

RESEARCH ARTICLE

Open Access



# Origination and selection of ABCDE and *AGL6* subfamily MADS-box genes in gymnosperms and angiosperms

Gangxu Shen<sup>1,2\*</sup>, Chih-Hui Yang<sup>3</sup>, Chi-Yen Shen<sup>1</sup> and Keng-Shiang Huang<sup>2,3\*</sup>

## Abstract

**Background:** The morphological diversity of flower organs is closely related to functional divergence within the MADS-box gene family. Bryophytes and seedless vascular plants have MADS-box genes but do not have ABCDE or *AGAMOUS-LIKE6* (*AGL6*) genes. ABCDE and *AGL6* genes belong to the subgroup of MADS-box genes. Previous works suggest that the B gene was the first ABCDE and *AGL6* genes to emerge in plant but there are no mentions about the probable origin time of ABCDE and *AGL6* genes. Here, we collected ABCDE and *AGL6* gene 381 protein sequences and 361 coding sequences from gymnosperms and angiosperms and reconstructed a complete Bayesian phylogeny of these genes. In this study, we want to clarify the probable origin time of ABCDE and *AGL6* genes is a great help for understanding the role of the formation of the flower, which can decipher the forming order of MADS-box genes in the future.

**Results:** These genes appeared to have been under purifying selection and their evolutionary rates are not significantly different from each other. Using the Bayesian evolutionary analysis by sampling trees (BEAST) tool, we estimated that: the mutation rate of the ABCDE and *AGL6* genes was  $2.617 \times 10^{-3}$  substitutions/site/million years, and that B genes originated 339 million years ago (MYA), CD genes originated 322 MYA, and A genes shared the most recent common ancestor with E/*AGL6* 296 MYA, respectively.

**Conclusions:** The phylogeny of ABCDE and *AGL6* genes subfamilies differed. The *APETALA1* (*AP1* or A gene) subfamily clustered into one group. The *APETALA3/PISTILLATA* (*AP3/PI* or B genes) subfamily clustered into two groups: the *AP3* and *PI* clades. The *AGAMOUS/SHATTERPROOF/SEEDSTICK* (*AG/SHP/STK* or CD genes) subfamily clustered into a single group. The *SEPALLATA* (*SEP* or E gene) subfamily in angiosperms clustered into two groups: the *SEP1/2/4* and *SEP3* clades. The *AGL6* subfamily clustered into a single group. Moreover, ABCDE and *AGL6* genes appeared in the following order: *AP3/PI* → *AG/SHP/STK* → *AGL6/SEP/AP1*. In this study, we collected candidate sequences from gymnosperms and angiosperms. This study highlights important events in the evolutionary history of the ABCDE and *AGL6* gene families and clarifies their evolutionary path.

**Keywords:** ABCDE gene, *AGL6*, MADS-box gene, Evolutionary events, Phylogenetic analysis

\*Correspondence: numbershen@yahoo.com.tw; huangks@isu.edu.tw

<sup>1</sup> Department of Electrical Engineering, I-Shou University, Kaohsiung, Taiwan

<sup>2</sup> The School of Chinese Medicine for Post-Baccalaureate, I-Shou University, Kaohsiung, Taiwan

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



## Background

MADS-box genes played a crucial role in the emergence of flower structures during plant evolution [1, 2]. Moreover, the role of MADS-box genes in controlling flower morphogenesis makes them ideal genetic tools for studying the development of various flower structures [3]. The number of MADS-box genes in terrestrial plants is higher than in any other group of eukaryotes [4–7]. The term MADS-box gene is derived from four of the earliest recognized family members: *MINICHROMOSOME MAINTENANCE 1* (*MCM1*) from *Saccharomyces cerevisiae*, *AGAMOUS* (*AG*) from *Arabidopsis thaliana*, *DEFICIENTS* (*DEF*) from *Antirrhinum majus*, and *SERUM RESPONSE FACTOR* (*SRF*) from *Homo sapiens* [4, 8]. An ancestral MADS-box gene was presumably duplicated before the most recent common ancestor (MRCA) of eukaryotes and evolved into two main clades, the *SRF*-like (type I) and *MEF2*-like (type II) MADS-box genes [9]. In Streptophyta (Charophyta algae and terrestrial plants), *MEF2*-like transcription factors (TFs) are often referred to as the MADS, intervening, keratin-like, and C-terminal type (*MIKC*-type) TFs, since their structures include a MADS (*M*)-domain that is followed by an intervening (*I*), a keratin-like (*K*), and a C-terminal (*C*) domains respectively [10, 11]. In terrestrial plants, the *MIKC*-type TFs form two main groups: the *MIKC\** and the *MIKC<sup>C</sup>* type [12]. After these genes emerged, flowering plants diversified substantially during the Cretaceous period to become the largest plant group on earth [7]. Their remarkable evolutionary success was primarily due to the newly evolved reproductive structures and is similar to the success of gymnosperms which use seeds as a new propagation system [13]. The *MIKC<sup>C</sup>* group can be further divided into 14 phylogenetic subfamilies [4, 6, 14], among which 10 are present in all angiosperms while 7 in all gymnosperms [6, 15]. Therefore, the appearance of *MIKC<sup>C</sup>*-type genes seems to be closely associated with the successful evolution of flowering plants.

Among *MIKC<sup>C</sup>*-type genes, the subgroups *ABCDE* and *AGAMOUS-LIKE 6* (*AGL6*) are key factors in flower development according to a proposed *ABCDE* model which suggests that combinations of various MADS-box genes determine the identity of flower organs [2, 8, 16]: *A*, *B* and *C* proteins function by interacting with *E* proteins which are necessary for all organ types [17]: *A* and *E* are present in sepals; *A*, *B* and *E* are present in petals; *B*, *C* and *E* are present in stamens; *C* and *E* are present in carpels [1, 3, 17–21]. However, some early studies in this field reported that the *E* gene is not expressed in sepals [22].

Related studies have reported MADS-box genes in gymnosperms [15, 23–27] and angiosperms [1, 3, 6, 18, 20, 28–30]. Selecting representative gymnosperm

species from a range of families, including Gnetaceae (*G. gnemon*), Pinaceae (*P. abies*), Podocarpaceae (*P. macrophyllus*), Araucariaceae (*W. nobilis*), Sciadopiaceae (*S. verticillata*), Taxaceae (*T. baccata*), Cupressaceae (*C. japonica*) and Ginkgoaceae (*G. biloba*), allowed us to estimate a precise evolutionary timeline. In gymnosperms, some MADS-box genes are only expressed in reproductive organs, whereas most MADS-box genes, are expressed in both vegetative and reproductive organs [31]. This difference indicates that an increase in the number of MADS-box genes and the subsequent recruitment of some MADS-box genes as homeotic selector genes are important for the evolution of complex reproductive organs [32]. When selecting angiosperms, we included species from the three groups: (1) basal angiosperm (*A. trichopoda*) (2) monocots (*M. acuminata*, *O. sativa*, *Z. mays*, and *P. aphrodite*) (3) magnoliopsida and eudicots. Since magnoliopsida and eudicots is the largest group of angiosperm, we chose to include 14 typical species from the different families in this group, so that they would be useful for validating the evolutionary timeline. We considered choosing these seed plants (gymnosperms and angiosperms) for complete gene evolution of plants, which is of critical importance for the phylogenetic analysis. In related studies, bryophytes and seedless vascular plants do not have *ABCDE* or *AGL6* genes but have MADS-box genes [33, 34].

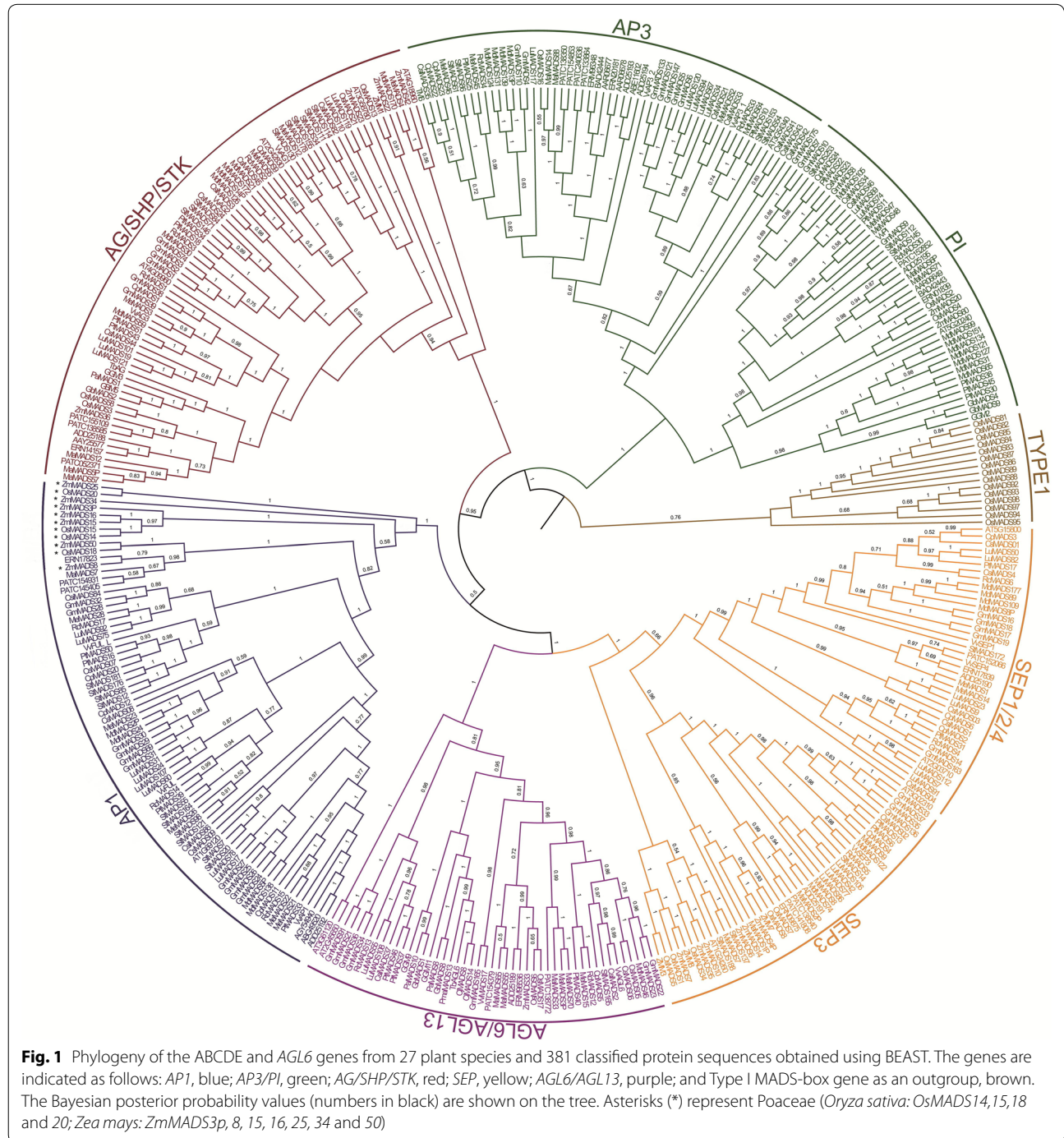
Many studies have examined the origin of type II MADS-box genes accompanying the divergence of major plant lineages [35], some of which suggest that the type II MADS-box gene clades originated about 300 to 400 million years ago (MYA) [15, 35–38]. Molecular clock-based dating methods deduced that the *B* and *C* gene lineages originated 660 and 570 MYA respectively [39, 40], a period before the separation of the lineages that led to mosses, ferns, and seed plants. Alternatively, the type II MADS-box genes in the lineage that led to extant ferns may have evolved faster than those in the seed plant lineage, such that orthology between genes from ferns and seed plants can no longer be recognized [35]. Previous works suggest that the *B* gene was the first *ABCDE* and *AGL6* genes to emerge [15, 35–38] but there are no mentions about the probable origin time of *ACDE* and *AGL6* genes. Clarifying the probable origin time of *ABCDE* and *AGL6* genes is a great help for understanding the role of the formation of the flower, which can decipher the forming order of MADS-box genes in the future. In this study, we collected *ABCDE* and *AGL6* 381 protein sequences and 361 coding sequences from gymnosperms and angiosperms, in order to understand the evolutionary history of the *ABCDE* and *AGL6* genes.

**Results**

**Identification of 381 ABCDE and AGL6 genes**

To examine the evolutionary history of ABCDE and AGL6 genes, we retrieved 381 sequences (Fig. 1, Table 1, Additional files 1, 2) from databases using known ABCDE and AGL6 protein sequences from *A. thaliana* and rice (*O. sativa*) as well as tomato MADS-box gene 6

(*TM6*) of *S. lycopersicum* as query sequences [2, 4, 6, 12, 29, 38, 41, 42] (Additional files 1, 2) in a BLAST search [43]. To verify the identities of the retrieved sequences before BLAST analyses, sequences were entered into the SMART to confirm the presence of basic MADS-box gene domains [44]. *AGL32* (B-sister genes) constitute a clade with a close relationship to class B genes [45].



**Table 1 Protein sequences of 27 plant species**

No	Species	AP1	AP3/PI	AG/SHP/STK	SEP	AGL6
1	<i>Gnetum gnemon</i> (GGM)	–	+	+	–	+
2	<i>Picea abies</i> (Pa)	–	–	+	–	+
3	<i>Podocarpus macrophyllus</i> (Pma)	–	–	–	–	+
4	<i>Wollemia nobilis</i> (Wn)	–	–	–	–	–
5	<i>Sciadopitys verticillata</i> (Sv)	–	–	–	–	–
6	<i>Taxus baccata</i> (Tb)	–	–	+	–	+
7	<i>Cryptomeria japonica</i> (Cj)	–	–	–	–	+
8	<i>Ginkgo biloba</i> (Gb)	–	+	+	–	+
9	<i>Amborella trichopoda</i> (ERN, AAR, BAD, AAY, ERM)	+	+	+	+	+
10	<i>Musa accuminata</i> (Ma)	+	+	+	+	+
11	<i>Oryza sativa</i> (Os)	+	+	+	+	+
12	<i>Zea mays</i> (Zm)	+	+	+	+	+
13	<i>Phalaenopsis aphrodite</i> (PATC)	+	+	+	+	+
14	<i>Solanum lycopersicum</i> (TM)	N	+	N	N	N
15	<i>Solanum tuberosum</i> (St)	+	+	+	+	+
16	<i>Vitis vinifera</i> (Vv)	+	+	+	+	+
17	<i>Citrus sinensis</i> (Csi)	+	+	+	+	+
18	<i>Carica papaya</i> (Cp)	+	+	+	+	+
19	<i>Malus domestica</i> (Md)	+	+	+	+	+
20	<i>Arabidopsis thaliana</i> (AT)	+	+	+	+	+
21	<i>Cucumis sativus</i> (Cs)	+	+	+	+	+
22	<i>Glycine max</i> (Gm)	+	+	+	+	+
23	<i>Nelumbo nucifer</i> (ADD, AGY, ABG, ABE)	+	+	+	+	+
24	<i>Populus trichocarpa</i> (Pt)	+	+	+	+	+
25	<i>Linum usitatissimum</i> (Lu)	+	+	+	+	+
26	<i>Ricinus communis</i> (Rc)	+	+	+	+	+
27	<i>Manihot esculent</i> (Me)	+	+	+	+	+

The “N” means uncollected related sequence, (+) means have protein sequence and (–) means unrelated protein sequence in this study. Gymnosperms: number 1–8. Angiosperms: number 9–27

Moreover, the B-sister and B genes arose 300–400 million years ago [45]. Therefore, we did not separate the B-sister and B genes in this study. The qualified sequences were aligned and included in the phylogenetic analyses. Sequences were arranged into subgroups according to the Bayesian phylogenetic tree in Fig. 1.

#### Phylogenetic analysis of the ABCDE and AGL6 genes

To depict the phylogenetic relationship among these 381 sequences, these genes were analyzed using Bayesian methods (Fig. 1). In previous studies, phylogenetic analysis of MADS-box genes in *Arabidopsis* and tomato was performed using the Bayesian methods for applied research [4, 46, 47]. In the present study, we used Bayesian method phylogenetic trees to sort individual sequences into subgroups (Fig. 1). The Bayesian method implemented in the Bayesian evolutionary analysis by sampling trees (BEAST) program was used to construct the phylogenetic tree (Fig. 1) representing the evolutionary relationship among all of the ABCDE and AGL6 gene

sequences, and to estimate the age of the ancestral node for each subgroup. Bayesian methods allow complex models of sequence evolution to be implemented [48]. According to Zhao et al. [49] the phylogenetic tree showing the relationships for the different functional gene clades of the MADS-box gene family ABCDE and AGL6 genes is the major clades of MIKC<sup>c</sup>-type group. In this study, our first aim was to clarify the origin of ABCDE and AGL6 genes.

#### Variations in the number of ABCDE and AGL6 genes in seed plants

The 381 ABCDE and AGL6 sequences from 27 seed plants clustered into five subgroups: *APETALA1* (AP1 or A gene, 74), *AP3/PISTILLATA* (AP3/PI or B genes, 101), *AG/SHATTERPROOF/SEEDSTICK* (AG/SHP/STK or CD genes, 75), *SEPALLATA* (SEP or E gene, 83), and *AGL6/AGL13* (AGL 6 gene, 48) (Fig. 1, Additional files 1, 2). The highest number of ABCDE and AGL6 genes in a flowering plant genome was observed in soybean

(*Glycine max*) (45) and the highest number among the gymnosperms was observed in *G. biloba* (6). The flowering plant *N. nucifera* had the fewest ABCDE and *AGL6* sequences (11). The A/E/*AGL6* MADS-box genes formed a monophyletic clade (posterior probability [PP]=0.5) that was larger (205) than the B (*AP3/PI*, 101) and CD (*AG/SHP/STK*, 75) clades (Fig. 1, Additional file 1).

#### Evolutionary patterns of ABCDE and *AGL6* genes in plants

Previous works suggest that the B gene (*AP3/PI*) was the first ABCDE and *AGL6* genes to emerge [15, 35–38] (Fig. 1). Our results show that plants that arose since gymnosperms appeared approximately 305 MYA [50] have both B/CD and *AGL6* genes (Table 1). Moreover, the B-sister and B genes arose 300–400 million years ago [45]. Therefore, we propose that the reasonable time of the B gene (*AP3/PI*) originated about 300 to 400 MYA. Kishino et al. [51] have proposed Bayesian methods of estimating the dates associated with branch points in a phylogenetic tree. Using the BEAST program, we set the origin of the B gene (*AP3/PI*) to about 350 MYA, and used this as a calibration point to estimate the appearance times of the ACDE and *AGL6* genes. In this study, we use B gene as the arising standard, which is sound and is expected to yield accurate information, and use BEAST for estimating the possible arising time is feasible. We are hopeful that using the origin time of a specific gene will accurately predict the origin time of other genes. With the comprehensive analysis, it is critical importance of the time of evolution for ABCDE and *AGL6* genes.

#### *AP1* patterns

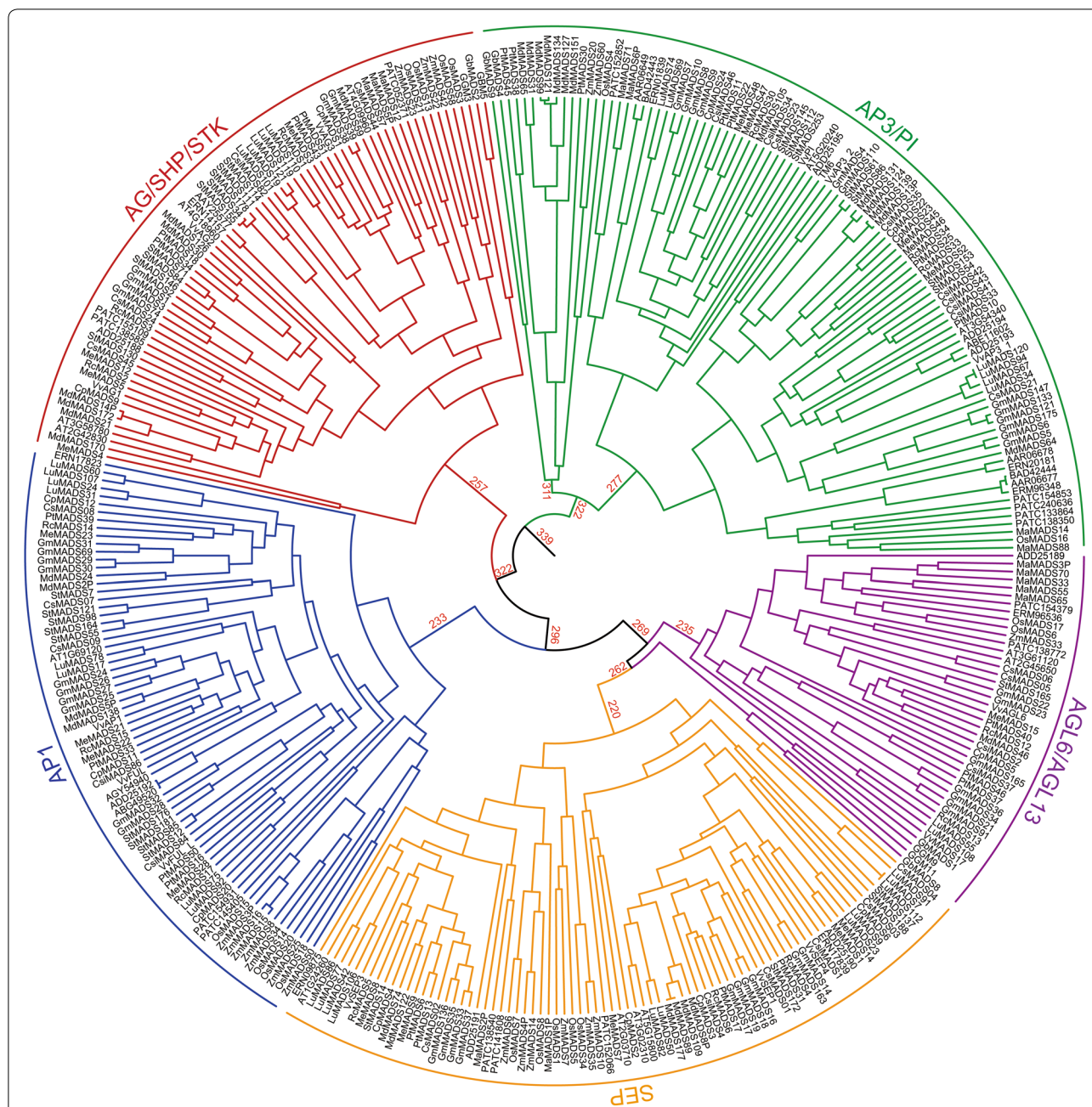
A-class genes are associated with sepal and petal development [17]. We found that only angiosperms possessed *AP1* genes (Table 1). According to our phylogenetic study (Fig. 1), the ancestral *AP1* diverged into one group. In monocots, the *AP1* genes seem to have undergone several duplication events. One duplication event appears to have occurred after the divergence of Poaceae (*O. sativa* and *Z. mays*, Fig. 1, Asterisks\*) from the other monocots, resulting in the duplicates *OsMADS18/20* (Fig. 1) and *OsMADS14/15* (Fig. 1, Additional file 3). The highest number of *AP1* was observed in *S. tuberosum* and *G. max* (Additional file 1). These results suggest that *AP1* replicated frequently in higher angiosperms and the restriction of MADS-box gene expression to specific reproductive organs and the specialization of MADS-box genes as homeotic genes in angiosperms were crucial aspects of floral organ evolution. Consistent with previous reports [23, 27, 52], the *AP1* gene has not been observed in gymnosperms (Table 1). Since there is more completed genome data and in our research there are comprehensive sequence collections, we have newly

discovered the sequences: *ZmMADS16* and *ZmMADS25* were in the *AP1* clade (Additional file 1, Asterisks\*), which consistent with the findings of previous *AP1* genes studies [5, 18, 53–56].

The clustering of *ERN17823* (Fig. 1) from the basal angiosperm *A. trichopoda* and *ZmMADS8* (Fig. 1) from *Z. mays* (PP=0.79) indicates that *AP1* emerged before the divergence of *A. trichopoda* and then developed into the *AP1* of *Z. mays* with the fewest changes. It remains unknown why these genes have undergone large expansions in *S. tuberosum* and *G. max* (Additional file 1). After a gene duplication, selection pressure favors gene retention only if the loss of a gene reduces the fitness of the organism [38]. Regardless, many duplicated genes appear to be redundant, since their loss-of-function mutants do not result in any detectable deviations in phenotype; however, there are known cases in which purifying selection constrains the divergence between redundant genes [38]. To trace the possible time when the ABCDE and *AGL6* genes emerged, we used the BEAST tool and set the origin time of the B gene (*AP3/PI*) at 350 MYA to obtain a mutation rate estimate of  $2.617 \times 10^{-3}$  substitutions/site/million years for ABCDE and *AGL6* genes. We found that A gene (*AP1*) shared an MRCA with *SEP/AGL6/AGL13* 296 MYA and diverged into its own lineage 233 MYA (Fig. 2).

#### *AP3/PI* patterns

B-class genes play an important role in petal and stamen development [37]. Furthermore, the paleo*AP3/DEF* lineage produced two additional lineages in the eudicots known as eu*AP3* and *TM6* [22]. *Malus domestica* had the highest number of *AP3/PI* genes (13 sequences, Additional file 1). Angiosperms have more *AP3/PI* genes than gymnosperms, which may help form more complex reproductive organs. Unlike the variable number of *AP1* genes among dicots and monocots, there were no obvious difference in the number of *AP3/PI* genes between the monocots and dicots, although they have distinctly different second whorl structures (lodicule vs. petal) [15, 28]. In Fig. 1, we found that most *AP3* genes located in a single cluster comprising homologs of both eudicot and monocot and the *A. trichopoda* [49]. In the study, we have newly discovered the sequence: *PtMADS25* was in the *AP3/PI* clade (Additional file 1, Asterisks\*), which consistent with the findings of previous *AP3/PI* genes studies [5, 18, 34, 52–57]. Among the 381 sequences from the 27 seed plants examined, the highest number of ABCDE and *AGL6* genes was observed in the *AP3/PI* with a total number of 101 (Additional file 1), suggesting that numerous *AP3/PI* gene expansions contribute to the evolution of reproductive organs. The MADS-box genes appear to have evolved mainly through gene duplication



**Fig. 2** Phylogeny of the ABCDE and AGL6 genes from 27 plant species and 361 classified coding sequences obtained using BEAST. The genes are colored as follows: AP1, blue; AP3/PI, green; AG/SHP/STK, red; SEP, yellow; and AGL6/AGL13, purple. The red values represent the probable origin and divergence time (million years ago, MYA) as calculated using BEAST of the AP1, AP3/PI, AG/SHP/STK, SEP, and AGL6/AGL13 genes in the tree. AP3/PI originated 339 MYA; AG/SHP/STK originated 322 MYA; and AP1, SEP, and AGL6/AGL13 originated 296 MYA. The divergence of SEP and AGL6/AGL13 occurred 269 MYA, and the divergence of AP1 occurred 233 MYA

events, followed by neofunctionalization and subfunctionalization, or in certain cases, the pseudogenization of the duplicated gene [6].

Different to the distinct evolution pattern of AP3 and PI, independent duplications of the B genes are being discovered in monocot and dicot species. These specific

duplications are predicted to be associated with morphological innovation such as the highly derived petals of the Orchid family [58]. In dicots, some species have only a single PI gene, such as *C. papaya* (*CpMADS24*), *C. sativus* (*CsMADS23*), *R. communis* (*RcMADS30*) and *Vitis vinifera* (*VvPI*) (Fig. 1). Although the evolutionary

patterns of the *PI* clade do not resemble the patterns of species expansion, *PI* genes from *G. gnemon*, *G. biloba*, *M. domestica*, and *P. trichocarpa* belong to one clade (PP=0.98; Additional file 4). One possible explanation for this possibility is that the ancestors of gymnosperms, *G. gnemon* and *G. biloba* possessed the *PI* gene, whereas the other species lose this homolog during the evolutionary process. Another possibility is that horizontal gene transfer might have occurred, in which microorganisms or insects transferred the *PI* gene from *G. gnemon* or *G. biloba* into *M. domestica* and *P. trichocarpa*. This process can occur between closely related eukaryotic species [59], can mediate the massive transfer of chloroplast–nuclear genes [60], and the inter-species movement of chloroplasts under stress [61]. Evidence of this may be found in the *PI* gene of *A. trichopoda* (AAR06649, BAD42443 and ERN01839) (Fig. 1), a close relative of *M. accuminata*, *N. nucifera* and *P. aphrodite* (PP=1; Additional file 4). Previously, the euAP3 lineage was a divergent paralogous group, only found in higher eudicots [62]. This study shows that AP3 lineage formed a monophyletic clade in monocot (PP=1; Fig. 1; *OsMADS16*, *MaMADS14*, *MaMADS88*, *PATC138350*, *PATC154853*, *PATC240636*, and *PATC133864*). Following examples set by other studies [15, 36–39], we used the BEAST tool to estimate when the B gene (*AP3/PI*) clade originated. We found that it arose 339 MYA and has a mutation rate of  $2.617 \times 10^{-3}$  substitutions/site/million years (Fig. 2), making it the first gene clade to have evolved that is involved in reproductive structure development (Fig. 2). Since gymnosperms appeared approximately 305 MYA [50], the *AP3/PI* genes might have appeared in the phase of evolution between seedless vascular plants and gymnosperms.

#### *AG/SHP/STK patterns*

CD-class genes are associated with stamen and carpel development [17]. As displayed in Fig. 1, the *AG/SHP/STK* genes formed a single clade (PP=1). By contrast, the five gymnosperm genes *TbAG*, *GGM3*, *PaMADS1*, *GBM5*, and *GbMADS2* (Fig. 1) formed a well-supported clade (PP=1; Additional file 5). Among the flowering plants, the highest number of *AG/SHP/STK* genes (9) was observed in *S. tuberosum*, whereas the lowest was observed in *N. nucifera* (1) (Additional file 1). Because of more completed genome data, we have newly discovered the sequences: *CsMADS24*, *CsMADS44*, *CsMADS45*, *GbMADS2* and *PtMADS34* were located in the *AG/SHP/STK* clade (Additional file 1, Asterisks\*). Our research is consistent with the findings of previous *AG/SHP/STK* genes studies [5, 18, 24, 26, 34, 52–56, 63–65].

BEAST analysis set the origin of the B genes (*AP3/PI*) to 350 MYA and yielded a mutation rate estimate of  $2.617 \times 10^{-3}$  substitutions/site/million years for the

ABCDE and *AGL6* genes. Based on these data, we found that the CD (*AG/SHP/STK*) originated 322 MYA, shortly after the appearance of B gene (Fig. 2).

#### *SEP patterns*

E-class genes are associated with the formation of all floral organ types during reproductive development [17]. The *SEP* has been isolated from a few plants and its homologs in *Arabidopsis* include *SEP1*, *SEP2*, *SEP3*, and *SEP4*, [66]. Some analyses place *SEP1* and *SEP2* closer to *SEP4* than to *SEP3* [66] (Fig. 1), whereas other studies conclude that *SEP3* is the closest relative of *SEP1* and *SEP2* [15, 67, 68]. We found that only angiosperms possessed *SEP*, and that these genes clustered into two groups: *SEP1/2/4* and *SEP3* (Fig. 1). Our results also suggest that *SEP1* and *SEP2* are more closely related to *SEP4* than to *SEP3*. *SEP3* formed a monophyletic clade in monocot [69]. However, the two *SEP3* genes *PATC141808* and *PATC138540* of *P. aphrodite* unexpectedly fell outside of this clade (Fig. 1). *SEP3* appears to have diverged more in the monocots than in the eudicots. In Fig. 1, most eudicot and monocot *SEP3* genes group as a distinct cluster [49]. In monocots, the *SEP3* lineage has undergone several duplication events (Fig. 1). One duplication event appears to have occurred after the divergence of Poaceae (*O. sativa* and *Z. mays*) from the remaining monocots, resulting in the duplicates *OsMADS7* and *OsMADS8* (Fig. 1). Our sampling was insufficient to determine whether this duplication is specific to the Poaceae or to all of the Poales [69]. This finding shows that the *SEP* genes of *S. tuberosum* and *M. esculenta* (*StMADS137*, *StMADS188* and *MeMADS7*) (Fig. 1) are closely related to the *SEP* genes in monocots. The *SEP1/2/4* of angiosperms clustered in a single clade (PP=1; Additional file 6). In Fig. 1, some eudicot species (e.g. *G. max* and *M. domestica*) had several copies that formed species-specific clades that reside inside a well-supported *SEP1/2/4* clade (PP=0.94; Additional file 6). The highest number of *SEPs* (11) was observed in *Linum usitatissimum* (Additional file 1). The flowering plants *A. trichopoda*, *M. accuminata*, and *N. nucifera* had the lowest number of *SEP* (Additional file 1); *SEPs* from these species underwent fewer duplications, implying that the E and ABCD genes are less involved in flower development in these species. The finding that basal angiosperms and monocots *M. accuminata* had less E gene expansion than did the other plants examined may indicate that the restriction of MADS-box gene expression to specific reproductive organs and the specialization of the MADS-box gene in the flowering plant lineage were crucial events in floral evolution. We have newly discovered the sequences: *ZmMADS6* and *ZmMADS7* were in the *SEP* clade (Additional file 1, Asterisks\*), which consistent

with the findings of previous *SEP* genes studies [5, 53, 54, 70].

The ABCDE and *AGL6* genes had an estimated mutation rate of  $2.617 \times 10^{-3}$  substitutions/site/million years using BEAST analysis set the origin time of the B genes (AP3/PI) at 350 MYA. We found that *SEP* shared an MRCA with *AGL6/13* and *API* genes 296 MYA, and with *AGL6/13* 269 MYA (Fig. 2).

#### *AGL6/AGL13* patterns

The *AGL6*-like genes are associated with floral development in angiosperms [29] and with cone formation in gymnosperms [71]. The *AGL6*-like genes in monocots and eudicots play essential roles in floral development [41, 72]. The *Arabidopsis* genome contains two *AGL6* genes, namely *AGL6* and *AGL13* [66], suggesting a potential functional redundancy between these two genes. Schauer et al. [47] argued that *AGL6* and *AGL13* exhibit signs of subfunctionalization, with different expression patterns, regulatory sequences, and possible functions. *AGL6/13* and *SEP* genes have a high degree of sequence similarity and form sister clades in phylogenetic trees [30] (Fig. 1). As displayed in Fig. 1, the *AGL6/13* is categorized into one class (*AGL6/13*) in which genes of gymnosperms formed a well-supported clade (PP=1; Additional file 7). Among the flowering plants, the highest number of *AGL6/13*s (7) was observed in *G. max*. Contrary to *SEP* genes which are only present in angiosperms, *AGL6/13* genes are ancient and widely distributed in gymnosperms and angiosperms (Additional file 1). In our research, there are comprehensive sequence collections, we have newly discovered the sequences: *CjMADS8*, *GmMADS91*, *GmMADS165*, *PaMADS10*, *PtMADS37*, *PtMADS 46* and *VvMADS17* were placed in the *AGL6/13* clade (Additional file 1, Asterisks\*). Consistent with the findings of previous *AGL6/13* genes studies [5, 18, 34, 52–55, 63–65, 70–73].

In this study, the *P. abies* gene *PaMADS8* was placed in the *AGL6/AGL13* subfamily (Fig. 1). *PaMADS8* (*DAL1*) was predicted to play a role in the transition from juvenile to adult plant (including the transition from reproductively incompetent to competent) [74]. This proposal was chiefly based on the observed expression pattern of *PaMADS8*; expression increased with the age of the tree and with the consecutive development of the vegetative structures within the tree. For instance, the relative expression of *PaMADS8* is highest in vegetative shoots in the apical part of the tree [74]. Both *PaMADS8* and other *AGL6* genes in gymnosperms are active in cones [52, 63]. These results revealed that *AGL6* gene redundancy and functional diversity also exist in gymnosperms. Using BEAST analysis and set the origin time of B genes (AP3/PI) at 350 MYA, and a mutation rate estimate of

$2.617 \times 10^{-3}$  substitutions/site/million years for ABCDE and *AGL6* genes. These results suggest that the *AGL6* family shared an MRCA with *SEP* and *API* genes 296 MYA (Fig. 2).

All calculations were implemented using codeml at PAML4.9. Different models were specified according to the software instruction. “np” refers to the number of parameters, “l=(ln L)” refers to the log value of the likelihood. The estimated parameters w refer to the dN/dS ratio. In the one-ratio model M0 and the Branch-specific two-ratio models, w (A), w (B), w (CD), w (E), and w (*AGL6*) stand for the w ratios in the 27 plant species.

#### Natural selection analysis

The assessment of synonymous (syn) and non-synonymous (non-syn) substitution ratios is important for understanding molecular evolution at the amino acid level [75]. To examine the intensity of natural selection acting on a specific clade, we examined the ratio (w) of non-syn substitutions to syn substitutions in our ABCDE and *AGL6* phylogeny. In this analysis, w < 1, w = 1, and w > 1 indicated purifying selection, neutral evolution, and positive selection, respectively. Based on our phylogeny, w assessments were conducted for five branches (w (A), w (B), w (CD), w (E), and w (*AGL6*) respectively). First, the branch-specific likelihood model [76] was applied to the ABCDE and *AGL6* data. As shown in Table 2, the one-ratio model revealed a w value of 0.29953, which is well below 1. This indicates a strong purifying selection pressure on the entire MADS-box gene family [77]. MADS-box proteins may be responsible for simultaneous increases in the ratio of nonsynonymous to synonymous substitutions early in angiosperm history and following concerted duplication events [77]. In contrast to the patterns of positive selection of *AP3/PI* reported by Hernandez–Hernandez et al. [37], however, we did not detect positive selection on ABCDE and *AGL6* genes within MADS box subfamilies. We used 361 classified coding sequences for our analysis of natural selection, and this would not affect the real relationships among these subfamilies as determined using ABCDE and *AGL6* genes to establish the phylogenetic tree. In the study, our results indicate that purifying selection has played an important role in the evolution of these MADS box gene subfamilies throughout seed plant history [77]. For the two-ratio model, when A, B, CD, E and *AGL6* genes were set as the out-branch, no significant differences were detected for any of the target genes ( $2\Delta l = 0$  or  $2\Delta l = 1.6$ ,  $p = 0.2059$ ,  $df = 1$ ), suggesting that the evolutionary rates of A, B, CD, E, and *AGL6* genes are not significantly different from each other. We have also tested the multiple ratio models including the five-ratio model w (A) ≠ w (B) ≠ w (CD) ≠ w (E) ≠ w (*AGL6*). The results showed that this model is



**Table 2** Natural selection test of plant MADS-box genes

Model	Np	$l = \ln L$	Estimates of parameters
M0: one-ratio $w(A) = w(B) = w(CD) = w(E) = w(AGL6)$	1	- 57551.8	$w(A) = w(B) = w(CD) = w(E) = w(AGL6) = 0.29953$
Branch-specific two-ratio models			
$w(B) = w(CD) = w(E) = w(AGL6) \neq w(A)$	2	- 57551.8	$w(B) = w(CD) = w(E) = w(AGL6) = 0.29953$ $w(A) = 2.13680$
$w(A) = w(CD) = w(E) = w(AGL6) \neq w(B)$	2	- 57551.0	$w(A) = w(CD) = w(E) = w(AGL6) = 0.29920$ $w(B) = 301.74451$
$w(A) = w(B) = w(E) = w(AGL6) \neq w(CD)$	2	- 57551.8	$w(A) = w(B) = w(E) = w(AGL6) = 0.29953$ $w(CD) = 2.12982$
$w(A) = w(B) = w(CD) = w(AGL6) \neq w(E)$	2	- 57551.8	$w(A) = w(B) = w(CD) = w(AGL6) = 0.29954$ $w(E) = 2.12305$
$w(A) = w(B) = w(CD) = w(E) \neq w(AGL6)$	2	- 57551.8	$w(A) = w(B) = w(CD) = w(E) = 0.29953$ $w(AGL6) = 2.11723$
$w(B) = w(CD) \neq w(AGL6) = w(A) = w(E)$	2	- 55162.3	$w(AGL6) = w(A) = w(E) = 0.30546$ $w(B) = w(CD) = 0.00010$
Five-ratio models			
$w(A) \neq w(B) \neq w(CD) \neq w(E) \neq w(AGL6)$	5	- 55190.1	$w(A) = 0.30530, w(B) = 1.74205, w(CD) = 0.78636,$ $w(E) = 0.64375, w(AGL6) = 1.69816$

not better than the two-ratio model  $w(B) = w(CD) \neq w(AGL6) = w(A) = w(E)$ . Our analyses support B and CD have undergone significantly different selection pressure from A, E and AGL6. Regarding the biased sampling of the MIKC<sup>c</sup>-type genes, since the dN/dS ratio is a pairwise characteristic, the w values we calculated for each branch represent the average dN/dS value for each specified branch. In our analyses, we specified CD genes altogether as a single branch and calculated their average dN/dS value. The current sampling of the MIKC<sup>c</sup>-type genes is sufficient and representative for the assessment of the selection pressure in this branch.

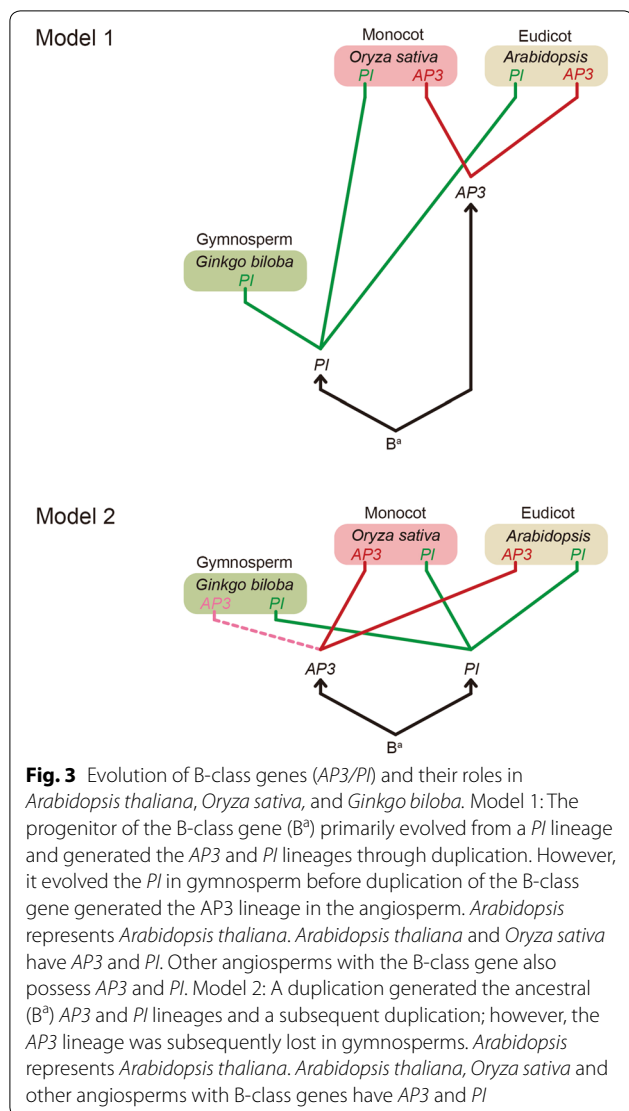
## Discussion

**B-class genes (*AP3/PI*) two possible evolutionary pathways**  
Bryophytes and seedless vascular plants do not have ABCDE or AGL6 genes [33, 34]. The gymnosperms have *PI* genes (*GGM2*, *GbMADS4* and *GbMADS9*; Fig. 1) but no *AP3* gene. Our study suggests that the *PI* clade probably evolved earlier than the *AP3* clade, and that the formation of gymnosperm cones depended on the presence of a *PI* ancestor. Therefore, the phylogeny of B-class genes lead us to infer two possible evolutionary pathways (Fig. 3). (1) The progenitor of the B gene (B<sup>a</sup>) first evolved through the *PI* lineage and then generated the *AP3* and *PI* lineages, since only the *PI* lineage was maintained in gymnosperms before the duplication of the B gene generated the *AP3* lineage in angiosperm. (2) An ancient duplication may have generated the ancestral (B<sup>a</sup>) *AP3* and *PI* lineages, and the *AP3* lineage was lost in gymnosperms after a subsequent duplication. Since the initial

duplication that generated the paralogous *PI* and *AP3* lineages predates the monocot and dicot division, monocots have both clades [58]. Therefore, a more complete collection of sequences in diverse species would provide a clearer understanding of B gene (*AP3/PI*) duplication.

### *AP3/PI* and *AG/SHP/STK* evolved earlier

Previously, Kim et al. [78] assumed that the B gene evolved relatively earlier than other flower identity genes. *AG*, *AGL6* and *DEF+GLO* (B genes) were present in the MRCA of angiosperms and gymnosperms approximately in 300 MYA [15]. Although some ancestral genes might have specialized 300 MYA in the development of male reproductive organs (*DEF/GLO*-like genes), female reproductive organs (*GGM13*-like genes) [79] or both (*AG* gene, *AGL2*, and *AGL6*-like genes), all of the MADS-box gene types were highly diversified before the establishment of the ovule approximately 300–400 MYA [36]. In this study, gymnosperms possessed *AP3/PI*. We estimated that *AP3/PI* originated 339 MYA (Fig. 2). Hence, we suggest that *AP3/PI* evolved before the appearance of gymnosperms, which in turn appeared approximately 305 MYA [50]. *AP3/PI* may have an ancestral function that is realized in extant gymnosperms in distinguishing male cones (form when the B gene is expressed) from female cones (form when the B gene is not expressed) [63]. The B gene is involved in the development of petals and stamens in angiosperms and male cones in gymnosperms [63, 80]. Kim et al. [78] also show that *AP3/PI* duplication occurred shortly after the divergence of extant gymnosperms and angiosperms, which is accordingly before the



age of the oldest flowering plant fossils. This implies that the joint expression of *AP3* and *PI* may not have resulted in the immediate formation of petals, which they presently control in the development of extant angiosperms. Therefore, the earliest angiosperms may have been biochemically flexible in their B gene function [78].

We estimate that *AG/SHP/STK* (CD genes) evolved 322 MYA, shortly after the appearance of *AP3/PI* (B genes) (Fig. 2). The C gene has a single function in reproductive organ development and the mechanisms controlling its expression domain and evolution were key factors in the emergence of flowering plants [81]. Jager et al. [64] show that *GBM5* (CD gene; Additional file 1) is expressed in the early stages of developing male and female organs, and persists in the female gametophyte. Parallel expression patterns have been detected for the orthologues of *GBM5* in coniferophytes: *DAL2* (*PaMADS1*, CD gene;

Additional file 1) in *P. abies* and *SAG1* in *Picea mariana*. *DAL2* and *SAG1* are expressed in male and female cones, but a gradual diminution was observed during the maturation of male cone, whereas female cones experience the development which maintained a great level of expression in respect of ovule maturation [82]. In contrast to this, the expression of *GGM3* (CD gene; Additional file 1) from *G. gnemon* persists in both male and female reproductive units in the late developmental stages [63]. In gymnosperms, some MADS-box genes are only expressed in reproductive organs, whereas most MADS-box genes, are expressed in both vegetative and reproductive organs [31]. This difference indicates that an increase in the number of MADS-box genes and the subsequent recruitment of some MADS-box genes as homeotic selector genes are important for the evolution of complex reproductive organs [32]. The expansion of the MIKC gene family in seed plants and increased plant complexity seem to be correlated [83]. Hence, CD genes (*AG/SHP/STK*) appeared to have evolved soon after the B genes (*AP3/PI*), and their emergence promoted reproduction in plants.

In different gymnosperms, *AG*-like genes are expressed in male and female reproductive organs, which may represent the ancestral state of gene expression [63]. These genes were suggested to function ancestrally in male and female cone formation, and in distinguishing them from the nonreproductive organs [63]. Several angiosperm *AG*-like genes probably have an ancestral function in specifying both male and female reproductive organs and have derived functions that are restricted to the stamen or pistil [84]. In male cones, microsporangia, which contain the pollen, develop at the base of the microsporophylls. By contrast, in female cones, uncovered ovules develop on the surface of megasporophylls, instead of being enclosed in a gynoecium [20]. Sporophylls are modified leaf-like organs that are the gymnosperm structures most closely related to carpels [20]. Therefore, angiosperm flowers and gymnosperm cones are homologous. Gymnosperms express B and CD genes, and the wide distribution of these genes throughout the gymnosperms shows that these genes were present when the gymnosperms first appeared. Since the B and CD genes are responsible for the formation of reproductive organs, the B and CD genes may have evolved before the A/E/AGL6 superclade (Fig. 2).

#### The A/E/AGL6 superclade evolved soon after *AG/SHP/STK*

In our study, two *G. gnemon* genes (*GGM9* and *GGM11*) were placed in the *AGL6/AGL13* subfamily (Additional file 1). Both of these *G. gnemon* genes are expressed in both male and female reproductive cones, but not in vegetative leaves [63]. Katahata et al. [71]

showed that *CjMADS14* (*AGL6*; Additional file 1) of *C. japonica* was expressed chiefly in male and female strobili. Together, these and other results suggest that *AGL6* is associated with reproduction [85, 86] and cone formation [71]. Our phylogenetic analysis revealed that *AGL6* family members are closely related to *SEP* with the MRCA occurring 296 MYA (Fig. 2). *Arabidopsis* *SEP* and *AGL6* genes were found to activate the expression of B and C genes [85]. Moreover, no *SEP* was found in gymnosperms (Table 1) [25]. *AGL6* does not directly influence floral structures; however, it is critical for the reproductive abilities of both gymnosperms and angiosperms. Thus, we propose that *AGL6* may have evolved after the formation of certain essential reproductive organs (e.g. flowers and cones) to aid in the formation of more complete reproductive structures in plants. *AGL6* genes may have played an important role in the evolution of unique flower features [86]. Li et al. [72] proposed that both *SEP* and *API* in angiosperms were derived from the common antecedent of *AGL6* within two duplication events, and another duplication event of *AGL6* genes likely arose before the derivation of grass. These findings suggest that *AGL6* genes may act in an ancient and conserved floral development pathway. Comparative analyses of spatiotemporal expression patterns of *AGL6* or genetic analyses on mutants are warranted to elucidate the functional redundancy of *AGL6* in lateral organ development and flowering [85]. Yoo et al. [87] showed that *AGL6* regulates the transcription of two critical flowering-time regulators: *FLC* and *FT*. Moreover, *AGL6* further enhanced *FT* expression in the absence of the *FLC* function, suggesting that *AGL6* regulates *FT* independently of *FLC* [87]. Thus, based on the concept of evolution, the plant must flower at an appropriate time, which implies that *AGL6* have emerged before *API*.

Wang and Melzer [88] suggested that the AG-like protein GGM3 (CD protein; Additional file 1) can form homotetramers and even more stable heterotetramers with the DEF/GLO-like protein GGM2 (B protein; Additional file 1). Therefore, the capacity of gymnosperm MADS-domain proteins to produce multimeric complexes is similar to their angiosperm counterparts. However, in contrast to angiosperms, multimeric complex formation does not depend on the E proteins orthologues, and *SEP* genes have not yet been identified in any gymnosperm [88]. Furthermore, the genomes of most gymnosperms examined in our study possess *AGL6*, but not *SEP* (Table 1). Consequently, *AGL6* and *SEP* genes estimated to have originated at a similar point in time (Fig. 2) but we hypothesize that *AGL6* evolved before the *SEP* genes. According to our phylogenetic analysis and hypothesis, the ABCDE and *AGL6* genes

may appear in the following order: *AP3/PI* → *AG/SHP/STK* → *AGL6* → *SEP* → *API*.

## Conclusions

We assembled a comprehensive dataset of ABCDE and *AGL6* genes of representative species from gymnosperms and angiosperms as well as used it to construct the phylogeny of plant ABCDE and *AGL6* genes. We have newly discovered the sequences: *API* (*ZmMADS16* and *ZmMADS25*); *AP3/PI* (*PtMADS25*); *AG/SHP/STK* (*CsMADS24*, *CsMADS44*, *CsMADS45*, *GbMADS2*, and *PtMADS34*); *SEP* (*ZmMADS6* and *ZmMADS7*); *AGL6/13* (*CjMADS8*, *GmMADS91*, *GmMADS165*, *PaMADS10*, *PtMADS37*, *PtMADS46* and *VvMADS17*) and these newly discovered the sequences are important for estimating the time of origin for ABCDE and *AGL6* genes, which can help to compensate the insufficient source of former researches. The phylogeny of ABCDE and *AGL6* genes subfamilies differed. The *API* subfamily clustered into one group. The *AP3/PI* subfamily clustered into two groups: the *AP3* and *PI* clades. The *AG/SHP/STK* subfamily clustered into a single group. The *SEP* subfamily in angiosperms clustered into two groups: the *SEP1/2/4* and *SEP3* clades. Finally, the *AGL6/13* subfamily clustered into a single group. The ABCDE and *AGL6* genes appeared in the following order: *AP3/PI* genes originated 339 MYA, *AG/SHP/STK* genes originated 322 MYA, *API* genes shared the MRCA with *AGL6/SEP* 296 MYA, and *AGL6/SEP* diverged in 269 MYA. Moreover, the phylogeny of B-class genes lead us to infer two possible evolutionary pathways. (1) The progenitor of the B gene ( $B^a$ ) first evolved through the *PI* lineage and then generated the *AP3* and *PI* lineages. (2) An ancient duplication may have generated the ancestral ( $B^a$ ) *AP3* and *PI* lineages, and the *AP3* lineage was lost in gymnosperms after a subsequent duplication. This study highlights important events in the evolutionary history of the ABCDE and *AGL6* gene families and clarifies their evolutionary path.

## Methods

### Identifying MADS-box sequences

To obtain sequences of model organisms, species-specific databases, including the *O. sativa* database (<http://rice.plantbiology.msu.edu/>), the *A. thaliana* database (<http://www.arabidopsis.org/>), and the *P. aphrodite* database (<http://orchidstra.abrc.sinica.edu.tw>). The sequences for *A. trichopoda* were obtained from the NCBI (<http://www.ncbi.nlm.nih.gov/>) and UNIPROT (<http://www.uniprot.org/uniprot/>). The sequences for angiosperms were obtained from the Gramene (<http://www.gramene.org/>) and Phytozome (<http://www.phytozome.net/>) databases. The sequences for gymnosperms were obtained

from NCBI, Phytozome, and UNIPROT. Some databases such as PANTHER (<http://www.pantherdb.org/>), PGDD (<http://chibba.agtec.uga.edu/duplication/>), and Ensembl Plants database (<https://plants.ensembl.org/index.html>) are references of this research. Known ABCDE and *AGL6* protein sequences from *A. thaliana* and *O. sativa* as well as the *TM6* sequence of tomato (*S. lycopersicum*) were used as the query sequences (Additional files 1, 2) [2, 4, 6, 12, 29, 38, 41, 42] for BLASTP [43]. We applied an E-value cutoff of less than  $10^{-10}$  for protein similarity.

### Confirming MADS-box sequences

First, all of the sequences obtained through BLAST were entered into SMART (<http://smart.embl-heidelberg.de/>) to confirm the presence of the basic domains from the MADS-box gene [44]. MIKC-type structures include M-domain that is followed by I, K, and C domains respectively [10, 11]. Subsequently, qualified sequences were aligned and subjected to phylogenetic tree analysis to determine their subfamily affiliations.

### Building the alignment and phylogenetic trees

The amino acid sequences were aligned using the program MULTIPLE Sequence Comparison by Log-Expectation (MUSCLE) in MEGA6. In addition, BEAST 2.2.1 was used to construct Bayesian phylogenies [4, 89]. The BEAST analysis was performed using a JTT substitution model and a Yule priors model. The stationary distribution of the MCMC chains and the convergence of runs were monitored using Tracer (v.1.6) to determine the appropriate MCMC chain length such that the effective sample size of every parameter was larger than 200 as recommended. Tree pictures were generated using TreeAnnotator (v. 1. 8. 2), with first 20000 trees discarded as burn-in. Trees were visualized using Figtree (v. 1. 4. 2).

### Natural selection analysis

Natural selective pressure on plant ABCDE and *AGL6* genes were examined by measuring the ratio of non-synonymous to synonymous substitutions ( $dN/dS=w$ ). Codon-based maximum likelihood estimates of  $w$  was performed using codeml in PAML4.9 [90]. Multiple-alignment of conserved domain sequences for those identified plant ABCDE and *AGL6* genes were carried out using ClustalW2 [91]. Significant insertions and gaps were removed manually. To facilitate the input data requirements of codeml, an additional Maximum Likelihood tree was constructed using a smaller data set where the ABCDE and *AGL6* genes with no identifiable conserved domain sequences were removed. The subtree including plant ABCDE and *AGL6* genes was used in codeml. Branch pattern specification was implemented using Treeview1.6.6 (<http://taxonomy.zoology.gla.ac.uk/>

[rod/treeview.html](#)). Five target clades were specified based on the present phylogenetic analysis: A, B, CD, E and *AGL6* genes. The  $w$  values for these clades were represented as  $w(A)$ ,  $w(B)$ ,  $w(CD)$ ,  $w(E)$ , and  $w(AGL6)$  respectively. Nested likelihood ratio tests were performed to assess the significance of the model under different hypotheses:  $w(B)=w(CD)=w(E)=w(AGL6)\neq w(A)$ ,  $w(A)=w(CD)=w(E)=w(AGL6)\neq w(B)$ ,  $w(A)=w(B)=w(E)=w(AGL6)\neq w(CD)$ ,  $w(A)=w(B)=w(CD)=w(AGL6)\neq w(E)$ ,  $w(A)=w(B)=w(CD)=w(E)\neq w(AGL6)$ , and  $w(B)=w(CD)\neq w(AGL6)=w(A)=w(E)$ . The corresponding  $p$  values were calculated using the online tool at <http://graphpad.com/quickcalcs/PValue1.cfm>.

### Additional files

**Additional file 1.** Sequence numbers of 27 plant species.

**Additional file 2.** Txt files of 27 species and 381 sequences.

**Additional file 3.** The *AP1* genes in Bayesian tree and the Bayesian posterior probability values in tree.

**Additional file 4.** The *AP3/PI* genes in Bayesian tree and the Bayesian posterior probability values in tree.

**Additional file 5.** The *AG/SHP/STK* genes in Bayesian tree and the Bayesian posterior probability values in tree.

**Additional file 6.** The *SEP* genes in Bayesian tree and the Bayesian posterior probability values in tree.

**Additional file 7.** The *AGL6/AGL13* genes in Bayesian tree and the Bayesian posterior probability values in tree.

### Abbreviations

AG: *AGAMOUS*; *AGL6*: *AGAMOUS-LIKE 6*; AP: *APETALA*; BEAST: Bayesian evolutionary analysis by sampling trees; BLAST: Basic Local Alignment Search Tool; C-domain: C-terminal domain; DEF: *DEFICIENS*; GLO: *GLBOSA*; I-domain: intervening domain; K-domain: keratin-like domain; M-domain: MADS-domain; *MCM1*: *minichromosome maintenance-1*; MRCA: most recent common ancestor; MUSCLE: Multiple Sequence Comparison by Log-Expectation; MYA: million years ago; NCBI: National Center for Biotechnology Information database; *Nenu*: *Nelumbo*; non-syn: non-synonymous; PANTHER: Protein ANalysis THrough Evolutionary Relationships; PGDD: Plant Genome Duplication Database; *PI*: *PISTILLATA*; *SEP*: *SEPALLATA*; *SHP*: *SHATTERPROOF*; SMART: Simple Modular Architecture Research Tool; *SRF*: *serum response factor*; *STK*: *SEEDSTICK*; syn: synonymous; TFs: transcription factors; *TM6*: tomato MADS-box gene 6.

### Authors' contributions

GS performed all the research and drafted the manuscript. CHY for her technical guidance and assistance in phylogenetic trees. CYS revised the manuscript and discussed the results. KSH supervised the study and revised the manuscript. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> Department of Electrical Engineering, I-Shou University, Kaohsiung, Taiwan. <sup>2</sup> The School of Chinese Medicine for Post-Baccalaureate, I-Shou University, Kaohsiung, Taiwan. <sup>3</sup> College of Medicine, I-Shou University, Kaohsiung, Taiwan.

### Acknowledgements

We thank Dr. Yong Jia (School of Agriculture, Food and Wine, the University of Adelaide) for his technical guidance and assistance in phylogenetic trees.

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

The sequences of 27 plant species were obtained from the *O. sativa* database (<http://rice.plantbiology.msu.edu/>), the *A. thaliana* database (<http://www.arabidopsis.org/>), the *P. aphrodite* database (<http://orchidstra.abrc.sinica.edu.tw/>), NCBI (<http://www.ncbi.nlm.nih.gov/>), UNIPROT (<http://www.uniprot.org/uniprot/>), Gramene (<http://www.gramene.org/>), Phytozome (<http://www.phytozome.net/>), PANTHER (<http://www.pantherdb.org/>), PGDD (<http://chibba.agtec.uga.edu/duplication/>), and Ensembl Plants database (<https://plants.ensembl.org/index.html>). The MEME tool ([meme.sdsc.edu/meme/meme-intro.html](http://meme.sdsc.edu/meme/meme-intro.html)) was used to identify the conserved motifs in the MADS-box proteins from the 27 species. MEGA; PAML; BEAST; TreeAnnotator; Figtree (Software).

**Consent for publication**

Not applicable.

**Ethics approval and consent to participate**

Not applicable.

**Funding**

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Publisher's Note**

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

Received: 20 December 2018 Accepted: 18 April 2019

Published online: 24 April 2019

**References**

- 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*. 2013;2:379–95.
- Krizek BA, Fletcher JC. Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet*. 2005;6:688–98.
- Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L. 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, Kapoor S. 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*. 2013;201:717–32.
- Jiang SC, Pang CY, Song MZ, Wei H, Fan S, Yu SX. Analysis of MIKCC-type MADS-box Gene family in *Gossypium hirsutum*. *J Integr Agric*. 2013;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 RG. MIKCC-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.
- Linkies A, Graeber K, Knight C, Leubner-Metzger G. The evolution of seeds. *New Phytol*. 2010;186:817–31.
- Chen F, Zhang X, Liu X, Zhang L. Evolutionary analysis of MIKCC-type MADS-box genes in gymnosperms and angiosperms. *Front Plant Sci*. 2017;8:895.
- Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol*. 2003;29:464–89.
- Dreni L, Zhang D. Flower development: the evolutionary history and functions of the AGL6 subfamily MADS-box genes. *J Exp Bot*. 2016;67:1625–38.
- Wellmer F, Graciet E, Riechmann JL. Specification of floral organs in *Arabidopsis*. *J Exp Bot*. 2014;65:1–9.
- Kim S, Koh J, Yoo MJ, Kong H, Hu Y, Ma H, Soltis PS, Soltis DE. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *Plant J*. 2005;43:724–44.
- Li HF, Liang WQ, Yin CS, Zhu L, Zhang DB. Genetic interaction of *OsMADS3*, *DROOPING LEAF* and *OsMADS13* in specifying rice floral organs identities and meristem determinacy. *Plant Physiol*. 2011;156:263–74.
- O'Maileidigh DS, Graciet E, Wellmer F. Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytol*. 2014;201:16–30.
- Yuan Z, Zhang DB. Roles of jasmonate signalling in plant inflorescence and flower development. *Curr Opin Plant Biol*. 2015;27:44–51.
- Kramer EM, Hall JC. Evolutionary dynamics of genes controlling floral development. *Curr Opin Plant Biol*. 2005;8:13–8.
- Fukui M, Futamura N, Mukai Y, Wang Y, Nagao A, Shinohara K. Ancestral MADS box genes in sugi, *Cryptomeria japonica* D. Don (Taxodiaceae), homologous to the B function genes in angiosperms. *Plant Cell Physiol*. 2001;42:566–75.
- Englund M, Carlsbecker A, Engstrom P, Vergara-Silva F. Morphological primary homology and expression of AG-subfamily MADS-box genes in pines, podocarps and yews. *Evol Dev*. 2011;13:171–81.
- Lovisetto A, Guzzo F, Tadiello A, Toffali K, Favretto A, Casadoro G. Molecular analyses of MADS-box genes trace back to gymnosperms the invention of fleshy fruits. *Mol Biol Evol*. 2012;29:409–19.
- Carlsbecker A, Sundström JF, Englund M, Uddenberg D, Izquierdo L, Kvarnheden A, Vergara-Silva F, Engström P. Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol*. 2013;200:261–75.
- Gramzow L, Weilandt L, Theissen G. MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann Bot*. 2014;114:1407–29.
- Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ. Molecular and genetic analyses of the silky1 gene reveal conservation in floral organ specification between eudicots and monocots. *Mol Cell*. 2000;5:569–79.
- Li H, Liang W, Hu Y, Zhu L, Yin C, Xu J, Dreni L, Kater MM, Zhang D. Rice MADS6 interacts with the floral homeotic genes SUPERWOMAN1, MADS3, MADS58, MADS13, and drooping leaf in specifying floral organ identities and meristem fate. *Plant Cell*. 2011;23:2536–52.
- Albert VA, Barbazuk WB, Der JP, Leebens-Mack J, Ma H, Palmer JD, Rounsley S, Sankoff D, Schuster SC, Soltis DE. The Amborella genome and the evolution of flowering plants. *Science*. 2013;342:1241089.
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H. A short history of MADS-box genes in plants. *Plant Mol Biol*. 2000;42:115–49.
- Hasebe M. Evolution of reproductive organs in land plants. *J Plant Res*. 1999;112:463–74.
- 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.
- Gramzow L, Barker E, Schulz C, Ambrose B, Ashton NG, Litt A. *Selaginella* genome analysis—entering the “Homoplasy Heaven” of the MADS world. *Front Plant Sci*. 2012;3:214.
- De Bodt S, Raes J, Van De Peer Y, Theissen G. And then there were many: MADS goes genomic. *Trends Plant Sci*. 2003;8:475–83.
- Becker A, Winter KU, Meyer B, Saedler H, Theissen G. MADS-box gene diversity in seed plants 300. *Mol Biol Evol*. 2000;17:1425–34.
- Hernández-Hernández T, Martínez-Castilla LP, Alvarez-Buylla ER. Functional diversification of B MADS-box homeotic regulators of

- flower development: adaptive evolution in protein–protein interaction domains after major gene duplication events. *Mol Biol Evol.* 2007;24:465–81.
38. Airoidi CA, Davies B. Gene duplication and the evolution of plant MADS-box transcription factors. *J Genet Genomics.* 2012;39:157–65.
  39. Nam J, de Pamphilis CW, Ma H, Nei M. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol Biol Evol.* 2003;20:1435–47.
  40. Singer SD, Krogan NT, Ash-ton NW. Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. *Plant Cell Rep.* 2007;26:1155–69.
  41. Ohmori S, Kimizu M, Sugita M, Miyao A, Hirochika H, Uchida E, Nagato Y, Yoshida H. Mosaic floral organs1, an *AGL6*-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell.* 2009;21:3008–25.
  42. Immink RG, Kaufmann K, Angenent GC. The 'ABC' of MADS domain protein behaviour and interactions. *Semin Cell Dev Biol.* 2010;21:87–93.
  43. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
  44. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* 2002;30:242–4.
  45. Erdmann R, Gramzow L, Melzer R, Theissen G, Becker A. GORDITA (*AGL63*) is a young paralog of the *Arabidopsis thaliana* B(sister) MADS box gene *ABS* (*TT16*) that has undergone neofunctionalization. *Plant J.* 2010;63(6):914–24.
  46. Hileman LC, Sundstrom JF, Litt A, Chen M, Shumba T, Irish VF. Molecular and phylogenetic analyses of the MADS-box gene family in tomato. *Mol Biol Evol.* 2006;23:2245–58.
  47. Schauer SE, Schlüter PM, Baskar R, Gheyselinck J, Bolaños A, Curtis MD, Grossniklaus U. Intronic regulatory elements determine the divergent expression patterns of *AGAMOUS*-LIKE 6 subfamily members in *Arabidopsis*. *Plant J.* 2009;59:987–1000.
  48. Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet.* 2003;4:275–84.
  49. Zhao T, Holmer R, de Bruijn S, Angenent GC, van den Burg HA, Schranz ME. Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. *Plant Cell.* 2017;29:1278–92.
  50. Reece JB, Urry AL, Cain ML, Wasserman SA, Minorsky PV, Jackson RB. *Campbell biology*. 10th ed. San Francisco: Pearson; 2013.
  51. Kishino H, Thorne JL, Bruno WJ. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol.* 2001;18:352–61.
  52. Becker A, Saedler H, Theissen G. Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnetum*. *Dev Genes Evol.* 2003;213:567–72.
  53. Zhao Q, Weber AL, McMullen MD, Guill K, Doebley J. MADS-box genes of maize: frequent targets of selection during domestication. *Genet Res.* 2011;93:65–75.
  54. Hu L, Liu S. Genome-wide analysis of the MADS-box gene family in cucumber. *Genome.* 2012;55:245–56.
  55. Smaczniak C, Immink RGH, Angenent GC, Kaufmann K. Developmental and evolutionary diversity of plant MADS domain factors: insights from recent studies. *Development.* 2012;139:3081–98.
  56. Zhang Z, Li H, Zhang D, Liu Y, Fu J, Shi Y, Li Y. Characterization and expression analysis of six MADS-box genes in maize (*Zea mays* L.). *J Plant Physiol.* 2012;169:797–806.
  57. Winter KU, Saedler H, Theissen G. On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *Plant J.* 2002;31:457–75.
  58. Mondragon-Palomino M, Theissen G. Conserved differential expression of paralogous *DEFICIENS*- and *GLOBOSA*-like MADS-box genes in the flowers of orchidaceae: refining the 'orchid code'. *Plant J.* 2011;66:1008–19.
  59. Xi Z, Wang Y, Bradley RK, Sug-umaran M, Marx CJ, Rest JS. Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genet.* 2013;9:e1003265.
  60. Stegemann S, Hartmann S, Ruf S, Bock R. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci USA.* 2003;100:8828–33.
  61. Stegemann S, Keuthe M, Greiner S, Bock R. Horizontal transfer of chloroplast genomes between plant species. *Proc Natl Acad Sci USA.* 2012;109:2434–8.
  62. Irish VF. Duplication, diversification, and comparative genetics of angiosperm MADS-box genes. *Adv Bot Res.* 2006;44:129–61.
  63. Winter KU, Becker A, Münster T, Kim JT, Saedler HG. MADS-box genes reveal that gymnosperms are more closely related to conifers than to flowering plants. *Proc Natl Acad Sci USA.* 1999;96:7342–7.
  64. Jager M, Hassanin A, Manuel M, Le Guyader H, Deutsch J. MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. *Mol Biol Evol.* 2003;20:842–54.
  65. Shu Y, Yu D, Wang D, Guo D, Guo C. Genome-wide survey and expression analysis of the MADS-box gene family in soybean. *Mol Biol Rep.* 2013;40:3901–11.
  66. Vandebussche MG, Van de Peer Y, Gerats T. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res.* 2003;31:4401–9.
  67. Yu H, Goh CJ. Identification and characterization of three orchid MADS-box genes of the *AP1/AGL9* subfamily during floral transition. *Plant Physiol.* 2000;123:1325–36.
  68. Nam J, Kim J, Lee S, An G, Ma H, Nei M. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc Natl Acad Sci USA.* 2004;101:1910–5.
  69. Yockteng R, Almeida AMR, Morioka K, Alvarez-Buylla ER, Specht CD. Molecular evolution and patterns of duplication in the *SEP/AGL6*-Like Lineage of the Zingiberales: a proposed mechanism for floral diversification. *Mol Biol Evol.* 2016;30:2401–22.
  70. Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ. Genome-wide analysis of MIKCC-type MADS box genes in grapevine. *Plant Physiol.* 2009;149:354–69.
  71. Katahata SI, Futamura N, Igasaki T, Shinohara K. Functional analysis of *SOC1*-like and *AGL6*-like MADS-box genes of the gymnosperm *Cryptomeria japonica*. *Tree Genet Genomes.* 2014;10:317–27.
  72. Li HF, Liang WQ, Jia RD, Yin CS, Zong J, Kong HZ, Zhang DB. The *AGL6*-like gene *OsMADS6* regulates floral organ and meristem identities in rice. *Cell Res.* 2010;20:299–313.
  73. Wong CE, Singh MB, Bhalla PL. Novel members of the *AGAMOUS*-LIKE 6 subfamily of MIKCC-type MADS-box genes in soybean. *BMC Plant Biol.* 2013;13:105.
  74. Carlsbecker A, Tandre K, Johanson U, Englund M, Engström P. The MADS-box gene *DAL7* is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *Plant J.* 2004;40:546–57.
  75. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Lett Nat.* 1977;267:275–6.
  76. Yang ZH. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 1997;13:555–6.
  77. Shan H, Zahn L, Guindon S, Wall PK, Kong H, Ma H, DePamphilis CW, Leebens-Mack J. 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.
  78. Kim S, Yoo MJ, Albert VA, Farris JS, Soltis PS, Soltis DE. Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *Am J Bot.* 2004;91:2102–18.
  79. Lovisetto A, Guzzo F, Busatto N, Casadoro G. Gymnosperm B-sister genes may be involved in ovule/seed development and in some species, in the growth of fleshy fruit-like structures. *Ann Bot.* 2013;112:535–44.
  80. Sundström J, Carlsbecker A, Svensson ME, Svensson M, Johansen UG, Engström P. MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev Genet.* 1999;25:253–66.
  81. Meyerowitz EM. Flower development and evolution: new answers and new questions. *Proc Natl Acad Sci USA.* 1994;91:5735–7.
  82. Rutledge R, Regan S, Nicolas O, Fobert P, Cote C, Bosnich W, Kauffeldt C, Sunohara G, Seguin A, Stewart D. Characterization of an *AGAMOUS*

- homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis*. *Plant J.* 1998;15:625–34.
83. Thangavel G, Nayar S. A survey of MIKC type MADS-box genes in non-seed plants: algae, bryophytes, lycophytes and ferns. *Front Plant Sci.* 2018;9:510.
  84. Theissen G. Shattering developments. *Nature.* 2000;404:711–3.
  85. Koo SC, Bracko O, Park MS, Schwab R, Chun HJ, Park KM, Seo JS, Grbic V, Balasubramanian S, Schmid M, Godard F, Yun DJ, Lee SY, Cho MJ, Weigel D, Kim MC. Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box gene *AGAMOUS-LIKE 6*. *Plant J.* 2010;62:807–16.
  86. Callens C, Tucker MR, Zhang D, Wilson ZA. Dissecting the role of MADS-box genes in monocot floral development and diversity. *J Exp Bot.* 2018;69:2435–59.
  87. Yoo SK, Wu X, Lee JS, Ahn JH. *AGAMOUS-LIKE 6* is a floral promoter that negatively regulates the FLC/MAF clade genes and positively regulates FT in *Arabidopsis*. *Plant J.* 2011;65:62–76.
  88. Wang YQ, Melzer RG. Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of 'floral quartets'. *Plant J.* 2010;64:177–90.
  89. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA 5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol Biol Evol.* 2011;28:2731–9.
  90. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
  91. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.

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](https://biomedcentral.com/submissions)

