

Letter

Ancestral duplication of MADS-box genes in land plants empowered the functional divergence between sporophytes and gametophytes

MADS-box transcription factors (TFs) have gained widespread recognition for their exceptional diversity and pivotal roles in various biological functions across eukaryotic organisms. Specifically, in land plants, the MADS-box gene family has undergone substantial expansion and provided the genetic raw material for many developmental novelties, including flowers, fruits and seeds. Therefore, understanding the origin of MADS-box genes is crucial for gaining insights into the evolutionary success of land plants. Land plant MADS-box TFs have been categorized into two groups. MIKC-type (Type II) TFs are named after the structural arrangement of the MADS (M), Intervening (I) and plant-specific Keratin-like (K) domain, followed by a variable C-terminal region. By contrast, M-type (Type I) TFs lack the K domain (Alvarez-Buylla *et al.*, 2000; Kofuji *et al.*, 2003; Nam *et al.*, 2004). The identification of mutants exhibiting distinct flower organ patterns led to the discovery of a multitude of MIKC-type MADS-box TFs in plants (Ng & Yanofsky, 2001; Nam *et al.*, 2003; Kaufmann *et al.*, 2005). By contrast, the identification of M-type genes occurred exclusively through bioinformatic analyses following the unveiling of the *Arabidopsis thaliana* genome (Alvarez-Buylla *et al.*, 2000). Within the MIKC-type of MADS-box TFs, the MIKC*-type is distinguished based on the different arrangements of exons that encode the K domain (Svensson *et al.*, 2000; Henschel *et al.*, 2002; Zobell *et al.*, 2010; Kwantes *et al.*, 2012; Rümpler *et al.*, 2023). Both the 'classic' MIKC^C- and the MIKC*-types are present across land plant lineages and form well-supported separate clades (Kwantes *et al.*, 2012; Gramzow & Theissen, 2013; Liu *et al.*, 2013), suggesting that they diverged before the diversification of land plant lineages (Henschel *et al.*, 2002; Kofuji *et al.*, 2003; Tanabe *et al.*, 2005; Kwantes *et al.*, 2012).

It was a long-standing prevailing view in the field that plant Type II genes are orthologs to myocyte enhancer factor-2 (MEF2) genes in animals. Those were presumed to have diverged from plant Type I genes and animal serum response factor (SRF) genes via an ancient duplication before the divergence of the extant eukaryotic lineages (Alvarez-Buylla *et al.*, 2000). Thus, with the shared presence of the K domain, MIKC*-type MADS-box genes were intuitively considered more closely related to MIKC^C-type genes, collectively called Type II. This adheres to the principle of parsimony, by which the acquisition of the K domain should ideally have occurred only

once during the evolution of plants (Kaufmann *et al.*, 2005; Thangavel & Nayar, 2018) (Fig. 1a). However, we recently showed that both Type I and Type II genes are land plant-specific MEF2 orthologs that arose from a duplication event predating the origin of land plants (Qiu *et al.*, 2023). Therefore, the plant Type I and Type II genes are more closely related than previously appreciated (Fig. 1a). Noticeably, the Type I and II gene duplication hypothetically took place around the time inferred for the rise of MIKC*-type genes (Kofuji *et al.*, 2003; Kwantes *et al.*, 2012). This raises the question whether the previously proposed monophyletic relationship between the MIKC*- and MIKC^C-type as the Type II clade is better supported than the alternative hypothesis proposing that MIKC*-type genes are close paralogs of the Type I (M-type) clade (Fig. 1a). Since the ancestral land plant MADS-box gene giving rise to all current subfamilies has been inferred to include the K domain (Kaufmann *et al.*, 2005; Thangavel & Nayar, 2018; Qiu *et al.*, 2023), this domain is not only shared by MIKC^C-type and MIKC*-type, but also inherited by the ancestral gene leading to the extant Type I clade. Therefore, MIKC^C- and MIKC*-types have just maintained the ancestral structures, which however does not provide sufficient evidence for their close relatedness. Regardless of whichever diverged first, MIKC*-, MIKC^C- or M-type, all scenarios equally assume an ancestral gain of the K domain and a subsequent specific loss in the M-type lineage (Fig. 1a).

To reappraise the origin of MIKC*-type TFs and to resolve the relatedness among MIKC*-, MIKC^C- and M-types, we performed phylogenetic analyses with broad sampling coverage across major lineages of land plants (Supporting Information Table S1). We aligned the MADS domain amino acid sequences of all MADS-box TFs from six bryophytes, four lycophytes, four ferns, five gymnosperms and three angiosperms (Figs 1b, S1; Table S2), with sequences from charophytes, the closest successive sister lineages of land plants, and green algae as outgroups. In the maximum-likelihood tree, the MIKC*-type TFs form a well-supported clade with Type I (M-type) TFs. We additionally generated a maximum-likelihood tree based on the codon alignments, which largely supports the new topology (Fig. S2). Since the MADS-box gene family has undergone several rounds of gene duplications in plants (Nam *et al.*, 2004), there are many redundant sequences in the above analysis, and especially the MIKC^C-type sequences are dominating the dataset. We therefore selected representative sequences of comparable sample sizes for the MIKC*-type, MIKC^C-type and Type I clades from divergent TF subfamilies in all those species. Phylogenetic trees generated with both maximum-likelihood and Bayesian inference show that MIKC*-type TFs constitute the nearest sister clade to the Type I (M-type) TFs, instead of MIKC^C-type (Figs S3, S4). These new phylogenetic analyses all suggest that in a recent streptophytic ancestor of land plants, an MIKC-type gene duplicated into the precursor of the MIKC^C clade and a precursor for the MIKC*/M-type clade. This

latter gene subsequently duplicated to give rise to the precursor of the MIKC* clade and the precursor of the M-type clade, which arose after the loss of the K-domain exons.

We noticed that not all the charophytic MADS-box sequences were positioned as successive sisters to all MADS-box sequences in land plants, but none of the charophytes we analyzed had a pair of MADS-box genes each clearly belonging to Type I or II clades. We thus refrain from concluding whether the charophytic lineages shared the abovementioned one or two rounds of MADS-box gene duplication. If the duplication took place in the early lineage of streptophytes, multiple independent gene losses must have occurred in the paraphyletic charophytes. We focused on the land plant MADS-box TFs and generated maximum-likelihood trees without sequences from charophytes (Fig. S5). With the sequences from chlorophytes as the outgroup, in land plants the MIKC* clade is still closer to the M-type clade rather than the MIKC^C clade.

We carried out approximate unbiased (AU) tests (Shimodaira, 2002) to compare the phylogenetic trees corresponding to competing hypotheses (Figs S6, S7). One topology represents our new phylogeny. The second one is a constraint phylogeny forcing the Type I clade to be the outgroup of the originally considered combined Type II clade of both MIKC^C- and MIKC*-type, reflecting the previous MIKC monophyly model. We also created a third constraint phylogeny resembling an alternative MIKC paraphyly hypothesis, assuming MIKC*-type diverged before the MIKC^C-type and Type I split. The AU tests suggest that the new phylogeny with monophyly of MIKC* and M-type is significantly better than the previous proposition of MIKC* and MIKC^C monophyly. Although the assumed third topology for MIKC^C and M-type monophyly is not completely rejected based on the topology tests, the new phylogeny is substantially better. Resolving the relationship between these clades was challenging, since the two inferred successive duplication events giving rise to the precursors of the three types likely took place in a very short time span in the history of streptophyte evolution, predating the split of bryophytes and vascular plants. This is reflected by low posterior probability supports in the Bayesian inferences (Fig. S4). Nevertheless, combining all available evidence from an up-to-date and comprehensive sampling and multiple analyses, we find the model that MIKC^C-type TFs diverged first and the M-type arose by loss of K-domain after the later divergence from the MIKC*-type with the highest probability.

The new phylogeny with the revised MIKC* origin aids to reconstruct the model depicting the functional evolution of MADS-box genes in land plants (Fig. 2). As suggested by several charophytic MADS-box genes (Tanabe *et al.*, 2005), and the sole green algal MADS-box gene that has been functionally characterized, CsubMADS1 in the haploid *Coccomyxa subellipsoidea* (Nayar & Thangavel, 2021), it is likely that ancestral MADS-box genes primarily served gametophytic functions, particularly in stress tolerance and gamete development. MIKC*-type genes likely have maintained the ancestral gametophytic function, consistent with expression and function of MIKC*-type genes in pollen, the male gametophyte (Verelst *et al.*, 2007a,b; Adamczyk & Fernandez, 2009; Liu *et al.*, 2013). Interestingly, the single-copy MIKC* ortholog in the liverwort *Marchantia polymorpha* MpMADS1 can

rescue mutants deficient of MIKC* in *Arabidopsis* (Zobell *et al.*, 2010), suggesting conservation of MIKC*-type function in male gametophytes across land plants. Supporting this hypothesis, MIKC*-type genes in mosses, lycophytes and ferns are preferentially associated with gametophytes, especially the structures bearing male gametophytes (Svensson *et al.*, 2000; Riese *et al.*, 2005; Zobell *et al.*, 2010; Kwantes *et al.*, 2012). In seed plants, MIKC*-type gene expression became nearly completely restricted to male gametophytes (Verelst *et al.*, 2007a,b; Adamczyk & Fernandez, 2009; Liu *et al.*, 2013; Gramzow *et al.*, 2014; Gu *et al.*, 2022). Similarly, M-type MADS-box TFs are functionally important in female gametophytes in seed plants, as well as in the endosperm in flowering plants (Bemer *et al.*, 2010; Masiero *et al.*, 2011; Qiu & Köhler, 2022), complementing the requirement of MADS-box function after MIKC*-type evolved to be male-specific. It has been proposed that the expansion of the MADS-box gene family contributed to the complexity of the plant body plan (Theissen *et al.*, 1996; Kaufmann *et al.*, 2005; Thangavel & Nayar, 2018). This is most pronounced for the MIKC^C clade, which is famous for its regulatory role in the patterning of reproductive organs (Smaczniak *et al.*, 2012; Theissen *et al.*, 2016). Nevertheless, the regulatory role of MIKC^C-type TFs in sporophytes is assumed to be derived, since ancestral MADS-box genes likely did not have a sporophytic function. This is inferred by charophytic MADS-box genes, which are not expressed in the zygote, the only diploid phase of charophytes (Tanabe *et al.*, 2005). In comparison with the MIKC^C-type subfamily, the copy number of MIKC*-type TFs remained moderately low (Kwantes *et al.*, 2012; Liu *et al.*, 2013). Likewise, the M-type TFs remained less duplicated and differentiated in vascular plants, until they evolved a function in the endosperm of angiosperms (Qiu & Köhler, 2022). Thus, number and differentiation patterns of MIKC* and M-type MADS-box genes align with the simplicity of gametophytes in vascular plants. Conversely, in the gametophyte-dominant bryophytes, the MIKC*-type subfamily largely expanded in several moss species, and the M-type subfamily underwent lineage-specific expansion in the *Anthoceros* hornworts, which may have contributed to the considerably more complex structures of mosses and hornworts compared with liverworts (Zobell *et al.*, 2010; Table S2).

Together, our new results suggest a reclassification that moves MIKC*-type TFs into the Type I clade (Fig. 2). Based on this revised phylogeny, the updated Type I TFs are a clade of MADS-box genes that primarily preserved their ancestral function in gametophyte development. During the evolution of seed plants, MIKC*-type and M-type adopted male- and female-specific functions, respectively. In parallel, the refined Type II clade, comprised of typical MIKC^C-type genes, gradually diverged to be the sporophytic MADS-box subfamily, which repeatedly duplicated and neofunctionalized to generate new genetic regulators underlying the diverse body architectures of sporophytes.

Materials and Methods

We extended the collection of genomes analyzed in Qiu *et al.* (2023) with more species that represent all major lineages of

land plants for MADS-box protein identification (Table S1). Amino acid sequences of MADS-box proteins in Arabidopsis (retrieved from TAIR10, <https://www.arabidopsis.org/>) were used

as queries in the BLASTP program to search for MADS-box proteins in these additional genomes. The MADS domains from identified MADS-box proteins were extracted based on the alignments to the

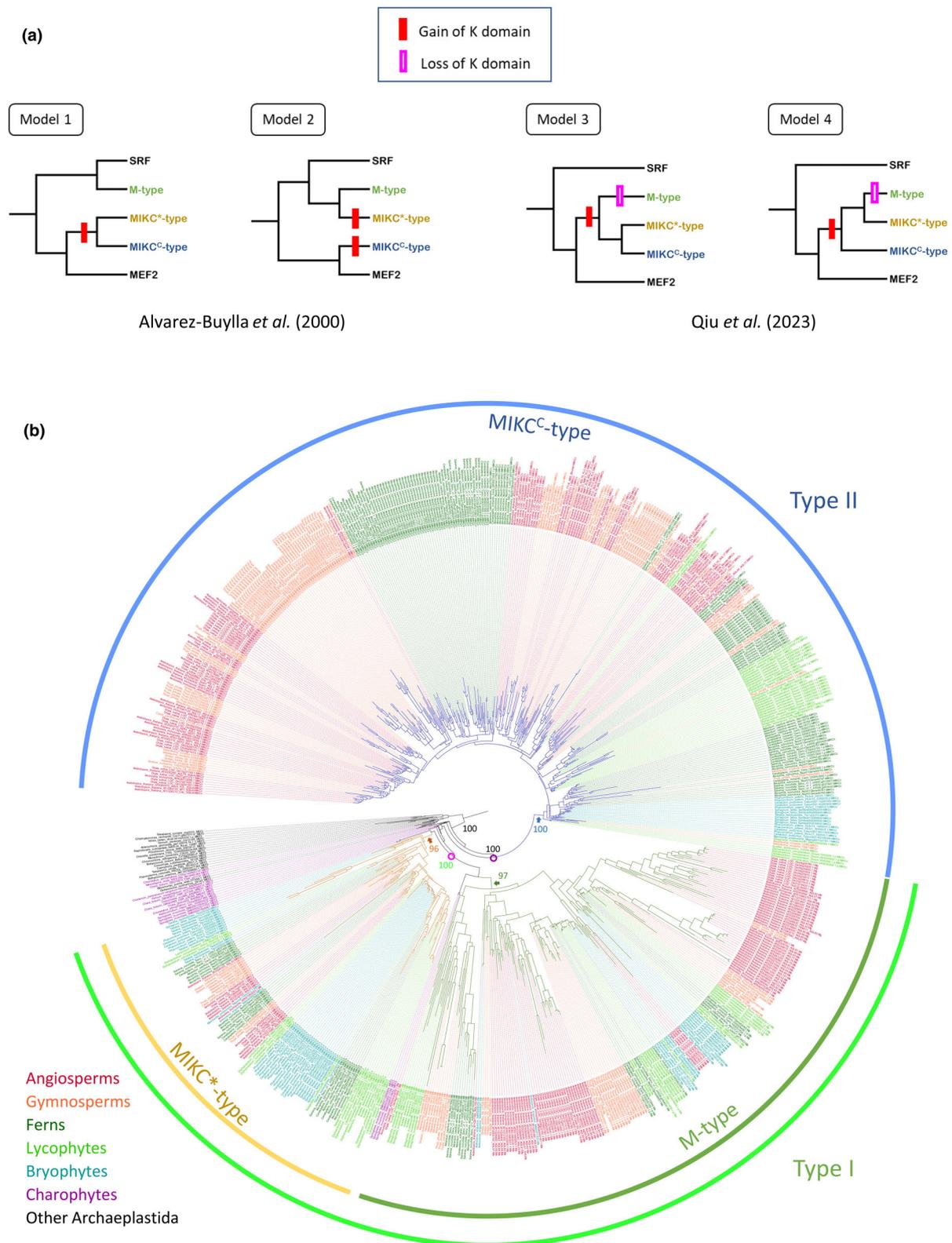


Fig. 1 Evolutionary models of MADS-box transcription factors in land plants. (a) Comparison between models depicting the evolution of MADS-box genes in land plants. A previous model (Alvarez-Buylla *et al.*, 2000) predicted that M-type genes are orthologous to serum response factor (SRF) genes in animals and distantly related to MIKCC-type genes, which are orthologous to MEF2-type in animals. Therefore, Model 1 hypothesizing MIKC*-type are closely related to MIKCC-type has been favored, because it assumes only a single shared acquisition of the K domain. The less parsimonious Model 2 needs to assume two independent acquisitions of the same K domain, if MIKC*-type is hypothesized to cluster with M-type instead. However, recent data (Qiu *et al.*, 2023) indicate that M-type, MIKCC-type and MIKC*-type were all derived from plant-specific MEF2-type gene duplications and the preduplicated ancestral gene had acquired the K domain. Therefore, Model 3, proposing a close relationship between MIKC*-type and MIKCC-type genes, by the principle of parsimony in evolution, is no longer more evident compared with the alternative Model 4, which suggests a closer relationship between MIKC*-type and M-type genes. Both Models 3 and 4 assume a single loss of the K domain in M-type genes after the ancestral gain. This study aimed to test which model explains the evolution of the three types of land plant MADS-box genes in light of the newly available information (Qiu *et al.*, 2023). (b) Maximum-likelihood tree of MADS domain sequences including all MADS-box TFs from representative land plants. Bootstrap support values for given branches of interest are labelled. Branches and clades are colored by category. The circles mark the ancestral gene duplication events and the arrows point to the branches of precursors giving rise to the land plant MIKCC-type, the MIKC*-type and the M-type clades. Sequence IDs and taxon color codes are the same as in the expanded tree in Supporting Information Fig. S1. Suffixes of sequence IDs denote the MIKCC-type (MIKCC), the MIKC*-type (MIKCs) or three subtypes, M α , M β and M γ , of the M-type (Ma, Mb and Mg, respectively).

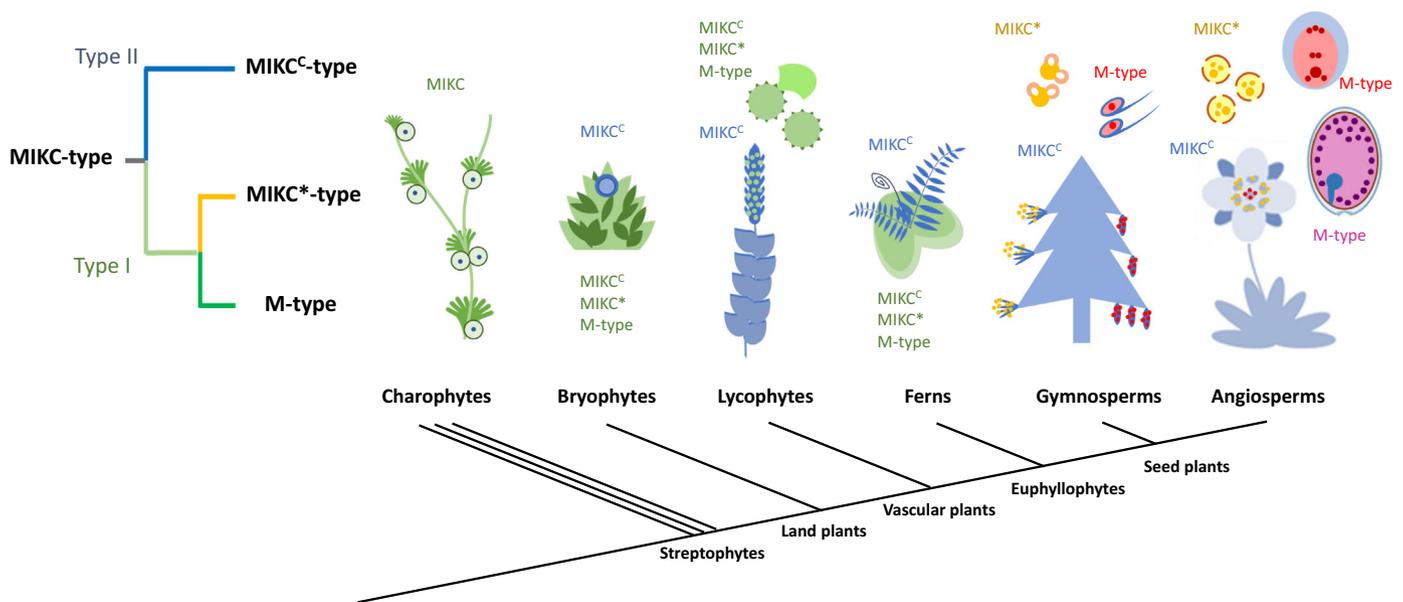


Fig. 2 Functional divergence of MADS-box transcription factors in land plants. Blue, sporophytes; green, gametophytes (nonseed plants); yellow, male gametophytes (seed plants); red, female gametophytes (seed plants); purple, endosperm (flowering plants).

MADS domain entries in the Conserved Domain Database (Lu *et al.*, 2020) by the conserved domain search tool, CD-Search (Marchler-Bauer & Bryant, 2004). We aligned the MADS domains with MUSCLE by the default settings (Edgar, 2004). We also used the amino acid alignment as a guide to generate the corresponding codon alignments (Notes S1). We first included all identified MADS-box TFs from a series of representative land plants: *Arabidopsis*, rice and *Amborella trichopoda* (angiosperm); *Thuja plicata*, *Cycas panzhibuaensis*, *Ginkgo biloba*, *Gnetum luofuense* and *Welwitschia mirabilis* (gymnosperms); *Ceratopteris richardii*, *Salvinia cucullate*, *Adiantum capillus-veneris* and *Alsophila spinulosa* (ferns); *Selaginella moellendorffii*, *Isoetes taiwanensis*, *Diphasiastrum complanatum* and *Lycopodium clavatum* (lycophytes); *Physcomitrium patens*, *Ceratodon purpureus*, *Sphagnum fallax*, *Takakia lepidozoioides*, *Marchantia polymorpha* and *Anthoceros angustus* (bryophytes) (Table S2). This enormous dataset

contains many more MIKCC-type sequences compared with MIKC*- and M-types. Therefore, we generated a downsized dataset with selected sequences in the three major MADS-box TF subfamilies across land plants. For the purpose of elucidating the deep branching pattern between the three major MADS-box clades, we specifically chose the genes with high confidence to be classified into a certain type as the representative sequences based on previous knowledge and the large maximum-likelihood tree generated in this study.

We applied IQ-TREE 2 to generate maximum-likelihood trees (Minh *et al.*, 2020). The implemented MODELFINDER (Kalyaanamoorthy *et al.*, 2017) determined the JTT substitution matrix (Jones *et al.*, 1992) to be the best substitution model in the tree inference for the large dataset of protein alignments and LG (Le & Gascuel, 2008) for the reduced dataset. For the nucleotide alignments, we allowed different evolutionary rates between partitions of the first, second

and third codon positions. We ran 1000 replicates of ultrafast bootstraps to estimate the support for reconstructed branches (Hoang *et al.*, 2018). We also employed PHYLOBAYES (v.3.2) to perform Bayesian inference under the CAT+GTR model with two chains. After ensuring that the two chains had converged with a maxdiff < 0.3, a consensus tree was created. The effective sample sizes of the different parameters were verified to be greater than 200, except for α parameter of the gamma distribution of rates across sites, which was 78 (Lartillot *et al.*, 2009). We further compared the topology of constraint phylogenetic trees fitting the competing hypotheses, by topology tests such as AU tests (Shimodaira, 2002) supported in IQ-TREE 2 (Minh *et al.*, 2020).

Acknowledgements

We thank Dr Jiaqi Wu for the helpful discussion. This research was funded by a grant from the Knut and Alice Wallenberg Foundation (2018-0206) to CK and the Max Planck Society. Open Access funding enabled and organized by Projekt DEAL.

Competing interests

None declared.

Author contributions

YQ, ZL and CK planned and designed the research. YQ and ZL analyzed data. YQ, ZL and CK wrote the manuscript. YQ and ZL contributed equally to this work.

ORCID

Claudia Köhler  <https://orcid.org/0000-0002-2619-4857>

Zhen Li  <https://orcid.org/0000-0001-8920-9270>

Yichun Qiu  <https://orcid.org/0000-0002-8145-4535>

Data availability

The data that support the findings of this study are openly available and resources are listed in Table S1 of this article.

Yichun Qiu^{1,*†} , Zhen Li^{2,3†}  and Claudia Köhler^{1,4,*} 

¹Department of Plant Reproductive Biology and Epigenetics, Max Planck Institute of Molecular Plant Physiology, Potsdam, 14476, Germany;

²Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, 9052, Belgium;

³VIB Center for Plant Systems Biology, VIB, Ghent, 9052, Belgium;

⁴Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Centre for Plant Biology, Uppsala, 75007, Sweden

(*Authors for correspondence: email: qiu@mpimp-golm.mpg.de (YQ); koehler@mpimp-golm.mpg.de (CK))

†These authors contributed equally to this work.

References

- Adamczyk BJ, Fernandez DE. 2009. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiology* 149: 1713–1723.
- 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. *Proceedings of the National Academy of Sciences, USA* 97: 5328–5333.
- Bemer M, Heijmans K, Airoidi C, Davies B, Angenent GC. 2010. An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiology* 154: 287–300.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
- Gramzow L, Theissen G. 2013. Phylogenomics of MADS-Box genes in plants – two opposing life styles in one gene family. *Biology* 2: 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. *Annals of Botany* 114: 1407–1429.
- Gu KJ, Lin CF, Wu JJ, Zhao YP. 2022. GINKGODB: an ecological genome database for the living fossil, Ginkgo biloba. *Database* 2022: baac046.
- Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theissen G. 2002. Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Molecular Biology and Evolution* 19: 801–814.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBOOT2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8: 275–282.
- Kalyaanamoorthy S, Minh BQ, Wong T, von Haeseler A, Jermin LS. 2017. MODELFINDER: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.
- 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. *Molecular Biology and Evolution* 20: 1963–1977.
- Kwantes M, Liebsch D, Verelst W. 2012. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. *Molecular Biology and Evolution* 29: 293–302.
- Lartillot N, Lepage T, Blanquart S. 2009. PHYLOBAYES 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25: 2286–2288.
- Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* 25: 1307–1320.
- Liu Y, Cui S, Wu F, Yan S, Lin X, Du X, Chong K, Schilling S, Theissen G, Meng Z. 2013. Functional conservation of MIKC*-Type MADS box genes in Arabidopsis and rice pollen maturation. *Plant Cell* 25: 1288–1303.
- 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 Research* 48(D1): D265–D268.
- Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Research* 32: W327–W331.
- 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: 865–872.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.
- Nam J, dePamphilis CW, Ma H, Nei M. 2003. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Molecular Biology and Evolution* 20: 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. *Proceedings of the National Academy of Sciences, USA* 101: 1910–1915.
- Nayar S, Thangavel G. 2021. CsubMADS1, a lag phase transcription factor, controls development of polar eukaryotic microalga *Coccomyxa subellipsoidea* C-169. *The Plant Journal* 107: 1228–1242.
- Ng M, Yanofsky MF. 2001. Function and evolution of the plant MADS-box gene family. *Nature Reviews. Genetics* 2: 186–195.
- Qiu Y, Köhler C. 2022. Endosperm evolution by duplicated and neofunctionalized Type I MADS-Box transcription factors. *Molecular Biology and Evolution* 39: msab355.
- Qiu Y, Li Z, Walther D, Köhler C. 2023. Updated phylogeny and protein structure predictions revise the hypothesis on the origin of MADS-box transcription factors in land plants. *Molecular Biology and Evolution* 40: msad194.
- Riese M, Faigl W, Quodt V, Verelst W, Matthes A, Saedler H, Münster T. 2005. Isolation and characterization of new MIKC*-type MADS-box genes from the moss *Physcomitrella patens*. *Plant Biology* 7: 307–314.
- 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. *Molecular Biology and Evolution* 40: msad088.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.
- Smaczniak C, Immink RG, Angenent GC, Kaufmann K. 2012. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139: 3081–3098.
- Svensson ME, Johannesson H, Engström P. 2000. The LAMB1 gene from the clubmoss, *Lycopodium annotinum*, is a divergent MADS-box gene, expressed specifically in sporogenic structures. *Gene* 253: 31–43.
- 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. *Proceedings of the National Academy of Sciences, USA* 102: 2436–2441.
- Thangavel G, Nayar S. 2018. A survey of MIKC Type MADS-Box genes in non-seed plants: algae, bryophytes, lycophytes and ferns. *Frontiers in Plant Science* 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. *Journal of Molecular Evolution* 43: 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: 3259–3271.
- Verelst W, Saedler H, Münster T. 2007a. MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiology* 143: 447–460.
- Verelst W, Twell D, de Folter S, Immink R, Saedler H, Münster T. 2007b. MADS-complexes regulate transcriptome dynamics during pollen maturation. *Genome Biology* 8: R249.
- Zobell O, Faigl W, Saedler H, Münster T. 2010. MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Molecular Biology and Evolution* 27: 1201–1211.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Maximum-likelihood tree based on the amino acid alignments including all MADS-box TFs from a series of representative land plants across all major lineages.

Fig. S2 Maximum-likelihood tree based on the codon alignments including all MADS-box TFs from a series of representative land plants across all major lineages.

Fig. S3 Maximum-likelihood (ML) tree of selected land plant MADS-box TFs and charophytic, green algal and Cyanophora paradoxa sequences.

Fig. S4 Bayesian inference of phylogeny of surveyed MADS domains.

Fig. S5 Maximum-likelihood tree of surveyed MADS domains without sequences from charophytes.

Fig. S6 Topology tests for phylogenetic trees constraint to certain evolutionary models with all the MADS-box TFs in Fig. 1(b).

Fig. S7 Topology tests for phylogenetic trees constraint to certain evolutionary models with the selected MADS-box TFs in Fig. S3.

Notes S1 The alignments used in the phylogenetic analyses.

Table S1 Species surveyed in this study.

Table S2 Land plant MADS-box sequences in the phylogenetic analyses corresponding to Fig. 1(b).

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Key words: gene duplication, land plants, MADS-box transcription factors, MIKC*, Type I.

Received, 12 July 2024; accepted, 30 July 2024.