Updated Phylogeny and Protein Structure Predictions Revise the Hypothesis on the Origin of MADS-box Transcription Factors in Land Plants

Yichun Qiu,[†,1,2] Zhen Li [†,3,4] Dirk Walther [1] and Claudia Köhler [*,1,2]

[1]Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany
[2]Swedish University of Agricultural Sciences & Linnean Center for Plant Biology, Uppsala BioCenter, Uppsala, Sweden
[3]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
[4]VIB Center for Plant Systems Biology, VIB, Ghent, Belgium
[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: Koehler@mpimp-golm.mpg.de.
Associate editor: Dr. Mary O'Connell

Abstract

MADS-box transcription factors (TFs), among the first TFs extensively studied, exhibit a wide distribution across eukaryotes and play diverse functional roles. Varying by domain architecture, MADS-box TFs in land plants are categorized into Type I (M-type) and Type II (MIKC-type). Type I and II genes have been considered orthologous to the SRF and MEF2 genes in animals, respectively, presumably originating from a duplication before the divergence of eukaryotes. Here, we exploited the increasing availability of eukaryotic MADS-box sequences and reassessed their evolution. While supporting the ancient duplication giving rise to SRF- and MEF2-types, we found that Type I and II genes originated from the MEF2-type genes through another duplication in the most recent common ancestor (MRCA) of land plants. Protein structures predicted by AlphaFold2 and OmegaFold support our phylogenetic analyses, with plant Type I and II TFs resembling the MEF2-type structure, rather than SRFs. We hypothesize that the ancestral SRF-type TFs were lost in the MRCA of Archaeplastida (the kingdom Plantae *sensu lato*). The retained MEF2-type TFs acquired a Keratin-like domain and became MIKC-type before the divergence of Streptophyta. Subsequently in the MRCA of land plants, M-type TFs evolved from a duplicated MIKC-type precursor through loss of the Keratin-like domain, leading to the Type I clade. Both Type I and II TFs expanded and functionally differentiated in concert with the increasing complexity of land plant body architecture. The recruitment of these originally stress-responsive TFs into developmental programs, including those underlying reproduction, may have facilitated the adaptation to the terrestrial environment.

Key words: **MADS-box transcription factors, land plants, MEF2, Type I/M-type, gene duplication.**

Introduction

MADS-box genes are a famous and intriguing gene family, broadly present in eukaryotic genomes. They encode transcription factors (TFs) which regulate diverse and important biological functions as reported in animals, fungi, plants, and protists (Messenguy and Dubois 2003). The name derives from the four founding members, Minichromosome maintenance 1 (Mcm1) from *Saccharomyces cerevisiae*, AGAMOUS from *Arabidopsis thaliana*, DEFICIENS from *Antirrhinum majus*, and Serum response factor (SRF) from *Homo sapiens* (Schwarz-Sommer et al. 1990). Animal genomes generally have two types of MADS-box genes, SRF and myocyte enhancer factor-2 (MEF2) genes that are present in one to a few copies. The budding yeast *Saccharomyces cerevisiae* has four MADS-box TFs; Mcm1 and Arg80 are related to the animal SRF, and Rlm1 and Smp1 are related to MEF2. Several phylogenetic analyses inferred the origin of SRF and MEF2 types through an ancient gene duplication event before the divergence of eukaryotes (fig. 1*a*) (Theissen et al. 1996; Alvarez-Buylla et al. 2000; Gramzow et al. 2010). The two types can be distinguished by the domains downstream of the MADS domain; while SRF-type TFs are characterized by a SAM domain (SRF, ARG80, and MCM1), the corresponding region in MEF2-type TFs is referred to as the MEF2 domain (Shore and Sharrocks 1995). The difference is reflected by the resolved crystal structures of several MADS-box TFs, including human SRF and budding yeast MCM1, human MEF2A, and mouse MEF2C. Required to build up the interface for TF dimerization and DNA-binding, the conserved MADS domain comprises an alpha helix and two antiparallel beta strands, while the SAM or MEF2 domain constitutes a second alpha helix. Remarkably, a kink present in the SAM domain of

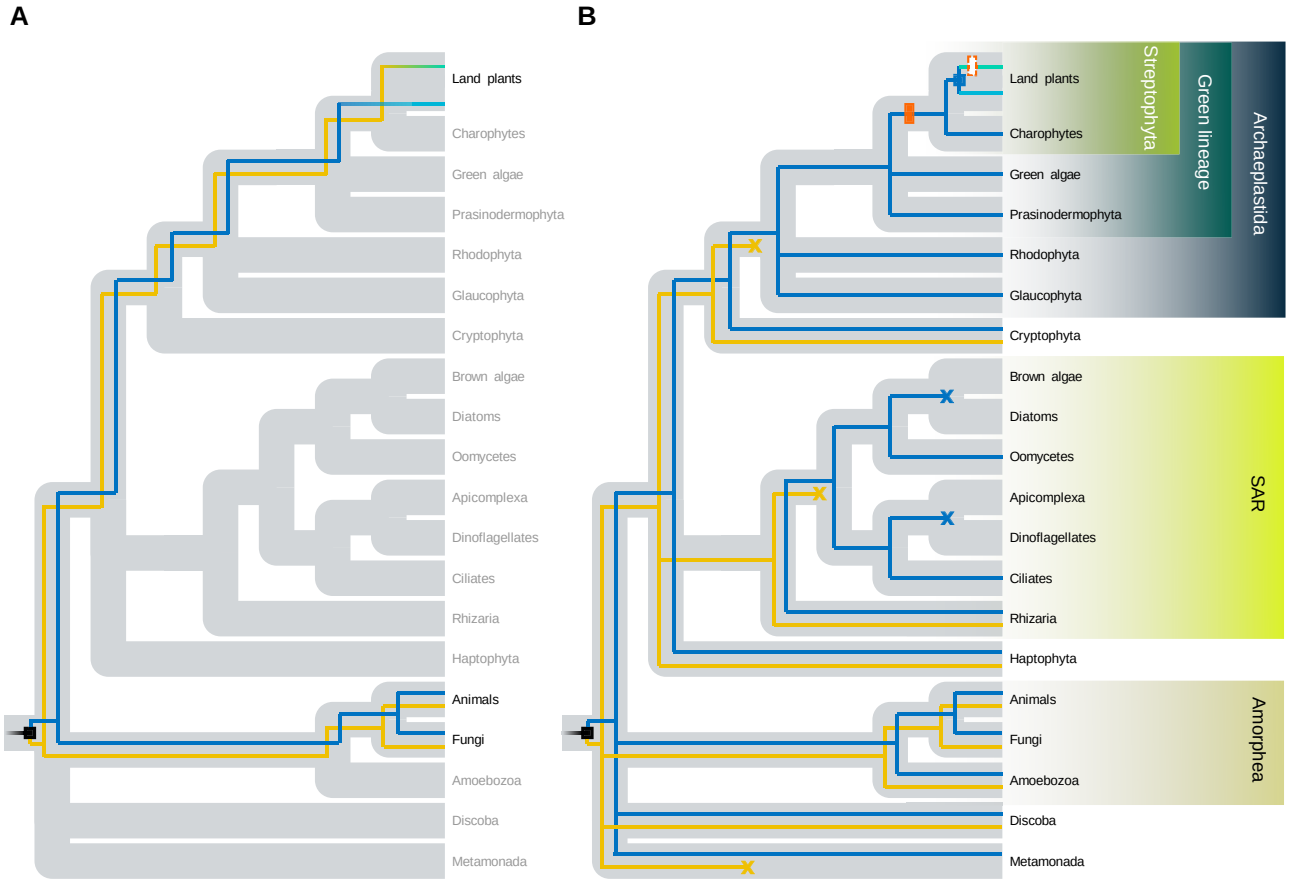

FIG. 1. The models of MADS-box transcription factor (TF) evolution. (*A*) The model by Alvarez-Buylla et al. (2000): Upon the identification of M-type MADS-box TFs in the genome of *Arabidopsis thaliana*, they were named Type I MADS-box TFs and considered orthologous to SRF-type TFs in animals and fungi, while MIKC-type in plants were grouped as Type II and clustered with MEF2-type in animals and fungi. Type I and II MADS-box genes originated by a hypothesized ancient duplication event predating the plant and animal divergence. (*B*) The model proposed in this study: With a broad survey across eukaryotes, the ancient duplication giving rise to SRF- and MEF2-type can be inferred as early as the origin of the most common recent ancestor of all living eukaryotic groups (consistent with model a). In several extant lineages both types have been retained; however, in the lineage of Archaeplastida, SRF-type TFs were lost, and only MEF2-type TFs continued evolving. Subsequently, Type I and II MADS-box TFs originated by duplication of a MEF2-type precursor in the most recent common ancestor of all land plants. Black, primitive MADS-box TFs; yellow, SRF-type; blue, MEF2-type; green, Type I; cyan, Type II. Square, gene duplication event; cross, gene loss. Orange bar, gain of the K domain; empty orange bar, loss of the K domain.

SRF-type TFs turns the orientation of the second helix in the opposite direction to that of MEF2-type TFs (fig. 2*a*).

In land plants, MADS-box TFs (also referred to as AGAMOUS-like in the model plant *Arabidopsis thaliana*) have evolved to be a flourishing family with typically as many as 50 to over 100 members in plants, especially angiosperm species (Gramzow and Theissen 2013), in sharp contrast to the much smaller family sizes of MADS-box TFs in other eukaryotes, typically comprising only two to five members (supplementary table S1, Supplementary Material online). According to the domain architecture, the MADS-box TFs in land plants can be specified into two types: the Type I TFs are usually referred to as M-type, since they share no well-characterized conserved domain following the MADS domain, while the Type II TFs typically have the MADS domain followed by the Intervening, Keratin-like and C-terminal domains, so they are also known as MIKC-type. The I domain is analogous to the SAM or MEF2 domain in animal MADS-box TFs in terms of location and function, and the K domain is likely specific for plants (Alvarez-Buylla et al. 2000; Parenicová et al. 2003). Due to their critical roles in establishing floral organ identity in angiosperms, the Type II MADS-box genes have been studied extensively. In contrast, Type I MADS-box genes have only been identified along with the first angiosperm genome of *A. thaliana*. Emerging studies have linked the functions of Type I genes in several angiosperm species mainly to the development of the female gametophyte and endosperm (Köhler et al. 2003; Kang et al. 2008; Bemer et al. 2010; Masiero et al. 2011; Paul et al. 2020; Qiu and Köhler 2022). Besides the distinct domain arrangements of their encoded proteins, land plants Type I and II MADS-box genes vary in expression patterns, numbers of exons, and noticeably, compared with Type II MADS-box genes, Type I genes have evolved faster and more frequently undergone duplication and loss (Parenicová et al. 2003; Nam et al. 2004). Collectively, among many known regulatory functions, MADS-box TFs act as

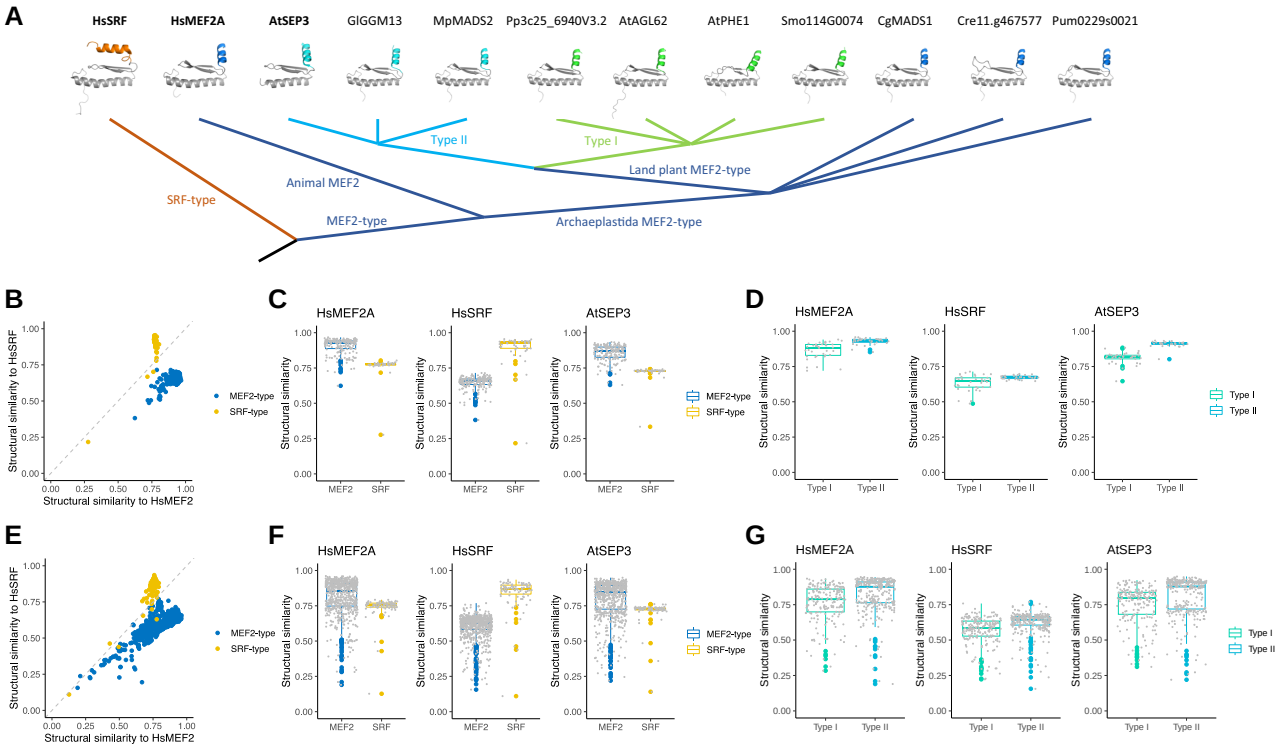

Fig. 2. Protein analyses of MADS-box TFs. (*A*) Crystal structures (bold) of human SRF and MEF2A, and Arabidopsis SEP3 that were used as templates for structural comparisons (A-chains only), and AlphaFold2-predicted structural models of Archaeplastida MADS-box proteins including Type I and II TFs in land plants. Models were trimmed to the relevant structural segments matching the template structures. Structures were drawn using Pymol (Schrödinger and DeLano 2020). (*B*,*C*) Similarity scores (TM-scores, with "1" indicating high (perfect), "0" no (random) agreement) of all predicted structural models by AlphaFold2 to human SRF and MEF2A, and Arabidopsis SEP3. SRF-type (yellow) and MEF2-type (blue) TFs were inferred by phylogeny. (*D*) Similarity scores of predicted structural models by AlphaFold2 for land plant MADS-box TFs to human SRF and MEF2A, and Arabidopsis AtSEP3. (*E*,*F*,*G*) Similarity scores of predicted structural models by OmegaFold.

major regulators of plant reproduction and have been closely connected with the rise of flowering plants to ecological dominance (Ng and Yanofsky 2001; Kaufmann et al. 2005). The family size of MADS-box genes has been linked to the complexity of the plant body plan (Theissen et al. 1996; Kaufmann et al. 2005; Thangavel and Nayar 2018). Thus, identifying the origin and resolving the subsequent diversification of plant MADS-box genes is required to understand the evolutionary success of land plants.

Upon the discovery of Type I MADS-box genes, a timely survey suggested that Type I and II genes in plants are orthologous to the SRF and MEF2 genes in animals, respectively (Alvarez-Buylla et al. 2000). Based on this model, an ancient duplication before the divergence of the extant eukaryotic lineages gave birth to the two classes of MADS-box genes in plants, likewise in animals and other eukaryotes (fig. 1*a*). This model has been influential in the field of MADS-box evolution and served as a basis for investigations of MADS-box gene evolution across all phylogenetic scales. As already noted by the authors of the model (Alvarez-Buylla et al. 2000), however, the clustering of Type I TFs in plants and SRF-type TFs in animals and fungi is not well supported in the original study based on only a few Arabidopsis, animal, and fungal sequences. Although caught by some thoughtful critique (De Bodt et al. 2003; Kaufmann et al. 2005), this pitfall has been neglected since then.

The burst of available genome sequences of plants as well as other major clades of eukaryotes encouraged us to revisit the origin of MADS-box genes in land plants. In particular the genomes of Charophytes, the paraphyletic algal relatives of land plants, have shed light on the evolution of gene families underlying the successful terrestrialization of land plants (Hori et al. 2014; Nishiyama et al. 2018). There, in the charophyte *Klebsormidium flaccidum*, the only present MADS-box gene belongs to Type II, coding for a MIKC-type TF (Hori et al. 2014). Similarly, in the *Chara braunii* genome, the three identified MADS-box genes belong to Type II, since they are all related to MEF2 genes (Nishiyama et al. 2018). Furthermore, in the genomes of the green algae *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Ostreococcus lucimarinus* and the red algae *Cyanidioschyzon merolae*, the annotated MADS-box genes are identified as MEF2-type, despite that the encoded TFs lack a K domain (Kaufmann et al. 2005; Thangavel and Nayar 2018). Surprisingly, SRF-type MADS-box genes have so far never been found in the charophycean, green, or red algae. If Type I MADS-box genes in land plants are descendants of the ancestral SRF-type genes as stated by the original model (Alvarez-Buylla et al. 2000), the orthologous

SRF-type genes would have been lost convergently and repeatedly in all of successive sister groups to land plants, the paraphyletic algal relatives. Thus, the systematic lack of SRF-type TFs in these lineages challenges the orthology between Type I MADS-box TFs in land plants and SRF-type TFs in animals and fungi. Enabled by the newly available genome sequences and the unprecedented power to predict protein structures based on amino acid sequences, we reassessed the origin of plant Type I MADS-box genes and propose an alternative model explaining the evolution of MADS-box genes in the green lineages.

Results

Ancient Duplication of SRF and MEF2 Clades is Supported Using Newly Available Genome Sequences

Capitalizing on newly available genome sequences spanning the broad phylogeny of eukaryotes, we re-evaluated the original phylogenetic model of MADS-box gene evolution with extended sample sequences. We collected published genomes of different lineages of eukaryotes and identified MADS-box genes in 175 species (supplementary table S1, Supplementary Material online). The species represent seven currently accepted eukaryotic groups: 1) Archaeplastida (the kingdom Plantae *sensu lato*, e.g., streptophytes including land plants and its closest algal sister group Zygnematophyceae, green algae, Prasinodermophyta, red algae, and glaucophytes), 2) Cryptista, 3) Haptista, 4) SAR supergroup which are Stramenopila (e.g., brown algae, diatoms, oomycetes)/Alveolata (e.g., ciliates, dinoflagellates, Apicomplexa)/Rhizaria, 5) Amorphea (e.g., animals, fungi, amoebae), 6) Discoba, and 7) Metamonada. The latter two were previously collectively referred to as Excavata (Burki et al. 2020).

We selected protein sequences of MADS-box TFs from the diverse eukaryotic groups. To perform multiple sequence alignments (MSAs), we extracted the conventional MADS domain sequences of about 60 amino acids in length as defined in previous studies (Shore and Sharrocks 1995) extended by the corresponding regions of SAM/MEF2/I domains, so that the extended MSAs structurally cover the functional unit of two alpha helices and the connecting beta strands. We inferred phylogenetic trees with maximum likelihood, Bayesian inference, and neighbor-joining. All three methods consistently found that the MADS domain sequences naturally form two major clades, corresponding to current SRF and MEF2 lineages, as referenced by the known SRF and MEF2 sequences from animals, fungi, and amebae (fig. 3 and 4; supplementary fig. S1 and S2, Supplementary Material online). This finding supports the overarching hypothesis that an ancient duplication of a MADS-box gene gave rise to the SRF-type and MEF2-type precursors. We found both SRF- and MEF2-types genes in nearly all of the surveyed Amorphea species, as well as Cryptista and Haptista. Furthermore, two species in Discoba, representing a distinct group distantly diverged from the plant and animal lineages, have both types of MADS-box genes (supplementary table S1, Supplementary Material online). Together, our results provide supporting evidence for the presence of SRF- and MEF2-type MADS-box genes early before the diversification of major groups of extant eukaryotes.

Land Plant Type I and II MADS-box TFs are Both MEF2-type

Previous studies concluded that Type I MADS-box genes in land plants are more closely related to the SRF genes in animals (fig. 1*a*) (Alvarez-Buylla et al. 2000; Nam et al. 2003). In contrast, our phylogenetic analyses using three different phylogenetic approaches consistently suggested that Type I MADS-box TFs in land plants clustered within the MEF2 clade, which includes plant Type II TFs (figs. 3 and 4; supplementary figs. S1–S3, Supplementary Material online). Thus, both Type I and II genes are inferred to be MEF2-type and no SRF-type gene is present in the extant land plants. In addition, we carried out approximate unbiased (AU) tests (Shimodaira 2002) to compare the two competing phylogenetic trees of MADS-box TFs. One topology represents our new phylogeny, which groups plant Type I and II both within the MEF2 clade; the other is a constraint phylogeny forcing the plant Type I clade into the SRF clade reflecting the previous model. The AU tests significantly rejected the topology with the clustering of plant Type I TFs in the SRF clade (supplementary fig. S4, Supplementary Material online).

MADS Domains of SRF-type and MEF2-type TFs Form Distinct Structures

The functional units of SRF- and MEF2-type (extended MADS domain) are known to form distinct protein structures (fig. 2*a*). We applied AlphaFold2 (Jumper et al. 2021) to predict the structures of identified eukaryotic MADS-box TFs and compared their overlay patterns with the resolved SRF- or MEF2-type MADS domain structures (fig. 2*b* and *c*; supplementary table S2, Supplementary Material online). For the structures of TFs that belong to the MEF2-type inferred by phylogeny, the human SRF (HsSRF) structure was frequently not considered most similar, but rather the human MEF2A (HsMEF2A) structure. In contrast, the sequences present in the SRF clade fitted better to the HsSRF structure than the HsMEF2A structure (fig. 2*b* and *c*; supplementary fig. S5, Supplementary Material online). We also applied OmegaFold (Wu et al. 2022) to predict the structures of a larger set of MADS domain sequences (supplementary table S3, Supplementary Material online). Different from AlphaFold2 which makes use of MSAs for prediction, OmegaFold is able to predict the structures based on their individual primary sequences, without an explicit alignment. The results of structure similarities obtained from OmegaFold predictions agree with the comparisons based on the AlphaFold2 predictions (fig. 2*e* and *f*; supplementary fig. S5, Supplementary Material online).

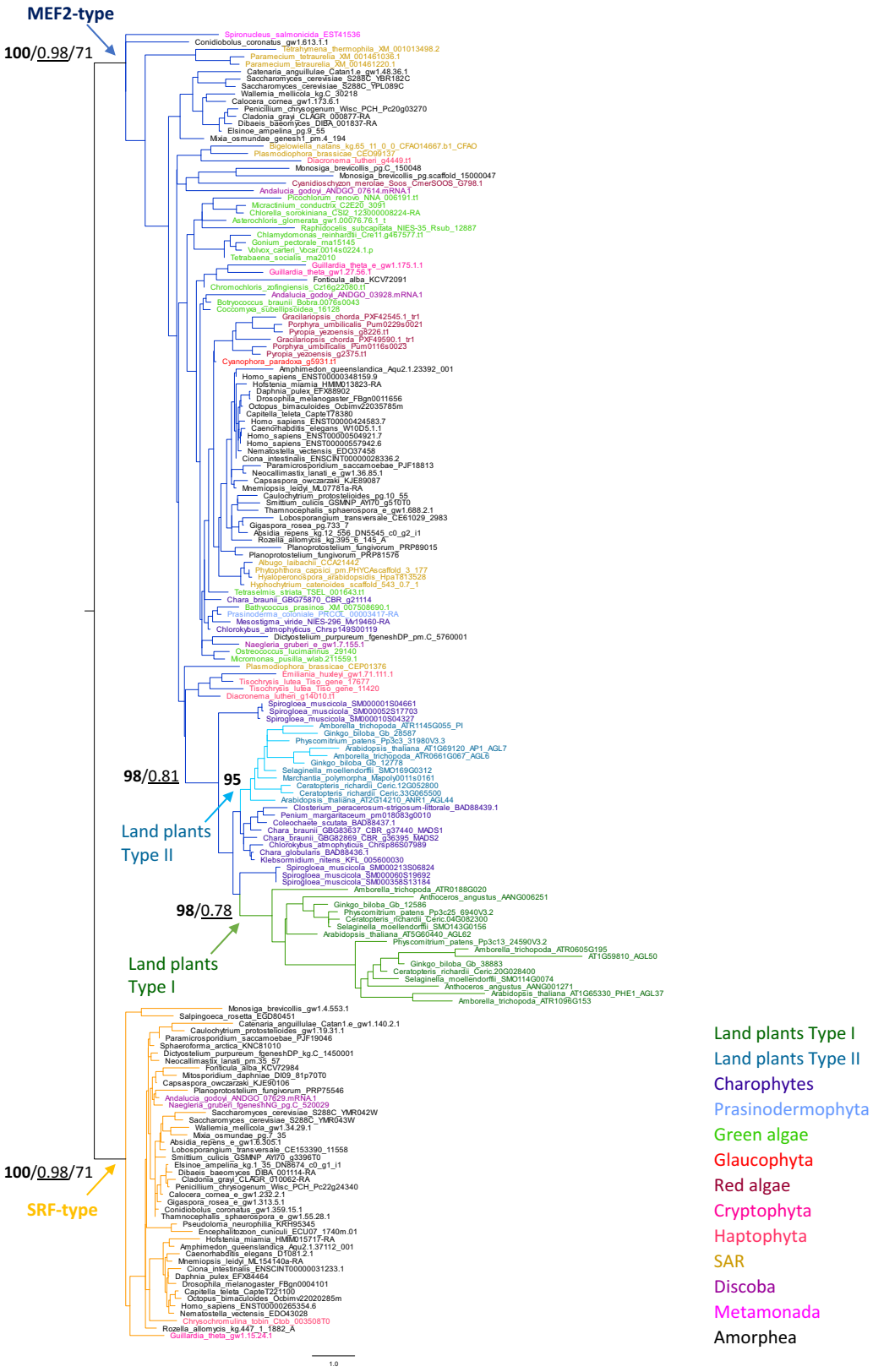

Fig. 3. Maximum-likelihood (ML) tree of selected MADS domains (extended definition, the first and second alpha helix and the connecting beta strands). Numbers for given branches of interest are support values: bootstrap values in ML trees (bold)/posterior probability in Bayesian inference (underlined)/bootstrap values in neighbor-joining trees. Branch color: yellow, SRF-type; blue, MEF2-type; green, Type I; cyan, Type II. Sequence IDs colored by category as listed on the right.

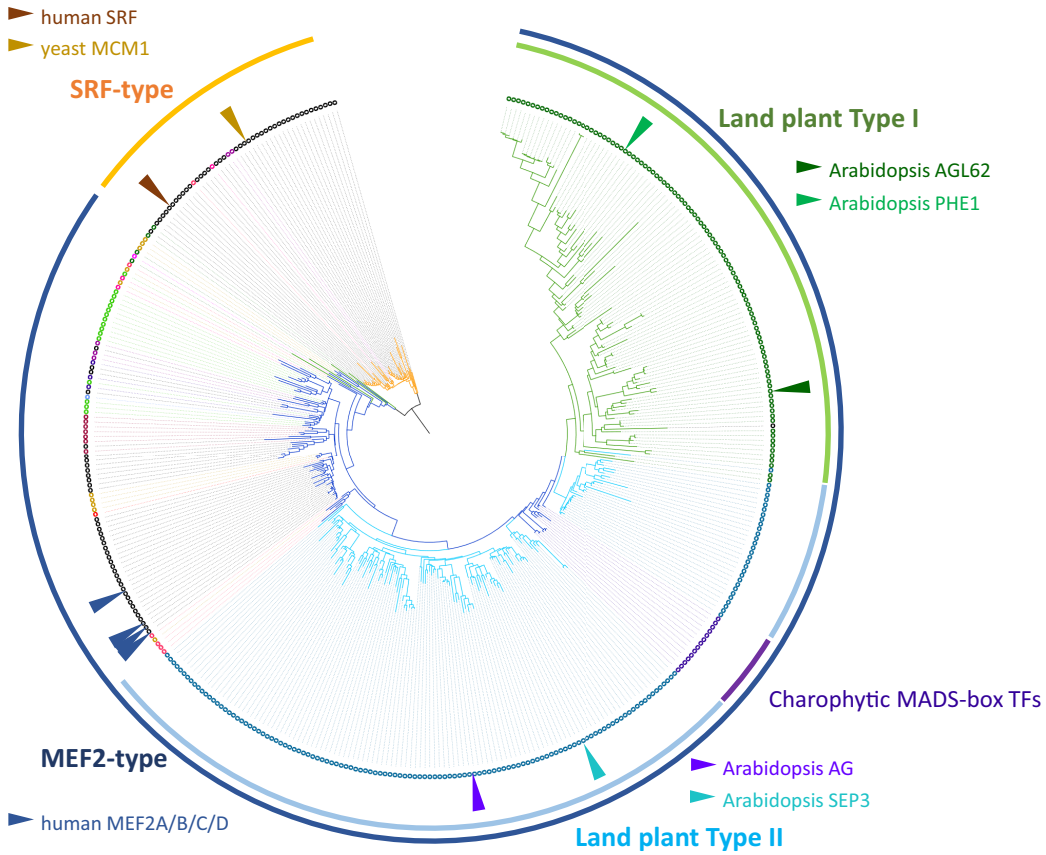

Fig. 4 Maximum-likelihood tree of extended MADS domain sequences including all MADS-box TFs from eight representative land plants. Branches and clades are colored by category as labels aside. Sequence IDs and taxon color codes are the same as the expanded tree in supplementary figure S2, Supplementary Material online. Arrows point to several TFs with known function as the landmarks for each type.

Hence, the predicted structure of the MADS domain supports the classification of proteins into SRF- or MEF2-type.

Based on AlphaFold2 and OmegaFold predictions, Type II TFs in land plants resemble the HsMEF2A structure (fig. 2*d* and *g*). Similarly, Type I TFs in land plants all mapped better to the HsMEF2A compared to the HsSRF structure. The second helix of Type I TFs was not predicted to be twisted by a kink, as found in SRF-type TFs. There is no experimentally resolved structure for a plant Type I MADS-box TFs available thus far. The only resolved crystal structure of a plant MADS-box protein is that of the Arabidopsis Type II TF SEPALLATA 3 (AtSEP3), which, as expected, displays a MEF2 structure (Lai et al. 2021). The predicted models for both Type I and II TFs in land plants, did align well with the AtSEP3 structure (fig. 2*d* and *g*), though Type II TFs had higher structural similarity scores to AtSEP3, consistent with the fact that they are more closely related. In Type II TFs, the I domain forms a helix, resembling the second helix formed by the MEF2 domain in the Amorphea MEF2-type TFs. Correspondingly, the second helix in the predicted structures of the Type I TFs is formed by an I-like domain region (Lai et al. 2021). While initially not defined in early studies (Alvarez-Buylla et al. 2000; Nam et al. 2004), the I-like domain in the Arabidopsis Type I TFs has been shown to be required for both dimerization and DNA-binding, functionally equivalent to the I domain in Type II TFs (Lai et al. 2021).

Loss of SRF-type Genes in the Most Recent Common Ancestor of Archaeplastida

The origin and divergence of Type I and II genes in land plants can be further inferred from the sister green lineages (figs. 3 and 4; supplementary figs. S1–S3, Supplementary Material online). In line with previous findings, the presence of only MEF2-type genes and the absence of SRF-type were observed in the genomes of streptophytic algae, a series of successive sister groups of land plants. Specifically, no SRF-type gene is present in three genomes of Zygnematophyceae (supplementary table S1, Supplementary Material online), the closest sister algal group of land plants (Cheng et al. 2019; Jiao et al. 2020). These findings suggest that the most recent common ancestor (MRCA) of land plants, as well as the MRCA of Streptophyta did not have SRFs. The loss of SRF-type genes may be tracked back to before the diversification of the whole Archaeplastida clade. In green algae and a third lineage of green plants, Prasinodermophyta, represented by *Prasinoderma coloniale*, there is no confidently predicted SRF-type gene either. In a few species belonging to the core Chlorophyta, we found some genes harboring an

open reading frame coding for a partial SRF-like MADS domain, but these SRF-like domains are quite divergent from SRF-type sequences found in other eukaryotic lineages, indicated by long branches, low support scores, and inconsistent positions in our phylogenetic analyses (supplementary fig. S6, Supplementary Material online). Their best BLASTP hits against all other MADS-box sequences were fungal SRF-type sequences, indicating that they probably arose from a horizontally transferred fragment in the MRCA of core chlorophytes. Likewise, in red algae, which are sister to the green plants, no SRF-type gene was identified, neither in the glaucophyte *Cyanophora paradoxa*. The lack of SRF-type genes in the Archaeplastida lineage is unlikely a consequence of multiple independent losses. Instead, it raises the more parsimonious hypothesis that the MRCA of the Archaeplastida lineage did not inherit an SRF-type gene, further supporting that only MEF2-type genes gave birth to Type I and II MADS-box genes in land plants.

In agreement with this evolutionary scenario, the predicted structures of Archaeplastida MADS-box TFs with or without a K domain all have higher similarity scores to the HsMEF2A and AtSEP3 structures than to the HsSRF structure (fig. 2*a*; supplementary fig. S5, Supplementary Material online). All these MEF2-type structures share the Intervening or MEF2 domain-like region, which constitutes the second helix with no kink. Thus, the predicted protein structures mirror the new phylogeny (figs. 1*b* and 2*a*).

Sporadic Losses of MADS-box TFs Across Eukaryotes

Except for Archaeplastida, some other eukaryotic lineages have only either SRF- or MEF2-type genes (supplementary table S1, Supplementary Material online). For example, three species in Microsporidia, a group of unicellular parasites closely related to fungi and *Sphaeroforma arctica* and *Thecamonas trahens*, belonging to successive sister groups of animals and fungi, they all have only SRF-type genes. In Haptista, *Chrysochromulina tobin* has only SRF-type genes, while three other species have only MEF2-type genes, suggesting reciprocal losses after the divergence from the MRCA comprising both types. Some SAR species, like ciliates, oomycetes and the cercozoan *Plasmodiophora brassicae* have only MEF2-type genes. Some surveyed species belonging to the green algae, the brown algae, diatoms, dinoflagellates and several Metamonada and Discoba protists among others have no extant MADS-box genes. To rule out the possibility that the observed absence of a certain type of MADS-box genes is a result of incomplete gene annotations, we scanned these genomes with the profile hidden Markov model for known MADS domains (PF00319) from Pfam (Paysan-Lafosse et al. 2023). We identified some unannotated and incomplete MADS domains in a few species (supplementary table S1, Supplementary Material online). Importantly however, no SRF-type gene was detected in the Archaeplastida. Therefore, the proposed SRF-type gene loss in the MRCA of Archaeplastida is not an artifact due to incomplete gene prediction.

Discussion

A New Hypothesis on the Origin of Type I and II MADS-box Genes in Land Plants

Supported by an updated phylogeny and predicted protein structures, we propose that in the land plant lineage Type I MADS-box TFs arose as a second clade of MEF2-type TFs, sister to the Type II TFs. The birth of Type I and II MADS-box genes was the result of a gene duplication event of a MEF2-type ancestral gene in the MRCA of land plants, which was followed by rounds of gene duplication largely expanding the MADS-box gene family. This new model of the MEF2-type origin is also favored by the principle of parsimony considering the absence of SRF-type genes in the sister lineages of Archaeplastida, specifically those of streptophytic algae (fig. 1*b*).

The major difference between our model and the previously proposed model is the origin of plant Type I MADS-box genes. Type I genes are known to have high substitution rates (Nam et al. 2004), which is reflected by long branches in previous phylogenetic analyses and ours as well (figs. 3 and 4). Previous studies claiming that Type I genes in Arabidopsis and rice are related to fungal and animal SRF-type genes relied on few available sequences (Alvarez-Buylla et al. 2000; Nam et al. 2004). Such limitation in sequence availability also affected another early model, which suggested that Type I genes (referred to as M-type in that study) are polyphyletic, while Type II genes might be MEF2-like genes (Kofuji et al. 2003). The upsurge of eukaryotic genomes has been filling the gaps between distantly related animal and plant sequences, making it possible to break the long branches and improve the phylogenetic resolution of the MADS-box gene family. Our new investigation covering diverse, previously underrepresented protist groups provides comprehensive support for the hypothesized ancient duplication of SRF- and MEF2-types before the divergence of extant eukaryotes. Besides, the Charophytic genomes serve as great references for the gene family evolution in land plants. We thus took the opportunity to revisit the evolution of the MADS-box gene family that was likely a key driver for plants adapting to land ecosystems.

The identification of Type I TFs in land plants as members of the MEF2-type lineage suggests that plant Type I and II should both be considered as "Type II" as defined by Alvarez-Buylla et al. (2000). Thereby, to avoid future confusion in nomenclature, we propose that referring to "Type I" and "Type II" should be restricted to MADS-box TFs in land plants, considering that these terms have been widely used in the literature in plant sciences. To differentiate the two clades of MADS-box TFs that originated from the ancient eukaryote-wide duplication, we suggest referring to them as "SRF-type" and "MEF2-type", respectively, as "SRF" and "MEF2" have been commonly used even before the proposed categories by Alvarez-Buylla et al. (2000). Additionally, we recommend to use "M-type" and "MIKC-type" only when distinguishing the MADS-box TFs in plants by domain architecture rather than their origin.

Evolution of I and K Domains

In contrast to the previous proposition that the I domain in plant Type II (MIKC-type) TFs and the MEF2 domains in animals were independently acquired, we propose that these domains constituting the second helix in all MEF2-type TFs have a common origin. Originally, the MADS domain was defined only to include the first helix and the antiparallel strands. However, we suggest an extended definition to include the second helix, since the helix–strand–helix structure functions as one unit that probably evolved together (Riechmann et al. 1996; Lai et al. 2021). Thus, the recently recognized I-like region in the Type I TFs (Lai et al. 2021) and the second helix in all other MEF2-type TFs are likely homologous, since they are structurally and functionally conserved. Meanwhile, the SAM domain in SRF-type TFs has gradually diverged from the precursor of the MEF2 domain. The turnover of the helix orientation was likely a key event establishing the two subclades, because SRF-type TFs do not heterodimerize with MEF2-type (Shore and Sharrocks 1995).

Admittedly, the predicted MEF2-type structure alone does not completely reject the previous hypothesis of an SRF-type origin of plant Type I TFs; it is possible that the hypothetically SAM-derived domain of plant Type I TFs changed the helix orientation convergently like in Type II TFs. Nevertheless, our phylogeny of the MADS domain, independent from the second helix, also suggests the MEF2-type origin of plant Type I TFs (supplementary figs. S3 and S4, Supplementary Material online), which is congruent with the most parsimonious evolution of their structures.

In terms of the Keratin-like domain, we confirmed the proposed streptophytic origin of the MIKC-type (Kaufmann et al. 2005; Thangavel and Nayar 2018), since K domains have only been identified in streptophytic MADS-box TFs (fig. 1*b*). Hence, the MEF2-type TF in the MRCA of Archaeplastida had no K domain and is thus an M-type. An ancestral M-type TF in the MRCA of Streptophyta acquired the K domain and continued evolving as a plant-specific MIKC-type. Subsequently, in the MRCA of land plants, a gene encoding MIKC-type TF duplicated into the paralogous precursors of Type I and II genes. The Type I TF precursor lost the K domain, leading to the extant Type I TFs as derived M-type (fig. 1*b*).

The loss of K domain in the Type I TFs partially explains the exon number differences between Type I and II genes (Kofuji et al. 2003; Parenicová et al. 2003). Type II genes, and most of Charophytic MADS-box genes, usually have multiple exons, with the first one coding for the MADS domain, the second one coding for the I domain and the next three to six exons coding for the K domain (Parenicová et al. 2003; Nishiyama et al. 2018; Rümpler et al. 2023). Type I genes, in contrast, typically have only one or two exons, where the MADS and I-like domains are encoded by a single exon (Parenicová et al. 2003). The lack of introns between the MADS domain and I-like domain coding regions suggests that the Type I precursor formed through spliced mRNA and cDNA intermediates by retroposition, as previously proposed (Kofuji et al. 2003).

From Stress Response and Reproductive Induction to Complex Structural Programming

Our study also confirmed the sharp contrast between MADS-box gene family size in land plants compared to that in other eukaryotes (Theissen et al. 1996; Thangavel and Nayar 2018). Most eukaryotes have only a few MADS-box genes (supplementary table S1, Supplementary Material online), revealing that the low-copy status remained constant during the evolution of protist-like stages, including early Archaeplastida. However, following the inferred duplication of Type I and II MADS-box TFs coupled with the terrestrialization of plants, the MADS-box gene family largely expanded, which provided the raw genetic material for subsequent functional differentiation. There have been extensive studies showing that MADS-box genes are key regulators of plant organ formation (Smaczniak et al. 2012; Theissen et al. 2016; Thangavel and Nayar 2018), similar to homeobox genes in animals (Nam et al. 2003; Lynch and Wagner 2008). The expansion of the MADS-box gene family has been proposed to be linked to the increasing complexity of extant land plants (Gramzow et al. 2014; Thangavel and Nayar 2018). Convergently and in concert with the evolution of multicellularity, while less abundant in copy number, SRF and MEF2 genes in metazoan animals are both functioning in embryo patterning and continue to regulate muscle development after maturity (Potthoff and Olson 2007). Nevertheless, since multicellularity evolved independently in animals and plants, the missing link for inferring the ancestral functional role of MADS-box genes and understanding their functional evolution lies in the unicellular, or under-differentiated multicellular eukaryotes.

Both SRF- and MEF2-type genes in unicellular and multicellular fungi, amoebae, and oomycetes have been shown to function in various stress responses (Messenguy and Dubois 2003; Galardi-Castilla et al. 2013; Rocha et al. 2016; Leesutthiphonchai and Judelson 2018; Wang et al. 2018; Ding et al. 2020). Thus, the regulation of stress-responsive programs is possibly the ancestral function of MADS-box genes, which has been maintained in multicellular metazoans, both invertebrates and vertebrates (Potthoff and Olson 2007; van der Linden et al. 2007; Blanchard et al. 2010; Vrailas-Mortimer et al. 2011). The stress-responsive rather than housekeeping function of ancestral MADS-box genes could explain the observed gene loss in several extant lineages. Originated from an ancestral stress-responsive TF, SRF- and MEF2-type TFs initially had presumably redundant functions upon duplication. Thus, the loss of SRF-type could have been compensated by MEF2-type TFs, which is likely the case in the unicellular ancestor of Archaeplastida. Supporting this assumption, the only MADS-box TF studied in microalgae, *Coccomyxa subellipsoidea* CsubMADS1, acts as a key regulator of stress tolerance (Nayar and Thangavel 2021). The colonization of the terrestrial habitat

most likely required an expansion of the genetic regulators responsive to the environment. Consistently, many MADS-box TFs have known function in regulating the response to stress, like FLOWERING LOCUS C, ARABIDOPSIS NITRATE REGULATED 1 (ANR1, AGL44), or AGL21 (Castelán-Muñoz et al. 2019).

In many unicellular organisms, the onset of reproduction is frequently induced by environmental stress, which may have facilitated the recruitment of MADS-box genes into the reproductive program (Escalante et al. 2003; Galardi-Castilla et al. 2013; Piccirillo et al. 2015; Leesutthiphonchai and Judelson 2018). In the land plant lineage, the evolution of spores and seeds that allow to withstand adverse environmental conditions may have been made possible by coupling the MADS-box TF regulation of stress resistance to reproductive development. This is seen for example in flowering plants, where MADS-box genes regulate floral patterning, but also the onset of flowering in responses to environmental cues (Castelán-Muñoz et al. 2019).

MADS-box TFs supposedly evolve functionally by rewiring gene regulatory networks. The DNA-binding sites recognized by SRF- and MEF2-type TFs have been extensively characterized as variants of CArG-box (Shore and Sharrocks 1995; Wu et al. 2011), which are also conserved among investigated plant MADS-box TFs (Aerts et al. 2018). Arabidopsis PHERES1 is currently the only plant Type I TF with known genome-wide DNA-binding sites in vivo, and the binding motifs are the same as that of Type II (Batista et al. 2019). Without dramatic sequence innovation in the DNA-binding sites, however, a given MADS-box TF achieves various collections of targets by different dimerization options (Lai et al. 2021; van Mourik et al. 2023). The combination of heterodimers got largely magnified in land plants, and specifically Type I TFs no longer homodimerize (de Folter et al. 2005). Moreover, the varying sequences in the C-terminus of Type I and II TFs further increased the diversity of protein–protein interaction and thus the potential to form regulatory complexes. Multiple rounds of duplication and diversification of Type I and II TFs likely have promoted the transition from a gametophyte-dominant to a sporophyte-dominant life cycle by equipping the sporophytic phase with developmental innovations such as flowers, fruits, and seeds.

Interestingly, while animal MEF2 subfamily did not expand as dramatically as the land plant orthologs, large numbers of splice variants have also increased the diversity of MEF2 TFs (Martin et al. 1994; Theissen et al. 1996). Animal MEF2 TFs are expressed predominantly within the early mesoderm (Potthoff and Olson 2007), which further differentiates into muscles, vascular, and neuronal tissues, so that its function greatly promotes the mobility, integrity and sensibility of metazoans. Convergently, the plant MEF2 genes got recruited into body patterning and reproduction in response to environmental stimuli. Thus, during the evolution of multicellularity in both animals and plants, MEF2-type TFs contributed to the formation of increasingly complex body plans.

In summary, we conclude that the duplication of an ancestral MEF2-type TF with a similar domain architecture to the extant MIKC-type TFs occurred in the MRCA of land plants. This event gave origin to the Type I and II precursors, which subsequently underwent gradual expansion, resulting in a vast family. This expansion played a crucial role in enabling the evolution of intricate structures tailored for the challenges of the terrestrial environment. In particular, importantly, among these structures relying on Type I and II MADS-box gene function are flowers and endospermic seeds, the two most prominent novelties of angiosperms. Gaining insight into the original functions of MIKC-type transcription factors and unraveling the impacts of the recent duplications and divergence of these transcription factors on driving developmental innovations across various plant lineages holds the promise of captivating future research pursuits.

Methods

Sequence and Phylogenetic Analyses

To search for MADS-box proteins in the investigated species (supplementary table S1, Supplementary Material online), amino acid sequences of MADS-box proteins of Arabidopsis (retrieved from TAIR10, https://www.arabidopsis.org/), human (retrieved from Ensembl, http://www.ensembl.org/Homo_sapiens/Info/Index), and yeast (retrieved from Mycocosm, https://mycocosm.jgi.doe.gov/Sacce1/Sacce1.home.html) were used as queries in BLASTP program runs. We also included three charophytic MADS-box sequences reported by Tanabe et al. (2005). The output sequences were aligned to the MADS domain entries in the Conserved Domain Database (Lu et al. 2020) by the conserved domain search tool, CD-Search (Marchler-Bauer and Bryant 2004), which guided the extraction of MADS domains in each species. To inspect any missed MADS-box genes, HMM searches were carried out with HMMER (Eddy 2011). Genomes of interest were scanned against the MADS-box TF associated profile hidden Markov model (PF00319) retrieved from Pfam (now hosted by InterPro, http://www.ebi.ac.uk/interpro/) (Paysan-Lafosse et al. 2023).

MUSCLE was used to generate the amino acid alignments of MADS domains extracted from the selected sequences with default settings (Edgar 2004). We first analyzed all MADS-box TFs from a series of representative land plants: Arabidopsis (angiosperm); *Cycas panzhihuaensis*, *Ginkgo biloba*, and *Gnetum luofuense* (gymnosperms); *Ceratopteris richardii* (fern), *Physcomitrium patens*, *Marchantia polymorpha*, and *Anthoceros angustus* (bryophytes) (supplementary fig. S2, Supplementary Material online). This enormous data set was dominated by recently duplicated and possibly redundant plant sequences. Therefore, we specifically generated a downsized data set for which the sample sizes from animals and plants were comparable, and the selected sequences of plant Type I and II TFs represented all major lineages of land plants and different well-established subfamilies. We prepared

two sets of alignments for the subsequent phylogenetic analyses: the alignments of only the conventional MADS-box region corresponding the first helix and the antiparallel strands; the alignments of the extended MADS domain definition to include the second helix additionally. We also applied two different alignment tools, T-coffee (Notredame et al. 2000), and MAFFT (Katoh et al. 2019), which returned similar alignments (supplementary fig. S7, Supplementary Material online).

IQ-TREE 1.6.7 was applied to perform phylogenetic analyses for maximum-likelihood trees (Nguyen et al. 2015). The implemented ModelFinder determined LG amino acid replacement matrix (Le and Gascuel 2008) to be the best substitution model in the tree inference (Kalyaanamoorthy et al. 2017). One thousand replicates of ultrafast bootstraps were applied to estimate the support for reconstructed branches (Hoang et al. 2018). We also selected the second-best suggestions of substitution models, JTT (Jones et al. 1992) and WAG (Whelan and Goldman 2001), in additional ML analyses (supplementary fig. S8, Supplementary Material online). Bayesian inference was carried out by Phylobayes (v3.2) under the CAT + GTR model with two chains. A consensus tree was built after the two chains were converged with the maxdiff less than 0.3 and the effective sample sizes of different parameters larger than 100 (Lartillot et al. 2009). MEGA11 (Tamura et al. 2021) was applied to generate neighbor-joining trees with *P*-distance (proportion of different amino acids), gamma distribution allowed for rate among sites and gaps treated by pairwise deletion. One thousand bootstrap replicates were generated and majority rule defines the consensus tree.

All these alternative approaches generated phylogenetic trees in agreement with each other, reflecting the robustness of our new hypothesis. We further compared the topology of constraint phylogenetic trees fitting the previous and the new hypotheses, by several tree topology tests such as AU tests (Shimodaira 2002) supported in IQ-TREE (Nguyen et al. 2015).

Protein Structure Prediction and Analyses

We predicted the protein structures of selected MADS-box TFs by the web-based service ColabFold (https://colab.esearch.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) (Mirdita et al. 2022). The top-ranked models were compared to resolved MADS-box protein structures, HsMEF2A (1EGW), HsSRF (1HBX), and AtSEP3 (7NB0), chains A respectively, downloaded from RCSB Protein Data Bank (https://www.rcsb.org/). The program "maxcluster" (http://www.sbg.bio.ic.ac.uk/~maxcluster/index.html) was used to perform structural comparisons based on computed TM-scores (Zhang and Skolnick 2005). Since AlphaFold2 prediction relies on MSAs, we left out nearly identical sequences which may result in the same MSAs and produce pseudoreplication. We also applied OmegaFold (https://github.com/HeliXonProtein/OmegaFold) (Wu et al. 2022) to predict the structures of all investigated MADS-box TFs based solely on their primary sequences.

Supplementary material

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

Acknowledgments

We thank Dr Elisabeth Hehenberger for her advice on the current taxonomy of eukaryotes. This research was funded by a grant from the Knut and Alice Wallenberg Foundation (2018-0206) to C.K. and the Max Planck Society.

Data Availability

All data are incorporated into the article and its online supplementary material.

References

Aerts N, de Bruijn S, van Mourik H, Angenent GC, van Dijk ADJ. 2018. Comparative analysis of binding patterns of MADS-domain proteins in *Arabidopsis thaliana*. *BMC Plant Biol*. **18**:131.

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martínez-Castilla L, Yanofsky MF. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci U S A*. **97**(10):5328–5333.

Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, Köhler C. 2019. The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife*. **8**:e50541.

Bemer M, Heijmans K, Airoldi C, Davies B, Angenent GC. 2010. An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiol*. **154**(1):287–300.

Blanchard FJ, Collins B, Cyran SA, Hancock DH, Taylor MV, Blau J. 2010. The transcription factor Mef2 is required for normal circadian behavior in Drosophila. *J Neurosci*. **30**(17):5855–5865.

Burki F, Roger AJ, Brown MW, Simpson A. 2020. The new tree of eukaryotes. *Trends Ecol Evol*. **35**(1):43–55.

Castelán-Muñoz N, Herrera J, Cajero-Sánchez W, Arrizubieta M, Trejo C, García-Ponce B, Sánchez MP, Álvarez-Buylla ER, Garay-Arroyo A. 2019. MADS-Box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants. *Front Plant Sci*. **10**:853.

Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, Sun W, Li X, Xu Y, Zhang Y, *et al.* 2019. Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell*. **179**(5):1057–1067.

De Bodt S, Raes J, Van de Peer Y, Theissen G. 2003. And then there were many: MADS goes genomic. *Trends Plant Sci*. **8**(10):475–483.

de Folter S, Immink RG, Kieffer M, Parenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, *et al.* 2005. Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell*. **17**:1424–1433.

Ding Z, Xu T, Zhu W, Li L, Fu Q. 2020. A MADS-box transcription factor FoRlm1 regulates aerial hyphal growth, oxidative stress, cell wall biosynthesis and virulence in *Fusarium oxysporum* f. sp. cubense. *Fungal Biol*. **124**(3–4):183–193.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. **7**(10):e1002195.

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

Escalante R, Moreno N, Sastre L. 2003. *Dictyostelium discoideum* developmentally regulated genes whose expression is dependent

on MADS box transcription factor SrfA. *Eukaryot Cell*. **2**(6): 1327–1335.
Galardi-Castilla M, Fernandez-Aguado I, Suarez T, Sastre L. 2013. Mef2A, a homologue of animal Mef2 transcription factors, regulates cell differentiation in *Dictyostelium discoideum*. *BMC Dev Biol*. **13**:12.
Gramzow L, Ritz MS, Theissen G. 2010. On the origin of MADS-domain transcription factors. *Trends Genet*. **26**(4): 149–153.
Gramzow L, Theissen G. 2013. Phylogenomics of MADS-Box genes in plants - two opposing life styles in one gene family. *Biology (Basel)*. **2**(3):1150–1164.
Gramzow L, Weilandt L, Theissen G. 2014. MADS Goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann Bot*. **114**(7):1407–1429.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. **35**(2):518–522.
Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N, *et al.* 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat Commun*. **5**:3978.
Jiao C, Sørensen I, Sun X, Sun H, Behar H, Alseekh S, Philippe G, Palacio Lopez K, Sun L, Reed R, *et al.* 2020. The *Penium margaritaceum* genome: hallmarks of the origins of land plants. *Cell*. **181**(5):1097–1111.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. **8**:275–282.
Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.* 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**:583–589.
Kalyaanamoorthy S, Minh BQ, Wong T, von Haeseler A, Jermiin LS. 2017. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. **14**(6):587–589.
Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN. 2008. The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell*. **20**(3): 635–647.
Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT Online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. **20**:1160–1166.
Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene*. **347**:183–198.
Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M. 2003. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol*. **20**(12): 1963–1977.
Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U. 2003. The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev*. **17**(12):1540–1553.
Lai X, Vega-Léon R, Hugouvieux V, Blanc-Mathieu R, van der Wal F, Lucas J, Silva CS, Jourdain A, Muino JM, Nanao MH, *et al.* 2021. The intervening domain is required for DNA-binding and functional identity of plant MADS transcription factors. *Nat Commun*. **12**(1):4760.
Lartillot N, Lepage T, Blanquart S. 2009. Phylobayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. **25**(17):2286–2288.
Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. **25**(7):1307–1320.
Leesutthiphonchai W, Judelson HS. 2018. A MADS-box transcription factor regulates a central step in sporulation of the oomycete *Phytophthora infestans*. *Mol Microbiol*. **110**(4):562–575.
Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, *et al.* 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. **48**(D1):D265–D268.
Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in the evolution of development. *Evolution*. **62**: 2131–2154.
Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res*. **32**:W327–W331.
Martin JF, Miano JM, Hustad CM, Copeland NG, Jenkins NA, Olson EN. 1994. A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Mol Cell Biol*. **14**:1647–1656.
Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM. 2011. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*. **23**(3):865–872.
Messenguy F, Dubois E. 2003. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene*. **316**:1–21.
Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. Colabfold: making protein folding accessible to all. *Nat Methods*. **19**:679–682.
Nam J, dePamphilis CW, Ma H, Nei M. 2003. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol Biol Evol*. **20**(9):1435–1447.
Nam J, Kim J, Lee S, An G, Ma H, Nei M. 2004. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc Natl Acad Sci U S A*. **101**(7):1910–1915.
Nayar S, Thangavel G. 2021. CsubMADS1, a lag phase transcription factor, controls development of polar eukaryotic microalga *Coccomyxa subellipsoidea* C-169. *Plant J*. **107**(4):1228–1242.
Ng M, Yanofsky MF. 2001. Function and evolution of the plant MADS-box gene family. *Nat Rev Genet*. **2**(3):186–195.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. **32**(1):268–274.
Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, *et al.* 2018. The Chara genome: secondary complexity and implications for plant terrestrialization. *Cell*. **174**(2):448–464.e24.
Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. **302**:205–217.
Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, *et al.* 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*. **15**(7):1538–1551.
Paul P, Dhatt BK, Miller M, Folsom JJ, Wang Z, Krassovskaya I, Liu K, Sandhu J, Yu H, Zhang C, *et al.* 2020. MADS78 and MADS79 are essential regulators of early seed development in rice. *Plant Physiol*. **182**(2):933–948.
Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, *et al.* 2023. Interpro in 2022. *Nucleic Acids Res*. **51**(D1):D418–D427.
Piccirillo S, Morales R, White MG, Smith K, Kapros T, Honigberg SM. 2015. Cell differentiation and spatial organization in yeast colonies: role of cell-wall integrity pathway. *Genetics*. **201**(4): 1427–1438.
Potthoff M, Olson EN. 2007. MEF2: a central regulator of diverse developmental programs. *Development*. **134**(23):4131–4140.
Qiu Y, Köhler C. 2022. Endosperm evolution by duplicated and neofunctionalized type I MADS-box transcription factors. *Mol Biol Evol*. **39**(1):msab355.
Riechmann JL, Krizek BA, Meyerowitz EM. 1996. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc Natl Acad Sci U S A*. **93**:4793–4798.
Rocha MC, Fabri JH, Franco de Godoy K, Alves de Castro P, Hori JI, Ferreira da Cunha A, Arentshorst M, Ram AF, van den Hondel CA, Goldman GH, *et al.* 2016. Aspergillus fumigatus MADS-box

transcription factor rlmA is required for regulation of the cell wall integrity and virulence. *G3 (Bethesda, MD)*. **6**(9):2983–3002.
Rümpler F, Tessari C, Gramzow L, Gafert C, Blohs M, Theissen G. 2023. The origin of floral quartet formation-ancient exon duplications shaped the evolution of MIKC-type MADS-domain transcription factor interactions. *Mol Biol Evol*. **40**:msad088.
Schrödinger L, DeLano W. 2020. PyMOL. Available from: http://www.pymol.org/pymol
Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science*. **250**:931–936.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. **51**:492–508.
Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. *Eur J Biochem*. **229**:1–13.
Smaczniak C, Immink RG, Angenent GC, Kaufmann K. 2012. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development*. **139**(17):3081–3098.
Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. **38**(7):3022–3027.
Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Münster T, Theissen G, Nozaki H, Ito M. 2005. Characterization of MADS-box genes in Charophycean green algae and its implication for the evolution of MADS-box genes. *Proc Natl Acad Sci U S A*. **102**(7):2436–2441.
Thangavel G, Nayar S. 2018. A survey of MIKC type MADS-box genes in non-seed plants: algae, bryophytes, lycophytes and ferns. *Front Plant Sci*. **9**:510.
Theissen G, Kim JT, Saedler H. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J Mol Evol*. **43**(5):484–516.
Theissen G, Melzer R, Rümpler F. 2016. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development*. **143**(18):3259–3271.
van der Linden AM, Nolan KM, Sengupta P. 2007. KIN-29 SIK regulates chemoreceptor gene expression via an MEF2 transcription factor and a class II HDAC. *EMBO J*. **26**(2):358–370.
van Mourik H, Chen P, Smaczniak C, Boeren S, Kaufmann K, Bemer M, Angenent GC, Muino JM. 2023. Dual specificity and target gene selection by the MADS-domain protein FRUITFULL. *Nat Plants*. **9**:473–485.
Vrailas-Mortimer A, del Rivero T, Mukherjee S, Nag S, Gaitanidis A, Kadas D, Consoulas C, Duttaroy A, Sanyal S. 2011. A muscle-specific p38 MAPK/Mef2/MnSOD pathway regulates stress, motor function, and life span in Drosophila. *Dev Cell*. **21**(4):783–795.
Wang Q, Du M, Wang S, Liu L, Xiao L, Wang L, Li T, Zhuang H, Yang E. 2018. MADS-Box transcription factor MadsA regulates dimorphic transition, conidiation, and germination of *Talaromyces marneffei*. *Front Microbiol*. **9**:1781.
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. **18**:691–699.
Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S, Su C, Wu Z, Xie Q, Berger B, *et al*. 2022. High-resolution de novo structure prediction from primary sequence. bioRxiv .
Wu W, Huang X, Cheng J, Li Z, de Folter S, Huang Z, Jiang X, Pang H, Tao S. 2011. Conservation and evolution in and among SRF- and MEF2-type MADS domains and their binding sites. *Mol Biol Evol*. **28**:501–511.
Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. **33**:2302–2309.