

RESEARCH

Open Access



Evolutionary divergence of motifs in B-class MADS-box proteins of seed plants

Gangxu Shen^{1,2*†}, Yong Jia^{3†} and Wei-Lung Wang^{2*}

Abstract

Background: MADS-box transcription factors function as homo- or heterodimers and regulate many aspects of plant development; moreover, MADS-box genes have undergone extensive duplication and divergence. For example, the morphological diversity of floral organs is closely related to the functional divergence of the MADS-box gene family. B-class genes (such as *Arabidopsis thaliana* *APETALA3* [*AP3*] and *PISTILLATA* [*PI*]) belong to a subgroup of MADS-box genes. Here, we collected 97 MADS-box B protein sequences from 21 seed plant species and examined their motifs to better understand the functional evolution of B proteins.

Results: We used the MEME tool to identify conserved sequence motifs in these B proteins; unique motif arrangements and sequences were identified in these B proteins. The keratin-like domains of *Malus domestica* and *Populus trichocarpa* B proteins differed from those in other angiosperms, suggesting that a novel regulatory network might have evolved in these species. The MADS domains of *Nelumbo nucifera*, *Glycine max*, and *Amborella trichopoda* B-proteins contained motif 9; in contrast, those of other plants contained motif 1. Protein modelling analyses revealed that MADS domains with motif 9 may lack amino acid sites required for DNA-binding. These results suggested that the three species might share an alternative mechanism controlling floral development.

Conclusions: *Amborella trichopoda* has B proteins with either motif 1 or motif 9 MADS domains, suggesting that these two types of MADS domains evolved from the ancestral domain into two groups, those with motif 9 (*N. nucifera* and *G. max*), and those with motif 1. Moreover, our results suggest that the homodimer/heterodimer intermediate transition structure first appeared in *A. trichopoda*. Therefore, our systematic analysis of the motifs in B proteins sheds light on the evolution of these important transcription factors.

Keywords: B gene, MADS-box gene, MADS domain, MEME, *Amborella trichopoda*

Background

The versatile MADS-box transcription factors (TFs) shape development in most multicellular eukaryotes; however, terrestrial plants have more MADS-box genes than other eukaryotic groups [1–4]. Indeed, during the evolution of flowering plants, MADS-box genes had key

roles in shaping the diverse structures of flowers [5, 6] and these diverse structures (as well as seed-based propagation) in turn had key roles in the success of flowering plants [7]. Therefore, studies of MADS-box genes provide insight into the development of the diverse flower types found in angiosperms [8]. The term MADS-box gene is derived from four of the earliest recognized family members: *MCM1*, *AGAMOUS*, *DEFICIENS*, and *serum response factor* (*SRF*), from *Saccharomyces cerevisiae*, *AGAMOUS* from *Arabidopsis thaliana*, *DEFICIENS* from *Antirrhinum majus*, and *serum response factor* (*SRF*) from *Homo sapiens* [8, 9], respectively. During evolution, MADS-box genes underwent multiple

*Correspondence: numbershen@yahoo.com.tw; wllwang@cc.ncue.edu.tw

†Gangxu Shen and Yong Jia have contributed equally to the study

¹ School of Chinese Medicine for Post-Baccalaureate, I-Shou University, Kaohsiung 84001, Taiwan

² Department of Biology, National Changhua University of Education, Changhua 500, Taiwan

Full list of author information is available at the end of the article



duplications and divergence, producing two major clades: the Serum Response Factor-like clade (type I) and the Myocyte Enhancer Factor-2-like clade (type II) [10, 11]. Charophyte algae and terrestrial plants have MIKC-type TFs that contain MADS, Intervening, Keratin-like, and C-terminal (MIKC) domains [12, 13]. MIKC-type TFs can be further divided into the MIKC* and the MIKC^C groups [14].

One subgroup of the MIKC^C-type MADS-box genes, the B genes, are defined from *Arabidopsis thaliana* *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), which

function in the specification of floral organs. B genes are not present in bryophytes and seedless vascular plants [15, 16]. Our previous research [17] confirmed the evolutionary history of the ABCDE and AGL6 genes and clarifies their evolutionary path. Here, to shed light on the evolution of B genes, we examined B proteins from gymnosperm, basal angiosperm, monocot, and magnoliopsida/eudicot species, using 97 B protein sequences from 21 seed plant species (Table 1) to identify conserved motifs.

Table 1 97 *AP3/PI* query sequences in 21 seed plants

No	Species	<i>AP3/PI</i>	No	Species	<i>AP3/PI</i>	No	Species	<i>AP3/PI</i>
1	<i>Gnetum gnemon</i>	GGM2	11	<i>Carica papaya</i>	CpMADS22	16	<i>Populus trichocarpa</i> *	PtMADS10
2	<i>Ginkgo biloba</i>	GbMADS4 GbMADS9			CpMADS23 CpMADS24			PtMADS11 PtMADS22
3	<i>Amborella trichopoda</i>	LOC18424280 LOC18448591 LOC18429933 LOC18436882	12	<i>Malus domestica</i> *	MdMADS13P MdMADS31 MdMADS64 MdMADS65*			PtMADS25 PtMADS30* PtMADS38* PtMADS45*
4	<i>Oryza sativa</i>	OsMADS2 OsMADS4 OsMADS16			MdMADS99* MdMADS105 MdMADS121*	17	<i>Linum usitatissimum</i>	LuMADS34 LuMADS67 LuMADS69
5	<i>Zea mays</i>	ZmMADS20 ZmMADS60			MdMADS124 MdMADS127*			LuMADS74 LuMADS94
6	<i>Phalaenopsis aphrodite</i>	PATC133864 PATC138350 PATC152852 PATC154853 PATC240636			MdMADS131 MdMADS134 MdMADS139 MdMADS151*	18	<i>Solanum tuberosum</i>	LuMADS117 LuMADS120 StMADS54 StMADS61 StMADS86
7	<i>Musa acuminata</i>	MaMADS6p MaMADS14 MaMADS71 MaMADS88	13	<i>Cucumis sativus</i>	CsMADS21 CsMADS23			StMADS112 StMADS145 StMADS153 StMADS253
8	<i>Arabidopsis thaliana</i>	AT3G54340 AT5G20240	14	<i>Glycine max</i>	GmMADS4 GmMADS5 GmMADS6 GmMADS7 GmMADS8 GmMADS9	19	<i>Ricinus communis</i>	RcMADS30 RcMADS33 RcMADS34
9	<i>Vitis vinifera</i>	VvAP3_1 VvAP3_2 VvPI			GmMADS10 GmMADS110 GmMADS121 GmMADS133 GmMADS147 GmMADS175	20	<i>Manihot esculenta</i>	MeMADS45 MeMADS46 MeMADS47 MeMADS48 MeMADS52
10	<i>Citrus sinensis</i>	CsiMADS33 CsiMADS34 CsiMADS35 CsiMADS41 CsiMADS42 CsiMADS43 CsiMADS46	15	<i>Nelumbo nucifera</i>	ABE11602 ADD25193 ADD25194 ADD25195	21	<i>Solanum lycopersicum</i>	TM6

The name of the bolded species are significantly different in M domain. The name of the asterisks* species are significantly different in K1 domain. Gymnosperms (number 1–2) and angiosperms (number 3–21): Basal angiosperm (3. *Amborella trichopoda*), Monocots (4. *Oryza sativa*, 5. *Zea mays*, 6. *Phalaenopsis aphrodite*, and 7. *Musa acuminata*), and magnoliopsida and eudicots (8–21)

Results

Identification of B class gene sequences

We retrieved 97 known B protein sequences (Table 1) using proteins from *A. thaliana* (*AT3G54340* and *AT5G20240*) and *Oryza sativa* (*OsMADS2*, 4, and 16) as query sequences [1, 3, 6, 18–20] in a Basic Local Alignment Search Tool (BLAST) search [20]. Subsequently, the retrieved sequences were entered into Simple Modular Architecture Research Tool (SMART) to confirm they have MADS-box domains [22]. Sequence alignment of the MADS domains was displayed in Additional file 2: Fig. S1.

Motif identification in B class genes

We used the MEME tool [3] to identify conserved sequence motifs in 97 B protein sequences in this study (Fig. 1 and Additional file 1). A total of 10 conserved motifs were identified in the AP3/PI proteins among 21 plant species (Fig. 1). The detailed amino acid sites conservation profiles are shown in Additional file 2: Fig. S2. Motif 1 (consensus sequence IEIKRIENPTNRQV-TYSKRRNGIFKKAHELTVLCLDAKVS-LIMFSS) and motif 9 (consensus sequence KAAELTVLCLDAKVS-LIMFSST) overlap with the MADS domain (M domain). In the B proteins of *Nelumbo nucifera* (*ABE11602*, *ADD25193*, *ADD25194*, and *ADD25195*), *Glycine max* (*GmMADS121*, *GmMADS133*, *GmMADS147*, and *GmMADS175*), and some of the B proteins in *Amborella trichopoda* (*LOC18429933* and *LOC18424280*), motif 1 was replaced by motif 9 (Fig. 1). However, as shown in Fig. 1, some *A. trichopoda* B protein sequences have motif 1 (*LOC18436882*, and *LOC18448591*). The M domains of gymnosperm B proteins belong to the motif 1 group (Fig. 1).

Gymnosperm B proteins (*GbMADS4*, *GbMADS9*, and *GGM2*) lack motif 4, which overlaps with the intervening domain (I domain), consists of approximately 30 amino acids and is a less-conserved region involved in protein dimerization [10].

Motifs 2, 3, 5, 7, and 10 were found in the keratin-like domain (K-domain). The K domain of AP3/PI consists of approximately 70 amino acids that are divided into K1, K2, and K3 subdomains [13]. K1 and K2 are required for dimer formation, whereas K3 may participate in multimerization [13, 23]. We found motif 2 and motif 10 in the K1 subdomain; motif 3 was found in K2; and motif

5, with the consensus sequence KYHVIKTQTDTCCK-KVRNLEE, or its alternative motif 7, with the consensus sequence QMEYWKMMKRNDKMLEDENKQLTF, were found in K3 (Figs. 1 and Additional file 2: Fig. S2).

Motif 10 of *Malus domestica* (*MdMADS65*, *MdMADS99*, *MdMADS121*, *MdMADS127*, and *MdMADS151*) and *Populus trichocarpa* (*PtMADS30*, *PtMADS38*, and *PtMADS45*) replaced motif 2 in the K1 domain of other angiosperms. Motifs 6 and 8 were found in the C-terminal domain (C domain) (Fig. 1).

Phylogeny analyses of plant AP3/PI genes

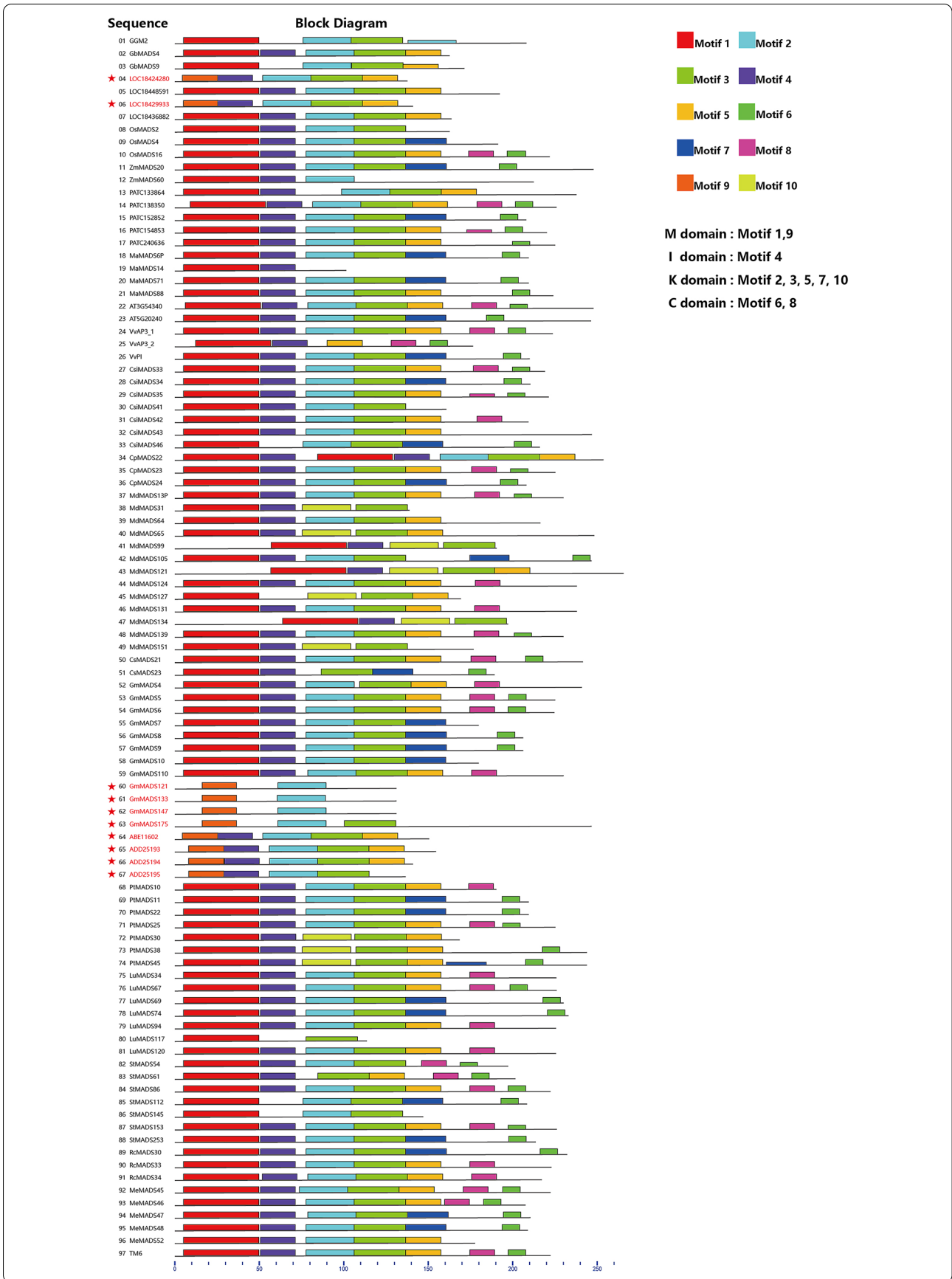
To investigate the phylogenetic relationship among these 97 AP3/PI sequences (Additional file 1), a Bayesian phylogeny was reconstructed (Fig. 2). Overall, the determined phylogeny was consistent with the species tree, indicating the phylogeny was reliable. As shown in Fig. 2, the target genes have divided into two major groups, representing AP3 and PI, respectively. Those MADS protein sequences containing motif 9 (highlighted in red) were identified within both AP3 group (*LOC18424280*, *ABE11602*, *ADD25194*, *GmMADS121*, *GmMADS133*, *GmMADS147*, and *GmMADS175*) and the PI group (*ADD25195* and *LOC18429933*), indicating motif 9 may have evolved independently within the two separate lineages. The position of *Amborella* MADS genes (gene IDs starts with LOC) was consistent with this species as an evolutionary intermediate between lower plants and core eudicot plants. Noteworthy, *Amborella* MADS genes *LOC18448591* and *LOC18436882* (highlighted in green) containing motif 1 were identified both in AP3 and PI groups (Fig. 2).

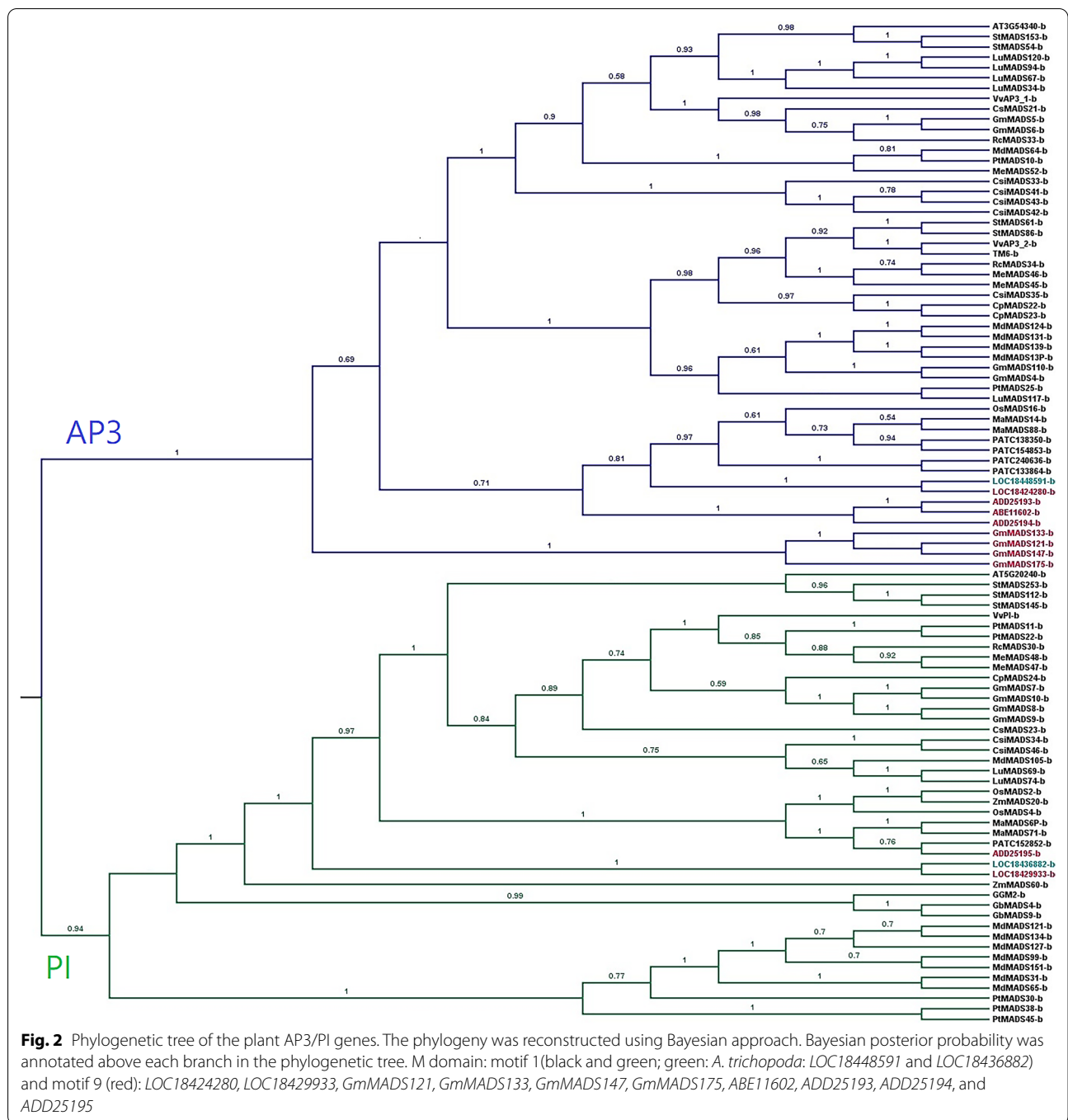
Protein structural modelling of MADS domain

To investigate the potential impact of the substitution between motif 1 and motif 9, protein structural modelling was performed for plant MADS domains using *AT5G20240* as a reference. The previously well-determined structure of human myocyte enhancer factor-2 (MEF2, PDB ID: 1TQE, MADS-box superfamily) in complex with DNA and its interacting protein (Fig. 3A) was used as a template for homology (~48% aa identity) structural modelling. Structural superimposition (Fig. 3B) showed the MADS domain of *AT5G20240* was well-conserved in comparison with MEF2. Based on 3D structural superimposition with MEF2, the

(See figure on next page.)

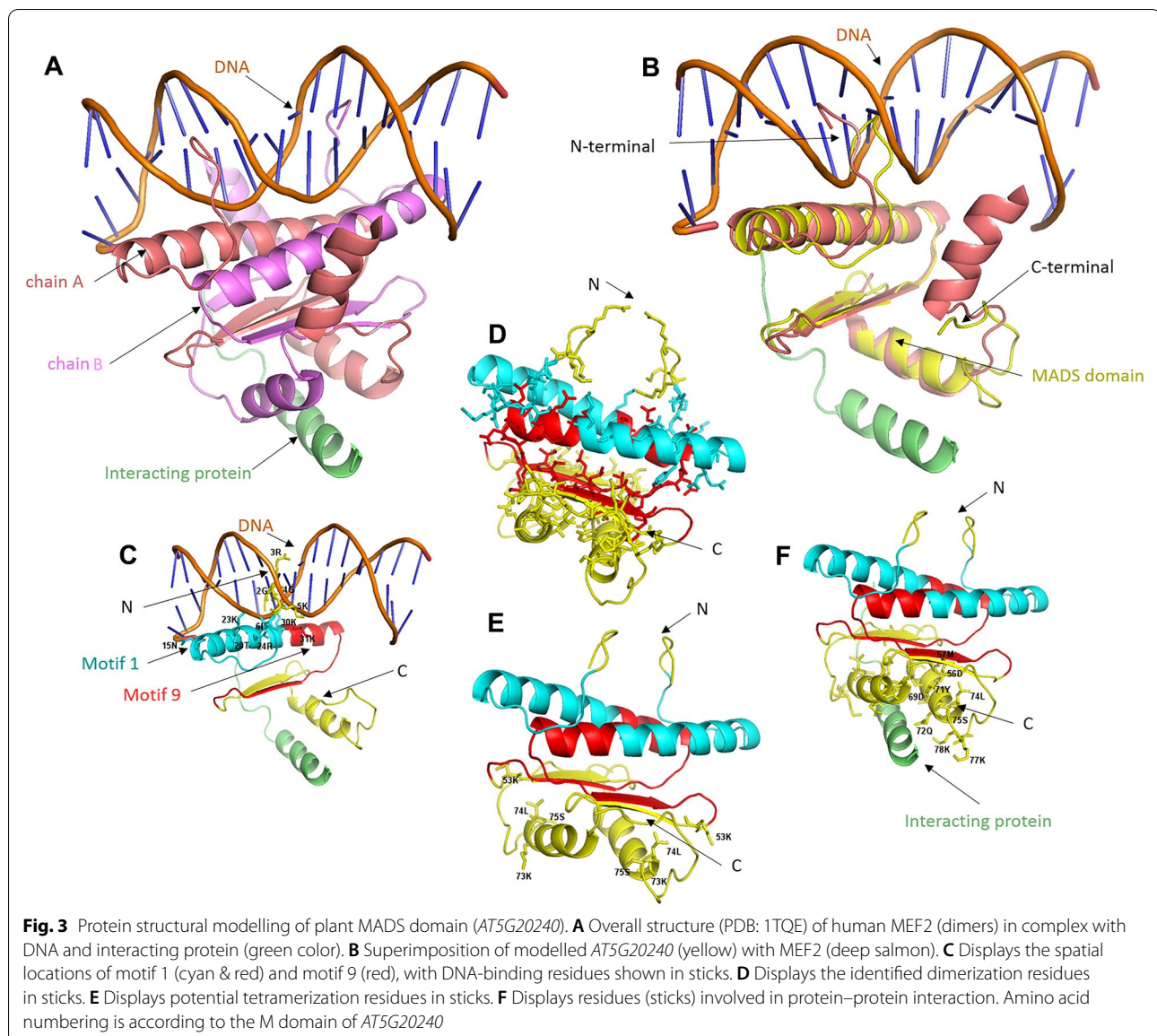
Fig. 1 MEME search results regarding the protein motifs in AP3/PI. The protein motifs of the B proteins in 97 sequences were separately obtained using the MEME motif search tool for each group. Ten motifs were identified, each of which is represented as a colored box. Asterisks and red letters: motif 9. Gymnosperms (numbers 1–3) and angiosperms (numbers 4–97): Basal angiosperm (*Amborella trichopoda*: 4–7), monocots (*Oryza sativa*: 8–10, *Zea mays*: 11–12, *Phalaenopsis aphrodite*: 13–17 and *Musa acuminata*: 18–21), and magnoliopsida and eudicots (22–97)





spatial positions of motif 1 and motif 9 of *AT5G20240* in reference to binding DNA were displayed (Fig. 3C). A total of 10 amino acids in *AT5G20240* were identified as potential DNA-binding sites (shown in sticks in Fig. 3C, Additional file 2: Fig. S3), six of which were located in motif 1 versus only one in motif 9, suggesting motif 1, instead of motif 9, may be responsible for DNA-binding.

As indicated in the MEF2 structure, the MADS domain binds DNA in the dimer form. Thus, amino acid sites for dimerization were also identified for *AT5G20240* (Fig. 3D). A total of 52 residues out of the 89 MADS domain amino acids were identified for dimerization interaction (Fig. 3D, Additional file 2: Fig. S3), which cover the majority of motif 9 residues and also a significant number of residues in motif 1. Motif 1, which



overlaps with motif 9, was intertwined for the MADS dimers to form a dimer complex (Fig. 3D & E), suggesting that both motif 1 and motif 9 were critical for the dimerization interaction. In addition, residues for protein–protein interaction (Fig. 3E) and potential tetramerization (Fig. 3F) were also identified and displayed. None of these interactions involved residues from motif 1 and motif 9.

Discussion

The MADS domains in *N. nucifera*, *G. max*, and *A. trichopoda* contain motif 9, whereas the MADS (M) domain of the B proteins from other plants contain motif 1. The two motifs overlap; motif 9 is a shorter version of motif 1, missing the first half. It has been suggested

that all MADS TFs function as biological dimers [24, 25]. Among the MIKC domains, residues in domains M, I, and K are involved in the dimerization process [25, 26]. In this study, protein structural modelling suggested motif 1 and motif 9 in the M domain were mainly involved in the dimerization process. Based on the determined structure of human MEF2 [26], it seems that the formation of a dimer is critical for the DNA-binding process. For example, our modelling results showed that the third residue arginine in the M domain was responsible for both DNA binding and dimerization. Indeed, a previous study has proven that this residue is critical for the recognition of DNA binding sites [27]. Notably, the third arginine residue and several other residues identified for

DNA binding in this study were missing in motif 9 compared to a complete motif 1. Therefore, the variation between motif 1 and motif 9 may affect the formation of functional MADS proteins. These observations suggested that either MADS proteins with motif 9 identified in *N. nucifera*, *G. max*, and *A. trichopoda* are dysfunctional or that these three species may share an alternative control mechanism for floral development. Moreover, differences in gene expression support the potential divergence of floral development in these species. In our study, three *N. nucifera* genes (*ABE11602*[*Nenu.AP3*], *ADD25193* [*Nenu.AP3-1*], and *ADD25194* [*Nenu.AP3-2*]) were placed in the *AP3* and *ADD25195* belongs to *PI* (Fig. 1). Indeed, three *N. nucifera* *PI* and *AP3* homologs (*Nenu.PI*, *Nenu.AP3-1*, and *Nenu.AP3-2*) are expressed in petals and stamens, similar to other B genes in eudicots [28]. *Nenu.PI* is additionally expressed in sepals, while *Nenu.AP3-1* is also detected in carpels. [28]. Therefore, sepal and petal differentiation in *N. nucifera* might be related to the alternative M domains in its B genes.

The observation that some *A. trichopoda* B protein sequences contain motif 9 and others have motif 1 suggests that the M domains of angiosperm B proteins evolved from an ancestor of *A. trichopoda* into two groups: a few species have B proteins with M domains that have motif 9 (*N. nucifera* and *G. max*) and most have motif 1. The M domains of gymnosperms belong to the motif 1 group (Fig. 1), so we speculate that the motif-1 M domain evolved earlier than motif 9. The B protein repertoire of *A. trichopoda* has both types of M domains, potentially increasing its capacity for floral organ diversity.

Studies of B proteins from across plant species have shed light on the evolution of their protein–protein interactions. For example, a gymnosperm B protein from *Gnetum gnemon* (motif 1) binds DNA as a homodimer [29]. Our protein modelling analyses also supported that plant MADS proteins function as a biological dimer, in which the dimerization is critical for target DNA-binding. The monocot *Lilium regale* has two classes of B proteins: the GLOBOSA (GLO)-like proteins (such as *PI*) homodimerize, but the DEFICIENS-like proteins (such as *AP3*) form heterodimers with GLO-like proteins [29]. These data suggest that B proteins evolved to form homodimers and then to form heterodimers, but in the evolution of these two structures, homodimer/heterodimer transition structures appeared. The MADS domain is important for DNA binding and dimerization [29]. As shown in Fig. 1, *A. trichopoda* has a complex repertoire of M domains (motifs 1 and 9). Based on structural modelling, this study identified the specific amino acid residues responsible for dimerization interactions, which involved a significant number of residues located in motif

1 and motif 9. Analysis of conserved motifs in B proteins reveals that the homodimer/heterodimer intermediate transition structure may have first appeared in the ancestors of *A. trichopoda* (Fig. 2), because *A. trichopoda* genes *LOC18448591* and *LOC18436882* (highlighted in green, Fig. 2) containing motif 1, as well as *LOC18424280* and *LOC18429933* (highlighted in red, Fig. 2) containing motif 9, were identified both in *AP3* and *PI* groups.

Many of the other motifs identified here affect domains involved in protein–protein interactions. For example, motif 4 is in the I domain, which is involved in protein dimerization [10]; it was surprising to find that motif 4 was absent in gymnosperm B proteins. Motifs 2, 3, 5, 7, and 10 were found in the K domain and the K1 and K2 subdomains are required for dimer formation, whereas K3 may participate in multimerization [13, 23]. Moreover, motif 10 of *Malus domestica* and *Populus trichocarpa* B proteins replaced motif 2 in the K1 domain of other angiosperms, suggesting that an alternative regulatory network might have evolved in these two species. Shan et al. have put forward similar views in a previous paper [30], but in this study, the related protein motif analysis was performed after *A. trichopoda* was sequenced [31]. Therefore, our research results can objectively confirm this result. Molecular evolutionary analyses provide a powerful approach for the identification of changes in amino acids, possibly associated with the evolution of gene function [30]. It is hypothesized that homodimerization or heterodimerization of the *AP1/SQUA*, *AP3/PI*, *AG*, and *SEP*-like proteins function as master regulators of the floral development in *Arabidopsis thaliana* [32].

Conclusions

According to our protein motif analysis, the B proteins of seed plants exhibited unique motif arrangements and sequences that might affect their floral development. For example, the differentiation of sepals and petals in *N. nucifera* might be related to the alternative M domains in their B proteins. *Amborella trichopoda* has a complex repertoire of M domains and this may affect its floral organ diversity. Moreover, our results shed light on B protein evolution, suggesting that the M domains of angiosperms evolved from an ancestor of *A. trichopoda* to form two groups: a small group with the motif 9 M domain (*N. nucifera* and *G. max*), and most angiosperms with the motif 1 M domain. Due to the M domains of gymnosperm B proteins, belonging to the motif 1 group, we speculated that the evolution of the motif 1 group occurred earlier than that of the motif 9 group. Finally, our results provide insight into the evolution of protein–protein interactions, suggesting that interaction and dimerization critically rely on both motif 1 and motif 9 groups. The homodimer/heterodimer intermediate

transition structure may have first appeared in *A. trichopoda*, because of the fact that the *A. trichopoda* motif 1 (*LOC18448591* and *LOC18436882*) and motif 9 (*LOC18424280* and *LOC18429933*) were both identified in *AP3* and *PI* groups. Moreover, the B proteins (K1 domain) of *M. domestica* and *P. trichocarpa* differ from those in other angiosperms, suggesting that an alternative regulatory network might have evolved in these two species. Our results will assist researchers in exploring MADS-box protein functional evolution.

Methods

Species selection

The selection of representative gymnosperm species was performed from a range of families, including Gnetales (*G. gnemon*), Pinaceae (*Picea abies*), Podocarpaceae (*Podocarpus macrophyllus*), Araucariaceae (*Wollemia nobilis*), Sciadopityaceae (*Sciadopitys verticillata*), Taxaceae (*Taxus baccata*), Cupressaceae (*Cryptomeria japonica*), and Ginkgoaceae (*G. biloba*); however, but we found that only *G. gnemon* and *G. biloba* were found to have the B gene [17]. When selecting angiosperms, we included species from the three groups: (1) basal angiosperms (*A. trichopoda*), (2) monocots (*Osativa*, *Zea mays*, *Phalaenopsis aphrodite*, and *Musa acuminata*), and (3) magnoliopsida and eudicots. Since magnoliopsida and eudicots are the largest group of angiosperms, we chose to include 14 typical species from the different families in this group. We considered choosing these seed plants (gymnosperms and angiosperms) to represent a complete evolution of plants, which is of crucial importance for the phylogenetic analysis. Furthermore, MADS-box genes are present in bryophytes and seedless vascular plants, whereas ABCDE and AGL6 genes have MADS-box genes but do not have ABCDE or AGL6 [33].

Sequence retrieval

Query sequences were obtained from the following databases: *A. thaliana*, <http://www.arabidopsis.org/>; *O. sativa*, <http://rice.plantbiology.msu.edu/>; *P. aphrodite*, <http://orchidstra.abrc.sinica.edu.tw/>; Gramene, <http://www.gramene.org/>; National Center for Biotechnology Information (NCBI), <http://www.ncbi.nlm.nih.gov/>; Phytozome, <http://www.phytozome.net/>; and UNIPROT, <http://www.uniprot.org/uniprot/>. B protein sequences from *A. thaliana* and *O. sativa* were used for BLAST-P queries against available genomic data.

MADS-box sequences validation

Online tool SMART (<http://smart.embl-heidelberg.de/>) was used to validate the presence of the MADS domains in the proteins encoded by the target genes [22].

Conserved motif prediction using MEME

MEME [3] online tool (meme.sdsc.edu/meme/meme-intro.html) was used to identify conserved motifs. A MEME search was performed with the following settings: (1) optimum motif width was set to ≥ 6 and ≤ 50 ; (2) the maximum number of motifs identified was set to 10 and (3) occurrences of a single motif were distributed among the sequences with the setting “zero” or “one per sequence” (-mod zoops).

Sequence alignment and phylogeny reconstruction

The amino acid sequences were aligned using the program MUSCLE version 3.8.31 [34] with 8 iterations. The phylogeny was searched by Bayesian simulation implemented using BEAST2 [35] program. The strict molecular clock assumption with an unlinked substitution model Yule + G (five categories) was used. A single Markov Chain Monte Carlo chain was run for 1,000,000 generations (sampled every 1000) with the first 1000 trees as pre burn-in until convergence. The final phylogenetic tree was inferred by treeannotator [35], with the first 5% trees discarded.

Protein structural modelling

Homology structural modelling of plant MADS domain was performed using Modeller [25] server based on sequence alignment in Chimera. Human MEF2 (PDB: 1TQE) [36] in complex with binding DNA and interacting protein was selected as the template (48% aa identity). The quality of the generated model was assessed based on lowest discrete optimized protein energy (DOPE) values and GA 341 score of 1, which indicated the reliability of these models. Potential interacting residues were identified based on distance less than 4 Å using PyMol 1.3r1 (Schrodinger, LLC; <https://pymol.org>). All structural visualizations were performed using PyMol.

Abbreviations

AGL6: AGAMOUS- LIKE 6; AP: APETALA; BEAST: Bayesian evolutionary analysis by sampling trees; BLAST: Basic Local Alignment Search Tool; DOPE: Discrete optimized protein energy; PI: PISTILLATA; SMART: Simple Modular Architecture Research Tool; SRF: Serum response factor.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40709-021-00144-7>.

Additional file 1. 97 B protein sequences.

Additional file 2: Fig. S1. Sequence alignment of MADS domains of plant AP3/PI genes ordered by the developed phylogeny. Motif 9 were highlighted in red box. **Fig. S2.** Conserved sequence profiles for motif 1-10. **Fig. S3.** Identified interacting residues in MADS domain modelling.

A) DNA-binding residues. B) Dimerization residues. C) Protein-protein interaction residues. D) Potential tetramerization residues. Cyan & red color indicate motif 1. Red color indicates motif 9. The other residues are in yellow color. The identified residues are highlighted in each row.

Acknowledgements

The authors would like to thank the relevant research community for making the genomic data available to the public.

Authors' contributions

GS performed all the research and drafted the manuscript. YJ performed protein structural modelling analyses and wrote the manuscript. WLW supervised the study and revised the manuscript. All authors have read and approved the manuscript.

Funding

No.

Availability of data and materials

The *O. sativa* database (<http://rice.plantbiology.msu.edu/>), the *A. thaliana* database (<http://www.arabidopsis.org/>), NCBI (<http://www.ncbi.nlm.nih.gov/>), Gramene (<http://www.gramene.org/>), UNIPROT (<http://www.uniprot.org/uniprot/>), Phytosome (<http://www.phytosome.net/>), and *P. aphrodite* database (<http://orchidstra.abrc.sinica.edu.tw/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declare that they have no competing interests.

Author details

¹School of Chinese Medicine for Post-Baccalaureate, I-Shou University, Kaohsiung 84001, Taiwan. ²Department of Biology, National Changhua University of Education, Changhua 500, Taiwan. ³College of Science, Health, Engineering and Education, Murdoch University, Murdoch, WA 6150, Australia.

Received: 13 January 2021 Accepted: 19 May 2021

Published online: 28 May 2021

References

- Pařenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 2003;15:1538–51.
- Leseberg CH, Li A, Kang H, Duvall M, Mao L. Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene*. 2006;378:84–94.
- Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics*. 2007;8:242.
- Dreni L, Kater MM. MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol*. 2014;201:717–32.
- Urbanus SL, de Folter S, Shchennikova AV, Kaufmann K, Immink RGH, Angenent GC. In planta localisation patterns of MADS domain proteins during floral development in *Arabidopsis thaliana*. *BMC Plant Biol*. 2009;9:5.
- Murai K. Homeotic genes and the abcde model for floral organ formation in wheat. *Plants (Basel)*. 2013;2:379–95.
- Linkies A, Graeber K, Knight C, Leubner-Metzger G. The evolution of seeds. *New Phytol*. 2010;186:817–31.
- Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*. 2011;23:865–72.
- Trobner W, Ramirez L, Motte P, Hue I, Huijser P, Lonig WE, et al. GLO-BOSA: a homeotic gene which interacts with DEFICIENS in the control of *Antirrhinum* floral organogenesis. *EMBO J*. 1992;11:4693–704.
- Jiang SC, Pang CY, Song MZ, Wei HL, Fan SL, Yu SX. Analysis of MIKC^C-type MADS-box gene family in *Gossypium hirsutum*. *J Integr Agr*. 2014;13:1239–49.
- Gramzow L, Ritz MS, Theißen G. On the origin of MADS-domain transcription factors. *Trends Genet*. 2010;26:149–53.
- Theißen G, Kim JT, Saedler H. 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*. 1996;43:484–516.
- Kaufmann K, Melzer R, Theißen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene*. 2005;347:183–98.
- Gramzow L, Theißen G. A hitchhiker's guide to the MADS world of plants. *Genome Biol*. 2010;11:214.
- Hasebe M. Evolution of reproductive organs in land plants. *J Plant Res*. 1999;112:463–74.
- Gramzow L, Barker E, Schulz C, Ambrose B, Ashton N, Theißen G, et al. *Selaginella* genome analysis—entering the “homoplasy heaven” of the MADS World. *Front Plant Sci*. 2012;3:214.
- Shen G, Yang CH, Shen CY, Huang KS. Origination and selection of ABCDE and AGL6 subfamily MADS-box genes in gymnosperms and angiosperms. *Biol Res*. 2019;52:25.
- Airoldi CA, Davies B. Gene duplication and the evolution of plant MADS-box transcription factors. *J Genet Genomics*. 2012;39:157–65.
- Ohmori S, Kimizu M, Sugita M, Miyao A, Hirochika H, Uchida E, et al. *MOSAIC FLORAL ORGANS1*, an *AGL6*-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell*. 2009;21:3008–25.
- Immink RG, Kaufmann K, Angenent GC. The “ABC” of MADS domain protein behaviour and interactions. *Semin Cell Dev Biol*. 2010;21:87–93.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
- Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, et al. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res*. 2002;30:242–4.
- Gramzow L, Theißen G. Phylogenomics of MADS-box genes in plants—two opposing life styles in one gene family. *Biology (Basel)*. 2013;2:1150–64.
- Tilly JJ, Allen DW, Jack T. The CARG boxes in the promoter of the *Arabidopsis* floral organ identity gene *APETALA3* mediate diverse regulatory effects. *Development*. 1998;125:1647–57.
- Lai X, Daher H, Galien A, Hugouvieux V, Zubieta C. Structural basis for plant MADS transcription factor oligomerization. *Comput Struct Biotechnol J*. 2019;17:946–53.
- Han A, He J, Wu Y, Liu JO, Chen L. Mechanism of recruitment of class II histone deacetylases by myocyte enhancer factor-2. *J Mol Biol*. 2005;345:91–102.
- Kappel S, Melzer R, Rumpler F, Gafert C, Theißen G. The floral homeotic protein *SEPALLATA3* recognizes target DNA sequences by shape read-out involving a conserved arginine residue in the MADS-domain. *Plant J*. 2018;95:341–57.
- Yoo MJ, Soltis PS, Soltis DE. Expression of floral MADS-box genes in two divergent water lilies: Nymphaeales and *Nelumbo*. *Int J Plant Sci*. 2010;171:121–46.
- Winter KU, Weiser C, Kaufmann K, Bohne A, Kirchner C, Kanno A, et al. Evolution of Class B Floral Homeotic Proteins: Obligate Heterodimerization Originated from Homodimerization. *Mol Biol Evol*. 2002;19:587–96.
- Shan H, Zahn L, Guindon S, Wall PK, Kong H, Ma H, et al. Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol Biol Evol*. 2009;26:2229–44.
- Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, Ma H, et al. The *Amborella* genome and the evolution of flowering plants. *Science*. 2013;342:1241089.

32. Yang Y, Fanning L, Jack T. The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J*. 2003;33:47–59.
33. Hasebe M, Wen CK, Kato M, Banks JA. Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proc Natl Acad Sci USA*. 1998;95:6222–7.
34. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:113.
35. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput Biol*. 2014;10:e1003537.
36. Wu Y, Dey R, Han A, Jayathilaka N, Philips M, Ye J, et al. Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for myocardin recruitment. *J Mol Biol*. 2010;397:520–33.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

