W-X-Y-X-W-) was identified in *C. vulgaris*, *Au. protothecoides*, *Co. subellipsoidea*, and *K. nitens* (supplementary fig. 9, Supplementary Material online). The presence of both DRIF domains in several green algae species from both chlorophytes and charophytes and their absence in distinct red algae species suggest that, similarly to DIV proteins, both of the DRIF domains must have evolved in the green algae ancestral lineage (fig. 3).

To assess the origin of the MYB domain of DRIF and the MYBI of DIV, all the MYB domains of MYB proteins from *K. nitens*, *Chlamydomonas reinhardtii*, and *V. carteri* (green algae species with DDR homologs having canonical and atypical MYB domains) were retrieved and aligned in order to obtain the phylogenetic tree shown in figure 4. The phylogenetic tree
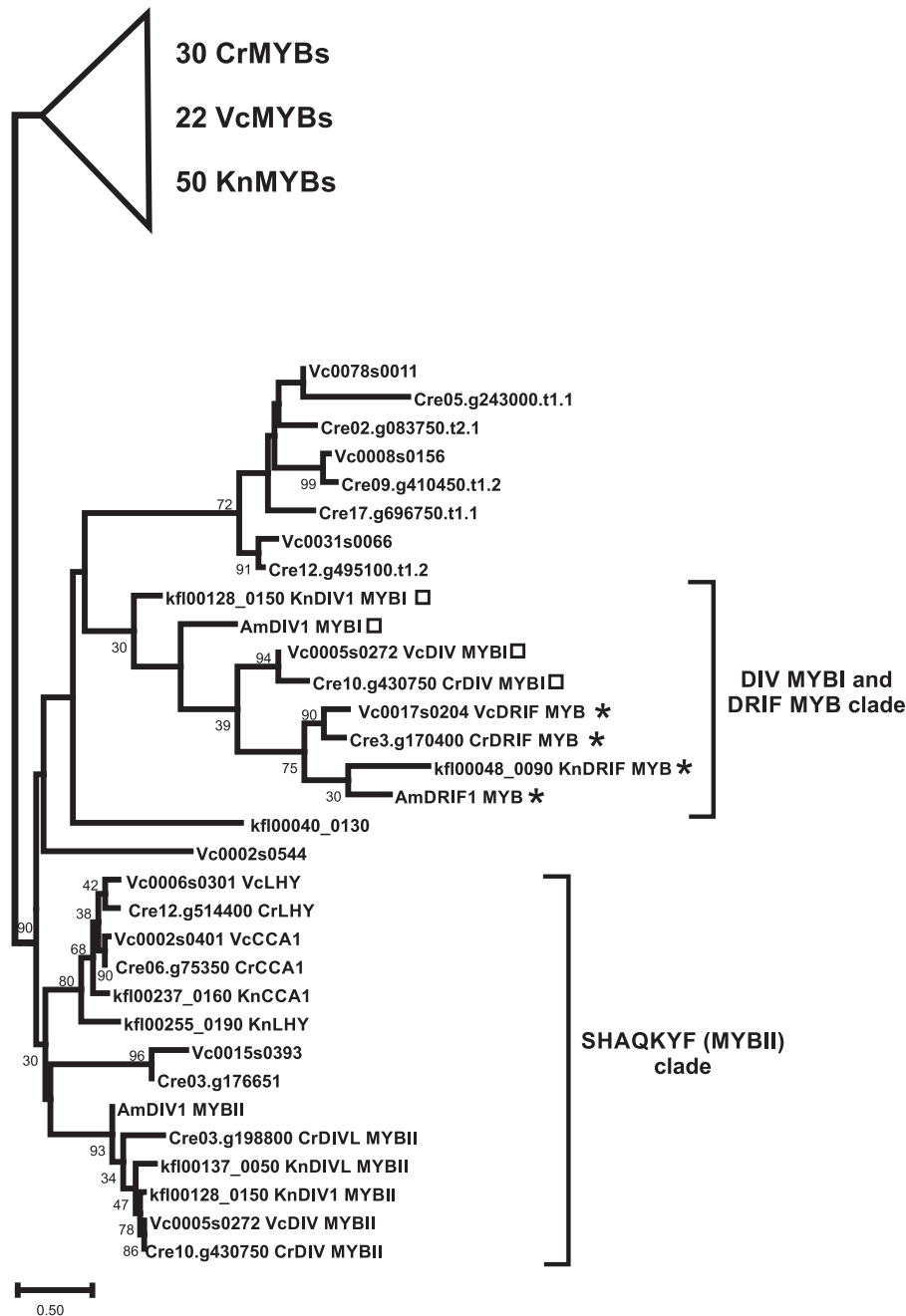
FIG. 4. Phylogenetic trees representing the evolutionary history of the MYB family in the green algae. Phylogenetic trees constructed using the maximum likelihood method on an alignment generated with the amino acidic sequences that constitute the MYB domains of all the MYB proteins present in *Chlamydomonas reinhardtii*, *Volvox carteri*, and *Klebsormidium nitens*. The MYB domain of DIV, DIVL, DRIF, CCA1, and LHY homolog proteins is signaled after the respective gene references. The MYBI domain of the DIV homologs (squares) and the MYB domain of the DRIF homologs (asterisks) cluster together suggesting a common evolutionary origin and a closer phylogenetic relationship. The MYBII domains from the DIV homologs are included in the clade of the SHAQKYF MYBs. The tree is supported by bootstrap values generated from 1,000 replicates indicated next to each node for values >30%, and the root was placed on the SHAQKYF clade. The tree represents the most likely tree generated by the algorithm. The scale bar indicates the evolutionary distance between the groups.

shows that the MYBI domain of DIV and the MYB domain of DRIF proteins cluster closely to the SHAQKYF MYB clade (fig. 4) suggesting that they might have a common origin through a duplication from an ancestral SHAQKYF MYB domain.

The analysis of the origin and evolution of the DDR protein families suggested that DIV and DIVL proteins had their origin on an ancient SHAQKYF MYB protein also present in the red algae, and that the MYBI domain and R motif from the DIV and DIVL proteins, respectively, were acquired after the divergence between the red algae and the green algae lineages. Most interestingly, these results have shown that the MYBI domain of DIV and the MYB domains of DRIF and RAD have a shared

evolutionary origin and might have been established through successive domain rearrangements.

DIV–DRIF Interaction Has Evolved Prior to the Antagonistic RAD–DRIF–DIV Interaction

The interaction between some members of the DRIF family with DIV and RAD proteins has been shown to be key to the establishment of distinct developmental programs in *Antirrhinum* and tomato (Machemer et al. 2011; Raimundo et al. 2013), where RAD behaves like an siPEP and competes with DIV for the interaction of a DRIF protein in an antagonistic subcellular mechanism. Moreover, Machemer et al. (2011) have also shown that FSB1, a DRIF protein from tomato, is also able to interact with five DIV and two RAD proteins from *Arabidopsis* and that FSM1, a RAD homolog, interacts with other two *Arabidopsis* DRIF proteins. These results suggest that the protein interactions, characteristic of the DDR module, are conserved, at least between homologs of angiosperm species. In order to determine the evolutionary key point that led to the establishment of the interactions between DRIF and DIV or RAD proteins, the open reading frames of all the homologs from *K. nitens*, *M. polymorpha*, *Ph. patens*, *Se. moellendorffii*, *P. pinaster*, and *Am. trichopoda* (fig. 2, circles) were cloned, and the interactions between these proteins were tested using a yeast two-hybrid (Y2H) assay (fig. 5). All proteins were cloned in fusion with the GAL4 activation or DNA-binding domain and interactions were assayed with proteins in both fusion forms, unless the proteins were able to promote transcription of reporter genes. In these cases, only the fusion protein to the GAL4 activation domain was assayed (supplementary fig. 11, Supplementary Material online).

According to the Y2H assay, the DIV and DRIF homologs identified in *K. nitens* were able to interact suggesting that this is an ancient interaction already present in the green algae. Both *M. polymorpha* DIVs were able to interact with the single DRIF homolog, showing that the interaction between DIV and DRIF proteins may also be conserved in early land plants. In *Ph. patens* there are two DIVs and three DRIFs, but the interaction was only detected with the PpDIV1/PpDRIF3 and PpDIV2/PpdDRIF3 protein combinations (fig. 5). In *Se. moellendorffii*, no interaction was detected between the two DIV homologs and the DRIF homolog. To understand which of the *Se. moellendorffii* proteins may have lost the ability to interact, the binding of SmDIV and SmDRIF proteins to DRIF and DIV homologs from *Antirrhinum* was tested. *Antirrhinum* DRIF1 protein is able to interact with both DIV proteins from *Selaginella* (supplementary fig. 10, Supplementary Material online). On the other hand, SmDRIF is unable to interact with AmDIV1. This shows that during the evolution of *Selaginella*, some changes in the DRIF homolog protein might have disabled its ability to interact with DIV proteins.

The gymnosperm *P. pinaster* contains two DRIF and four DIV proteins. While PinpDRIF2 does not interact with any of the PinpDIVs, PinpDRIF1 is able to interact with two of the four PinpDIVs indicating that the DIV–DRIF interaction is partly conserved in the gymnosperms (fig. 5). Closer examination to the protein sequence of PinpDRIF2 revealed that the central tyrosine (Y) that composes the DRIF MYB domain topology (-W-X-Y-X-W-) was replaced with a cysteine (C). This event most likely renders the PinpDRIF2 protein unable to interact with any DIV protein. The lack of interaction between PinpDRIF1 and PinpDIV4 is explained by the loss of PinpDIV4 ability to interact with DRIF proteins in general, as it also does not interact with AmDRIF1 (supplementary fig. 10, Supplementary Material online). On the other hand, the inability of PinpDRIF1 to interact with PinpDIV1 suggests that some of the DRIF proteins (such as SmDRIF) might have evolved in such a way that prevented an interaction with a specific DIV protein. In the basal angiosperm *Amborella*, the two DRIF proteins AmbtDRIF1 and AmbtDRIF2 interact with three of the four AmbtDIV homologs (fig. 5), confirming that the DIV–DRIF interaction was inherited by the basal angiosperms.

To determine when the RAD–DRIF interaction was established, the intraspecific interactions between homologs from a gymnosperm (*P. pinaster*) and from an early angiosperm (*Am. trichopoda*) were tested. No interaction was detected between the two PinpDRIFs with any of the six PinpRADs (fig. 5). Curiously, PinpRAD1 and PinpRAD2 interact with the *Antirrhinum* DRIF1, and PinpDRIF1 and PinpDRIF2 interact with *Antirrhinum* RAD, suggesting that both PinpRADs and PinpDRIFs have the ability to establish the RAD–DRIF interaction with the *Antirrhinum* homologs (supplementary fig. 10, Supplementary Material online). Therefore, the inability of the *P. pinaster* RADs and DRIFs to interact might be correlated to a technical problem associated with the coexpression and/or folding of the *P. pinaster* RADs and DRIFs in yeast. To test whether RAD and DRIF homologs from other gymnosperms are able to interact, the *G. biloba* RAD (GbRAD1 and GbRAD2), DIV (GbDIV1 and GbDIV2), and DRIF (GbDRIF1 and GbDRIF2) homologs were cloned and the interactions tested using Y2H. Both GbDIV1 and GbDIV2 were able to interact with GbDRIF1 and GbDRIF2, showing that, similar to *P. pinaster*, the interaction between DIV and DRIF proteins is conserved in gymnosperms (fig. 5). Contrary to *P. pinaster*, however, the *G. biloba* RADs were able to interact with both GbDRIFs, suggesting that the interaction between these proteins is conserved in the gymnosperms (fig.5). The two *Amborella* DRIF proteins, AmbtDRIF1 and AmbtDRIF2, can interact with the RAD homolog AmbtRAD1 (fig. 5), which is suggestive of an interaction between RAD and DRIF being conserved in the early angiosperm lineage.

Taken together, these results suggest that the DIV–DRIF interaction has evolved in the ancestral green lineage and was likely established at the same time that the new DIV and DRIF families and their respective MYB domains emerged. The RAD–DRIF interaction, on the other hand, has evolved later than the DIV–DRIF interaction.

Discussion

The sequencing of a large number of genomes and transcriptomes provides a relevant contribution to the understanding of how transcriptional networks may have been rewired
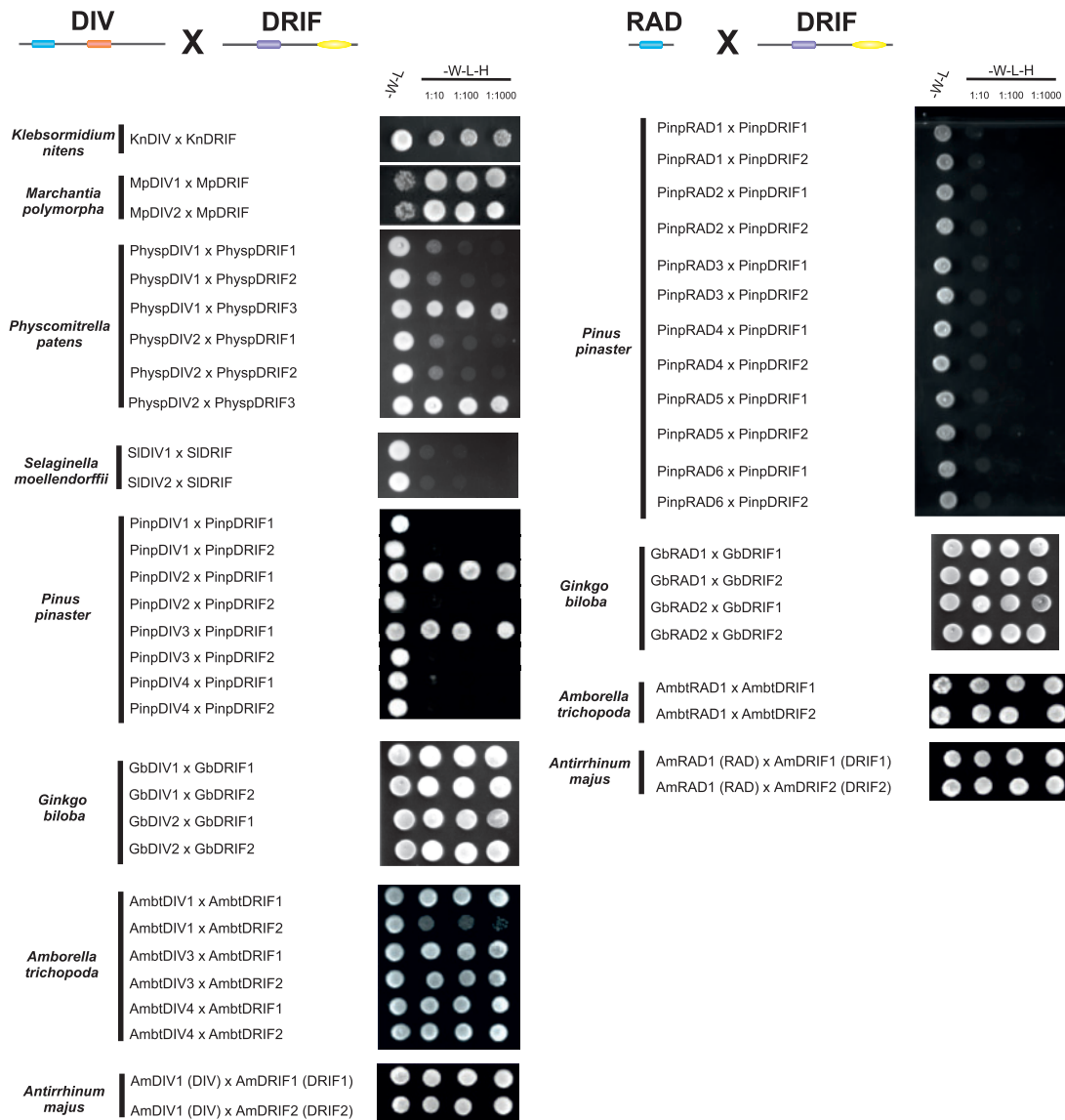
FIG. 5. Evolutionary history of the DIV–DRIF and RAD–DRIF interactions. Intraspecific interactions between the DIV and DRIF (left) and the RAD and DRIF (right) homolog proteins. The species tested were *Klebsormidium nitens* (Kn), *Marchantia polymorpha* (Mp), *Physcomitrella patens* (Physp), *Selaginella moellendorffii* (Sm), *Pinus pinaster* (Pinp), *Ginkgo biloba* (Gb), and *Amborella trichopoda* (Ambt). Yeast growth in -W-L-H medium denotes a positive interaction between the two tested proteins. (-W-L, medium lacking tryptophan and leucine; -W-L-H, medium lacking tryptophan, leucine and histidine; 1:10, 1:100 and 1:1,000 represent the dilution factor applied to the yeast inoculate).

during evolution. However, despite the extensive focus on genomic and transcriptomic data, there are few studies exploring specific molecular mechanisms that drove the establishment of new regulatory modules. The DDR regulatory module has been recruited during evolution to regulate diverse traits in plant biology. The unique domain structure of the DDR proteins and the dynamic interactions between them provide an excellent model to understand how new regulatory modules evolve.

Several genes belonging to the DIV, DRIF, or RAD families have been implicated in the regulation of diverse processes such as the establishment of flower asymmetry in *Antirrhinum* (Almeida et al. 1997; Galego and Almeida 2002;

Corley et al. 2005; Raimundo et al. 2013), the control of cell expansion in the tomato fruit pericarp (Machemer et al. 2011), the regulation of α-amylase gene expression in rice (Lu et al. 2002), or the repression of flowering in *Arabidopsis* (Li et al. 2015). The versatility of the proteins that compose the DDR module is a common attribute of the MYB superfamily of transcription factors (Dubos et al. 2010), such as the regulation of cell cycle (Ito 2001), plant metabolism (Stracke et al. 2007; Gonzalez et al. 2008), cell fate and identity (Lai 2005; Kang et al. 2009), and the response to abiotic and biotic stresses (Cominelli et al. 2005; Raffaele et al. 2008). The functional flexibility of the MYB superfamily is likely associated with variations in the MYB topology and

in the organization and number of MYB repeats that promote the evolution of new DNA and/or protein interactions among new MYB proteins, thus facilitating the emergence of new regulatory networks which, in turn, drives functional diversity (Lynch and Wagner 2008).
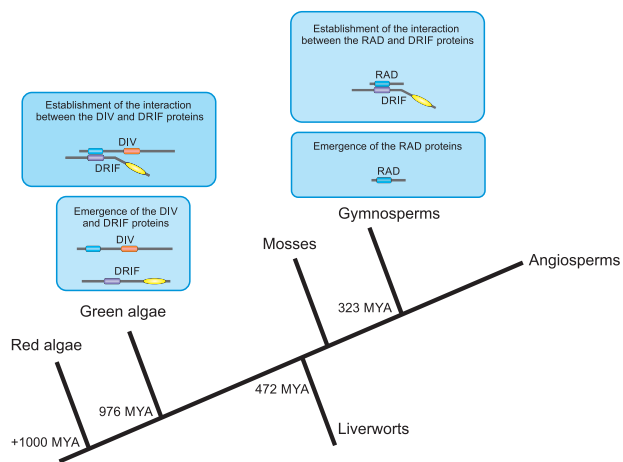
The DIV, DRIF, and RAD protein families are composed of domains with a unique topology that allows for DDR proteins to be easily distinguished from other MYB subfamilies. DIV proteins contain two MYB domains with distinct topologies and functions. The N-terminal MYBI domain is responsible for protein–protein interactions, namely with DRIF proteins, while the C-terminal SHAQKYF-type MYB domain is capable of binding DNA, specifically to an I-box sequence, and of transcription activation (Rose et al. 1999; Raimundo et al. 2013). Members of the SHAQKYF-type MYBs are found in unicellular eukaryotes such as the slime mold *Dictyostelium discoideum*, where a single SHAQKYF domain MYB protein, mybE, plays an important role in cell differentiation (Fukuzawa et al. 2006). SHAQKYF MYB domains are highly conserved across the plant lineage and unicellular eukaryotes, suggesting that they may have conserved an ancestral role throughout the evolution of the plant lineage (Feller et al. 2011). The DIV SHAQKYF-type MYB domain can also be found in other proteins with a single MYB domain, denominated by DIVL, that also contain an R motif (R/KLFGV), identified as a repressor of transcription and present in various transcription factor families (Ikeda and Ohme-Takagi 2009). So far, the characterized DIVL proteins seem to have similar or complementary functions relatively to the DIV proteins (Lu et al. 2002), suggesting that the SHAQKYF-type MYB domain of the two protein families is highly conserved and binds to the same regulatory sequences. Interestingly, we showed that both DIV and DIVL protein family members are first present in green algae suggesting that the evolution of DIV and DIVL may be functionally interconnected. Thus, the appearance of the MYBI domain of the DIV proteins might have contributed to add another level of regulatory plasticity by establishing interactions with other proteins such as members of the DRIF protein family. The small size of the RAD proteins (∼99 a.a.) and their particular mode of action as antagonistic agents in the establishment of the DIV–DRIF complex classify them as siPEP. siPEPs have their origin on transcription factors that have lost the DNA-binding domain but maintain a protein interaction domain and act as postregulatory inhibitors by forming nonfunctional heterodimers with their targets. siPEPs have evolved under two different mechanisms, they can lose the DNA-binding domain by alternative splicing (Roman et al. 1991) or, alternatively, can be derived from a duplication of a functional transcription factor followed by deletion of the DNA-binding domain (Wenkel et al. 2007). Similar to most siPEPs (Seo et al. 2011), the RAD protein family has emerged in an ancestral lineage positioned between the ferns and the gymnosperms and has duplicated throughout the seed plants.

A critical event in the establishment of the DDR transcription module was the emergence of the MYB domains that allowed for the combinatorial interactions to occur between the DIV, DRIF, and RAD proteins. Surprisingly, the MYB domains, which show a different topology in DIV and DRIF proteins, have most likely emerged from the duplication of a common MYB ancestor during the evolution of the green algae (fig. 4). The phylogenetic analysis performed on the MYB domains from green algae proteins has revealed a closer homology between the MYBI domain of DIV and the MYB domain of DRIF with the SHAQKYF MYB group rather than with other MYB subfamilies (fig. 4). Interestingly, these results imply that a domain responsible for protein–protein interactions might have evolved from a structurally similar domain but with a DNA-binding function. It is then possible that a whole DIV gene duplicated, lost the SHAQKYF MYB domain and gave rise to the DRIF ancestor, or only the MYBI DIV domain has duplicated into a new locus, which by gene fusion added the DUF3755 domain. The latter hypothesis is supported by the fact that no traces of the MYBII DIV sequence can be found on ancient DRIF genes and by the presence of a completely formed DUF3755 domain in a protein containing also a MYBI domain in green algae species. Gene fusion has been described as the dominant mechanism in creating domain rearrangements. After duplication, a gene encoding a protein with one or more domains is joined with an adjacent gene thus increasing the variability and number of domain combination in the proteome (Buljan et al. 2010). However, the DUF3755 domain was never identified in algae proteins without an adjacent MYB domain, thus the gene fusion theory still needs additional evidence.

The RAD protein family has most likely arisen by gene duplication and domain loss, which seems to be one of the main evolutionary processes driving the evolution of siPEPs together with alternative splicing (Seo et al. 2011). Before the emergence of seed plants, a DIV ancestral gene most likely duplicated and progressively lost the C-terminal SHAQKYF-type MYB domain. The RAD genes underwent a high number of independent duplications especially after the split between gymnosperms and angiosperms. The fast duplication of the RAD protein family indicates that the novel siPEP might have acquired new functions that were positively selected. This observation is in line with the suggestion that protein sequestration generated by antagonistic factors such as RAD can generate flexible and highly tuneable ultrasensitive responses in genetic networks and therefore promotes a big impact on the evolution of genetic circuits (Buchler and Cross 2009).

The functionality of the DDR regulatory module as a molecular switch is based on the interaction dynamic established between the DIV, DRIF, and RAD proteins. Our results revealed that members of the DIV and DRIF protein families are already present in the green algae and that they are able to interact. The physical interaction between DIV and DRIF homologs is maintained in early land plants *M. polymorpha* and *Ph. patens*. The key event that led to the establishment of the DIV–DRIF interaction was most likely a rearrangement of an ancient MYB domain during the evolution of the green algae that created the interaction between these two MYB proteins (fig. 6). The RAD–DRIF interaction has likely been established simultaneously with the appearance of the RAD family in the gymnosperm lineage, thus, our results suggest

FIG. 6. Diagram of the origin of the DIV, DRIF, and RAD protein families and the establishment of the DDR regulatory module. The DIV and DRIF protein families have their origin in an ancestral lineage that originated the green algae, and a DIV–DRIF interaction is established before the emergence of the first land plants. The RAD protein family has evolved much later in a lineage that gave rise to the gymnosperms thus forming the DDR regulatory module.

that the DIV–DRIF regulatory complex evolved much earlier than the antagonistic RAD–DRIF interaction (fig. 6).

A similar evolutionary process to the DDR module has occurred with the emergence of the WD40–bHLH–MYB regulatory module. This module is composed by members of three protein families (WD40, bHLH, and MYB) that precede the plant lineage. However, the complex between these proteins only appears to have been recruited in land plants to specify epidermal cell fate (Serna 2004; Ramsay and Glover 2005). The plasticity of the WD40–bHLH–MYB is mainly conferred by variations on the type of MYB proteins that interact with the bHLH proteins, which in turn are bound to the WD40 protein that serves as a scaffold to the whole complex. The analysis of the function of the DDR proteins in early plant species such as *M. polymorpha* or *Ph. patens* will contribute to reveal the ancient role of the DIV and DRIF proteins and of the transcription module and help to understand how their function was modulated during plant evolution.

In conclusion, our work provides a molecular depiction of how a new regulatory module can evolve by determining the origin of the protein families that it comprises and the timing of the establishment of their interactions. Our results indicate that the successive rearrangement and divergence of a single MYB domain gave rise, at different evolutionary points, to the DIV, DRIF, and RAD protein families. The members of the DIV and DRIF protein families have appeared in the green algae lineage and the DIV–DRIF interaction was established as early as in the green algae and was conserved during the evolution of the first land plants. The interaction between RAD and DRIF, however, is first detected in the gymnosperms, showing that the DIV–DRIF regulatory heterodimer has evolved prior to the antagonistic RAD–DRIF interaction (fig. 6). The study of the evolution of the DDR module thus provides a deeper and detailed

understanding of the molecular mechanisms underlying the establishment of novel regulatory modules.

Materials and Methods

Sequence Retrieval

DIV, DIVL, DRIF, and RAD homologous protein sequences were obtained by performing BLAST searches on the plant gene index database (http://compbio.dfci.harvard.edu/tgi/), on the *Am. trichopoda* genome database (http://amborella.huck.psu.edu, the phytozome portal (http://phytozome.jgi.doe.gov), the *K. nitens* genome project (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/), the Gymno PLAZA 1.0 database and the NCBI database (http://www.ncbi.nlm.nih.gov/), http://medicinalplantgenomics.msu.edu, and the fern database (https://www.fernbase.org/). The searches for the sequences for DIV, DIVL, DRIF, and RAD homologous protein sequences were performed using the conserved domains for each of the protein families. All the sequences are identified on supplementary table 1, Supplementary Material online, with the respective gene codes and corresponding gene names. To search for the entire MYB family in *Chla. reinhardtii*, *V. carteri*, *K. nittens*, *P. pinaster*, *Gnetum montanum*, *G. biloba*, and *Pic. sitchensis*, a profile search on HMMER3 (http://hmmer.janelia.org/) was performed using the seed alignments generated from Pfam (Finn et al. 2014) for the MYB domain (PF00249). The list of retrieved green algae MYB proteins was complemented with the sequences provided by Du et al. (2013) and sequences deposited in the PlantTFDB (http://planttfdb.cbi.pku.edu.cn/).

Phylogenetic Analysis

Protein sequences were aligned with MUSCLE (Edgar 2004) and the ambiguously aligned regions excluded to produce the final protein alignments used to construct the phylogenetic trees. Evolutionary relationships, using all the amino acids within the conserved domains, were inferred by maximum likelihood under the Jones–Taylor–Thornton substitution model, assuming a gamma distribution and 1,000 bootstrap replicates using the MEGA6 software (Tamura et al. 2013). Alignments were analyzed using the JALVIEW software (Waterhouse et al. 2009). The presented phylogenetic trees were the most likely trees.

Plasmid Construction and Yeast Two-Hybrid Analysis

Total RNA was extracted from *M. polymorpha*, *P. pinaster*, *G. biloba*, and *A. majus* using TRIzol reagent (Invitrogen). SuperScript III (Invitrogen) and oligo (dT) were used to retro-transcribe 1 μg of RNA, according to the manufacturer's instructions.

Open reading frames of the different *DRIF*, *RAD*, and *DIV* homologous genes were amplified from plasmids containing synthesized coding sequences of *Se. moellendorffii* and *K. nitens* or from cDNA samples of *M. polymorpha* (archegonia and antheridia), *Ph. patens* (protonema), *P. pinaster* (needles and flower buds), *G. biloba* (leaves), *Am. trichopoda* (leaf, root, shoot tip, stem and flower), and *A. majus* (flower) using

specific primers (supplementary table 2, Supplementary Material online). Amplified sequences were cloned into pGBT9 (bait vector; Clontech) and pGAD424 (prey vector; Clontech) using restriction enzymes or gap-repair cloning in yeast (Ma et al. 1987). Protein–protein interactions were analyzed using a GAL-4-based yeast hybrid system (Matchmaker two-hybrid system; Clontech). Proteins fused to the binding domain of GAL4 were tested for self-activation by monitoring growth of transformed cells in SD medium without histidine (plus 5 mM of 3-amino triazole). Different prey and bait vector combinations were then used to transform *Saccharomyces cerevisiae* strain AH109 using the LiAc/DNA/PEG transformation method (Gietz et al. 1995). Each experiment was replicated three times. Selection of positive interactions was performed according to Causier and Davies (2002).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

References

Almeida J, Rocheta M, Galego L. 1997. Genetic control of flower shape in *Antirrhinum majus*. *Development* 124(7): 1387–1392.

Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. 2008. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci*. 33(5): 220–229.

Bartlett ME, Whipple CJ. 2013. Protein change in plant evolution: tracing one thread connecting molecular and phenotypic diversity. *Front Plant Sci*. 4(382): 1–16.

Boyden GS, Donoghue MJ, Howarth DG. 2012. Duplications and expression of RADIALIS-like genes in Dipsacales. *Int J Plant Sci*. 173(9): 971–983. Available from: http://www.jstor.org/stable/10.1086/667626

Breuninger H, Thamm A, Streubel S, Sakayama H, Nishiyama T, Dolan L. 2016. Diversification of a transcription factor family led to the evolution of antagonistically acting genetic regulators of root hair growth. *Curr Biol*. 26(12): 1622–1628.

Brkljacic J, Grotewold E. 2017. Combinatorial control of plant gene expression. *Biochim Biophys Acta Gene Regul Mech*. 1860(1): 31–40. Available from: http://linkinghub.elsevier.com/retrieve/pii/ S1874939916301377

Buchler NE, Cross FR. 2009. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol Syst Biol*. 5(272): 1–7.

Buljan M, Frankish A, Bateman A. 2010. Quantifying the mechanisms of domain gain in animal proteins. *Genome Biol*. 11(7): R74. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/ gb-2010-11-7-r74

Causier B, Davies B. 2002. Analysing protein-protein interactions with the yeast two-hybrid system. *Plant Mol Biol*. 50(6): 855–870.

Cho HT. 2017. An ancient transcriptional regulatory module for tip growth has been conserved throughout the vascular plant lineage. *Plant Signal Behav*. 12(3): e1294300.

Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL, Tonelli C. 2005. A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Curr Biol*. 15(13): 1196–1200.

Corley SB, Carpenter R, Copsey L, Coen E. 2005. Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. *Proc Natl Acad Sci U S A*. 102(14): 5068–5073.

Costa MMR, Fox S, Hanna AI, Baxter C, Coen E. 2005. Evolution of regulatory interactions controlling floral asymmetry. *Development* 132(22): 5093–5101.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401(6749): 157–161.

Du H, Wang Y-B, Xie Y, Liang Z, Jiang S-J, Zhang S-S, Huang Y-B, Tang Y-X. 2013. Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res*. 20(5): 437–448.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci*. 15(10): 573–581.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5): 1792–1797.

Eguen T, Straub D, Graeff M, Wenkel S. 2015. MicroProteins: small size–big impact. *Trends Plant Sci*. 20(8): 477–482.

Feller A, MacHemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J*. 66(1): 94–116.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: The protein families database. *Nucleic Acids Res*. 42 (D1): D222–230.

Fukuzawa M, Zhukovskaya NV, Yamada Y, Araki T, Williams JG. 2006. Regulation of *Dictyostelium* prestalk-specific gene expression by a SHAQKY family MYB transcription factor. *Development* 133(9): 1715–1724. Available from: http://www.ncbi.nlm.nih.gov/pubmed/ 16571632

Galego L, Almeida J. 2002. Role of DIVARICATA in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Genes Dev*. 16(7): 880–891.

Gao A, Zhang J, Zhang W. 2017. Evolution of RAD- and DIV-like genes in plants. *Int J Mol Sci*. 18(9): 1961.

Gietz RD, Schiestl RH, Willems AR, Woods RA. 1995. Studies on the transformation of intact yeast cells by the LiAc/SSDNA/PEG procedure. *Yeast* 11(4): 355–360.

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J*. 53(5): 814–827.

Hamaguchi A, Yamashino T, Koizumi N, Kiba T, Kojima M, Sakakibara H, Mizuno T. 2008. A small subfamily of *Arabidopsis* RADIALIS-LIKE SANT/MYB genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. *Biosci Biotechnol Biochem*. 72(10): 2687–2696.

Han JDJ, Berlin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, et al. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430: 88–93.

Howarth DG, Donoghue MJ. 2009. Duplications and expression of DIVARICATA-like genes in dipsacales. *Mol Biol Evol*. 26(6): 1245–1258.

Ikeda M, Ohme-Takagi M. 2009. A novel group of transcriptional repressors in *Arabidopsis*. *Plant Cell Physiol*. 50(5): 970–975.

Ito M. 2001. G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. *Plant Cell*

13(8): 1891–1905. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.13.8.1891

Jin J, Xie X, Chen C, Park JG, Stark C, James DA, Olhovsky M, Linding R, Mao Y, Pawson T. 2009. Eukaryotic protein domains as functional units of cellular evolution. Sci Signal. 2(98): ra76. Available from: http://stke.sciencemag.org/cgi/doi/10.1126/scisignal.2000546

Kang YH, Kirik V, Hulskamp M, Nam KH, Hagely K, Lee MM, Schiefelbein J. 2009. The MYB23 gene provides a positive feedback loop for cell fate specification in the Arabidopsis root epidermis. Plant Cell 21(4): 1080–1094. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.108.063180

Kosugi S, Ohashi Y. 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J. 30(3): 337–348.

Lai LB. 2005. The Arabidopsis R2R3 MYB proteins FOUR LIPS and MYB88 restrict divisions late in the stomatal cell lineage. Plant Cell 17:2754–2767. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.105.034116

Li C, Zhou Y, Fan LM. 2015. A novel repressor of floral transition, MEE3, an abiotic stress regulated protein, functions as an activator of FLC by binding to its promoter in Arabidopsis. Environ Exp Bot. 113:1–10.

Lu C-A, Ho TD, Ho S-L, Yu S-M. 2002. Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of alpha-amylase gene expression. Plant Cell 14(8): 1963–1980.

Lu SX, Knowles SM, Andronis C, Ong MS, Tobin EM. 2009. CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of Arabidopsis. Plant Physiol. 150(2): 834–843. Available from: http://www.plantphysiol.org/cgi/doi/10.1104/pp.108.133272

Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E. 1999. Control of organ asymmetry in flowers of Antirrhinum. Cell 99(4): 367–376.

Luo D, Carpenter R, Vincent C, Copsey L, Coen E. 1996. Origin of floral asymmetry in Antirrhinum. Nature 383(6603): 794–799.

Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in developmental evolution. Evolution (N. Y) 62(9): 2131–2154.

Ma H, Kunes S, Schatz PJ, Botstein D. 1987. Plasmid construction by homologous recombination in yeast. Gene 58(2–3): 201–216.

Machemer K, Shaiman O, Salts Y, Shabtai S, Sobolev I, Belausov E, Grotewold E, Barg R. 2011. MYB factors interplay in differential cell expansion and consequences for tomato fruit development. Plant J. 68(2): 337–350. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21707804

Morgenstern B, Atchley WR. 1999. Evolution of bHLH transcription factors: modular evolution by domain shuffling? Mol Biol Evol. 16:1654–1663. Available from: papers3://publication/uuid/0B6BEB8B-009E-4DED-9E02-687890D0430D

Pasek S, Risler JL, Brézellec P. 2006. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics 22(12): 1418–1423.

Pawson T. 1995. Protein modules and signalling networks. Nature 373:573–580. Available from: http://www.nature.com/doifinder/10.1038/373573a0

Pires ND, Yi K, Breuninger H, Catarino B, Menand B, Dolan L. 2013. Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. Proc Natl Acad Sci U S A. 110(23): 9571–9576. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1305457110

Proost S, Mutwil M. 2016. Tools of the trade: studying molecular networks in plants. Curr Opin Plant Biol. 30:143–140.

Raffaele S, Vailleau F, Leger A, Joubes J, Miersch O, Huard C, Blee E, Mongrand S, Domergue F, Roby D. 2008. A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in Arabidopsis. Plant Cell 20(3): 752–767. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.107.054858

Raimundo J, Sobral R, Bailey P, Azevedo H, Galego L, Almeida J, Coen E, Costa MMR. 2013. A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in Antirrhinum flower asymmetry. Plant J. 75(4): 527–538.

Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10(2): 63–70.

Roman C, Cohn L, Calame K. 1991. A dominant negative form of transcription activator mTFE3 created by differential splicing. Science 254(5028): 94–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1840705

Rose A, Meier I, Wienand U. 1999. The tomato I-box binding factor LeMYBI is a member of a novel class of Myb-like proteins. Plant J. 20(6): 641–652.

Schmidt EE, Davies CJ. 2007. The origins of polypeptide domains. BioEssays 29(3): 262–270.

Seo PJ, Hong S-Y, Kim S-G, Park C-M. 2011. Competitive inhibition of transcription factors by small interfering peptides. Trends Plant Sci. 16: 541–549. Available from: http://www.sciencedirect.com/science/article/pii/S1360138511001270%5Cnhttp://pdn.sciencedirect.com/science?_ob=MiamiImageURL&_cid=271203&_user=607988&_pii=S1360138511001270&_check=y&_origin=article&_zone=toolbar&_coverDate=31-Oct-2011&view=c&originContent

Serna L. 2004. A network of interacting factors triggering different cell fates. Plant Cell 16(9): 2258–2263.

Serrano-Bueno G, Romero-Campero FJ, Lucas-Reina E, Romero JM, Valverde F. 2017. Evolution of photoperiod sensing in plants and algae. Curr Opin Plant Biol. 37:10–17.

Shikata M, Narumi T, Yamaguchi H, Sasaki K, Aida R, Oshima Y, Takiguchi Y, Ohme-Takagi M, Mitsuda N, Ohtsubo N. 2011. Efficient production of novel floral traits in torenia by collective transformation with chimeric repressors of Arabidopsis transcription factors. Plant Biotechnol. 28(2): 189–199.

Staudt A-C, Wenkel S. 2011. Regulation of protein function by 'microProteins'. EMBO Rep. 12(1): 35–42. Available from: http://embor.embopress.org/cgi/doi/10.1038/embor.2010.196

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. 2007. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. Plant J. 50(4): 660–677.

Su S, Xiao W, Guo W, Yao X, Xiao J, Ye Z, Wang N, Jiao K, Lei M, Peng Q, et al. 2017. The CYCLOIDEA–RADIALIS module regulates petal shape and pigmentation, leading to bilateral corolla symmetry in Torenia fournieri (Linderniaceae). New Phytol. 215(4): 1582–1593.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30(12): 2725–2729.

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al. 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403(6770): 623–627. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10688190

van der Graaff E, Laux T, Rensing SA. 2009. The WUS homeobox-containing (WOX) protein family. Genome Biol. 10(12): 248. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-12-248

Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. 2004. Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol. 14(2): 208–216.

Wang ZY. 1997. A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis Lhcb gene. Plant Cell 9(4): 491–507. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.9.4.491

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2: a multiple sequence alignement editor and analysis workbench. Bioinformatics 25(9): 1189–1191. Available from: http://www.zotero.org/herimbert/items/139086727

Wenkel S, Emery J, Hou B-H, Evans MMS, Barton MK. 2007. A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes. Plant Cell 19(11): 3379–3390. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.107.055772