JSE Journal of Systematics and Evolution

Journal of Systematics and Evolution 49 (6): 528–538 (2011) doi: 10.1111/j.1759-6831.2011.00158.x

Research Article

Evolutionary pattern of the regulatory network for flower development: Insights gained from a comparison of two *Arabidopsis* species

[1,2]Yang LIU [1,2]Chun-Ce GUO [1]Gui-Xia XU [1]Hong-Yan SHAN [1]Hong-Zhi KONG*

[1](*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences,* Beijing 100093, China)

[2](*Graduate University of Chinese Academy of Sciences,* Beijing 100049, China)

Abstract Previous studies on *Arabidopsis thaliana* and other model plants have indicated that the development of a flower is controlled by a regulatory network composed of genes and the interactions among them. Studies on the evolution of this network will therefore help understand the genetic basis that underlies flower evolution. In this study, by reviewing the most recent published work, we added 31 genes into the previously proposed regulatory network for flower development. Thus, the number of genes reached 60. We then compared the composition, structure, and evolutionary rate of these genes between *A. thaliana* and one of its allies, *A. lyrata*. We found that two genes (*FLC* and *MAF2*) show 1 : 2 and 2 : 2 relationships between the two species, suggesting that they have experienced independent, post-speciation duplications. Of the remaining 58 genes, 35 (60.3%) have diverged in exon–intron structure and, consequently, code for proteins with different sequence features and functions. Molecular evolutionary analyses further revealed that, although most floral genes have evolved under strong purifying selection, some have evolved under relaxed or changed constraints, as evidenced by the elevation of nonsynonymous substitution rates and/or the presence of positively selected sites. Taken together, these results suggest that the regulatory network for flower development has evolved rather rapidly, with changes in the composition, structure, and functional constraint of genes, as well as the interactions among them, being the most important contributors.

Key words *Arabidopsis*, evolution, flower development, regulatory network.

Flowers are reproductive structures and morphological innovations of angiosperms. The morphology and structure of flowers have diversified tremendously since their origins approximately 200 million years ago (Mya) (Wikström et al., 2001; Soltis & Soltis, 2004). According to the functional and genetic studies on genes regulating flower development in model plants (such as *Arabidopsis thaliana* (L.) Heynh), the development of a flower is genetically controlled by a regulatory network formed by genes and the interactions among them (summarized in Theissen & Saedler, 2001; Zhao et al., 2001; Soltis et al., 2002; Kaufmann et al., 2005; Higgins et al., 2010). Early-acting genes in this network are flowering time genes that cause the switch from vegetative growth to reproductive growth in response to environmental and endogenous flowering signals. Late-acting genes, mostly floral organ identity genes, specify the identities of different floral organs (such as sepals, petals, stamens, and carpels) by activating their downstream organ-building genes. Other genes, including meristem identity genes and intermediate genes, function to determine the formation of inflorescence/floral meristems or regulate the expression of floral organ identity genes.

Previous studies have shown that genes with different functions in the network tend to have distinct evolutionary patterns (Theissen et al., 2000; Irish & Litt, 2005; Soltis et al., 2007). Some genes have been highly conserved and kept low copy during evolution. For example, *LFY* acts as an integrating gene to specify floral meristem identity and maintains as a single-copy gene in almost all investigated angiosperms (Himi et al., 2001). Other genes, which are not as conserved as *LFY*, show evolutionary histories that are tightly associated with plant evolution. A-, B-, C/D-, and E-classes of MADS-box genes, which specify the identities of floral organs, belong to the *AP1/SQUA*, *AP3/PI*, *AG/STK*, and *SEP* subfamilies, respectively (Becker & Theissen, 2003), and have experienced several rounds of duplication events at the bases of angiosperms, core eudicots, and grasses (Shan et al., 2009). There are still some genes that are highly variable evolutionarily and suffered duplication events irrelevant to plant

Received: 4 May 2011 Accepted: 6 July 2011

* Author for correspondence. E-mail: hzkong@ibcas.ac.cn; Tel.: 86-10-62836489; Fax: 86-10-62590843.

evolution. *FLC*-like MADS-box genes, for instance, are core eudicot-specific and have been subject to multiple lineage-specific duplication events during evolution (Diaz-Riquelme et al., 2009).

In addition to the different evolutionary patterns among floral genes, recent studies have also suggested that homologous or even orthologous genes do not necessarily have the same function. *LFY*, for example, is positively regulated by *SOC1* in *Arabidopsis*, whereas its ortholog in rice, *RFL*, upregulates the expression of a *SOC1*-like gene, *OsMADS50* (Rao et al., 2008). Similarly, *CO* promotes flowering of *Arabidopsis* during long days, whereas its rice counterpart, *Hd1*, represses flowering during long days (Yano et al., 2000). These observations, together with many others, clearly indicate the correlations between changes in phenotypes and those in genotypes and highlight the importance of studies on the evolution of the floral regulatory network. However, due to complicated evolutionary histories and the function divergence of floral genes in angiosperms, the gene composition and regulatory relationships between homologous genes are sometimes quite different in the network of distantly related species, as we have learned from *Arabidopsis* and rice (Higgins et al., 2010). It has been difficult to reveal the evolutionary dynamics of the network through comparisons of floral genes at large scales.

For these reasons, in the past few years, researchers have turned to investigate the differences of floral genes between closely related species or even ecotypes. They have discovered that flowering time genes tend to evolve rapidly through different ways (e.g. Flowers et al., 2009; Takahashi et al., 2009; Higgins et al., 2010). For instance, by comparing floral genes of different *Arabidopsis* species or ecotypes, it has been found that, although most of the genes tend to evolve under strong purifying selection, some have displayed signatures of adaptive evolution (Lawton-Rauh et al., 1999; Le Corre et al., 2002; Moore et al., 2005; Flowers et al., 2009). Studies on six flowering time genes of 64 rice cultivars showed that obvious variations have happened in Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels (Takahashi et al., 2009). Comparative genomic studies on floral genes of rice and *Brachypodium distachyon* (L.) P. Beauv indicated that gain/loss of genes and the modification of the core photoperiod pathway have led to the divergence of flowering time pathways between the two grass species (Higgins et al., 2010). Despite recent advances in elucidating the evolutionary patterns and mechanisms of floral genes in different plant groups, a comprehensive and comparative study is still wanted.

Arabidopsis thaliana and *A. lyrata* (L.) O'Kane & Al-Shehbaz (hereafter called *thaliana* and *lyrata*, respectively) are congeneric species in the Brassicaceae family. They have a recent evolutionary origin with the divergence time being ~10 Mya (Hu et al., 2011). However, they show significant differences in flower size, mating system, number of chromosomes, and genome size (Beaulieu et al., 2007). In recent years, great progress has been made in functional studies on floral genes of *Arabidopsis*. Particularly, many new genes and interactions among genes have been well characterized. These advances, together with the completion of whole genome sequencing for *thaliana* and *lyrata*, have provided an excellent opportunity to study the evolutionary dynamics of the network. In this study, we added 31 genes into the floral regulatory network summarized by Kaufmann et al. (2005) and investigated the evolutionary pattern of the network by comparing floral genes in *thaliana* and *lyrata*. We discovered that variations in gene composition, gene structure and molecular evolutionary rate have contributed to the evolution of the network in *Arabidopsis*.

1 Material and methods

1.1 Data retrieval

Genomic DNA, cDNA, and amino acid sequences of *thaliana* genes used in this study were retrieved from TAIR (http://www.arabidopsis.org). The corresponding sequences of their orthologous genes in *lyrata* were obtained by searching against the *lyrata* genome database at the Phytozome website (http://www.phytozome.net/search.php?method=Org_Alyrata). Because gene gain and loss after speciation will influence the identification of orthologous genes, only genes from the two species were considered when the following criteria were met: (i) they show high sequence identity and have the same functional domains; (ii) they are sister to each other in the phylogenetic tree; and (iii) they show an evident microsyntenic relationship.

To this end, we first carried out BLASTP searches against the protein databases of *thaliana* and *lyrata*, with the protein sequences of *thaliana* as queries. The cut-off of the e-value was set as 10^{-50}. Meanwhile, TBLASTN against the genome of *lyrata* was also carried out to search potential sequences that were not annotated in the current released protein database. After removing redundant sequences from BLASTP and TBLASTN output results, we constructed an amino acid dataset for each floral gene for further sequence alignments and phylogenetic analyses. The Pfam (http://pfam.sanger.ac.uk/) webserver was used to examine domains of obtained proteins.

1.2 Sequence alignment and phylogenetic construction

Protein sequences for each dataset were aligned with CLUSTAL X 1.83 (Thompson et al., 1997) and adjusted manually using GeneDoc (Nicholas & Nicholas, 1997). Phylogenetic analyses for each protein matrix were carried out using the neighbor joining (NJ) method with p-distance as the substitution model, pairwise deletion of gaps, 1000 bootstrap replications, and other default parameters in MEGA version 4.0 (Tamura et al., 2007).

Note that the phylogenetic tree of *MAF2* homologs showed lower internal support for the *MAF2*/*MAF3* lineage, whatever protein or cDNA matrices, NJ or maximum likelihood (ML) methods were used (Fig. S1). Therefore, the nucleotide sequences of introns 2–5 of these genes, which possess more informative sites, were used for the phylogenetic tree construction with both NJ and ML methods. The ML analysis was carried out using PhyML version 2.4 (Guindon & Gascuel, 2003) with the general time reversible model. The proportion of invariable sites and the gamma distribution parameter for rate variation across variable sites were optimized and a BIONJ tree was used as the initial tree for ML searches. Bootstrap analyses were carried out with 100 replicates.

Finally, the genes of *thaliana* and *lyrata* that show the closest sister relationship in the phylogenetic tree were selected as the candidate orthologous genes. The colinearity of each gene pair was determined through comparing the gene order and organization of the genomic regions containing the focal genes, which was carried out with Genome Browsers at the JGI website (http://genome.jgi-psf.org/Araly1/Araly1.home.html).

1.3 Detection of sequence divergence between orthologs

Divergence between orthologs was first detected based on the alignment of protein and the corresponding nucleotide sequences. If obvious sequence difference exists in the $5'/3'$ end or exon–intron boundary, careful comparisons at the genomic level were carried out. For some genes having multiple transcripts, only the transcripts showing the highest sequence identity to each other were selected. Insertions/deletions (hereafter called indels) were determined if gaps were found within the exons of either of the orthologous genes. Exonizations/pseudoexonizations were deduced if the exonic sequence of one gene was aligned well with the intronic or untranslated region of its orthologs, or the intergenic region flanking its orthologs.

1.4 Estimation of ω ratio

The ratio of nonsynonymous to synonymous substitution rates ($\omega = d_N/d_S$) is a widely used indicator of natural selection (Kimura, 1983; Hughes & Nei, 1988). $\omega < 1$ indicates negative (purifying) selection, $\omega = 1$ neutral selection, and $\omega > 1$ positive (adaptive) selection (Hughes & Nei, 1988). To explore the molecular evolutionary patterns of floral genes, we estimated d_S, d_N and ω ratio using the method of Nei & Gojobori (1986) with the Jukes and Cantor correction for multiple hits in MEGA version 4.0 (Tamura et al., 2007). Standard errors for d_S and d_N were obtained by 500 bootstrap replications. The significant differences of the ω ratios between flowering time genes, meristem identity genes, intermediate genes, and organ identity genes were evaluated according to the P-values, which were calculated through a Mann–Whitney U-test in SPSS version 13.0 (SPSS, Chicago, IL, USA).

Because the above methods estimate the average ω ratios for all sites of the sequence, only the positive selection acting on the majority of the sites can produce an overall ω value above 1.0. However, in most cases, positive selection affects a few sites along the sequence, which could not be detected by estimating the average ω ratios. Therefore, we used sliding window analyses to detect positive selection at a certain region over the whole coding sequence, which was carried out with DnaSP version 4.10 (Rozas et al., 2003). Window length was set as 45 bp with step size 9 bp or 15 bp. Only genes showing $\omega > 1$ by the two step size analyses were considered as having internal regions under positive selection.

2 Results

2.1 Update of the regulatory network for flower development in *Arabidopsis*

The floral regulatory network was first reviewed by Theissen et al. (1996) and has been updated several times over the last decade (Theissen, 2001; Zhao et al., 2001; Soltis et al., 2002; Kaufmann et al., 2005). Considering that many genes involved in flower development have been functionally identified recently, a new summary is still needed for revealing a general picture of the evolutionary pattern of the network. In this study, we added 31 genes, most of which are flowering time genes and intermediate genes, to the latest network of Kaufmann et al. (2005) (Fig. 1). Thus, the number of genes involved in flower development reached 60 in the current network. Among the new genes, *AGL15* and *AGL18* act redundantly as repressors of the floral transition and delay flowering through inhibiting the

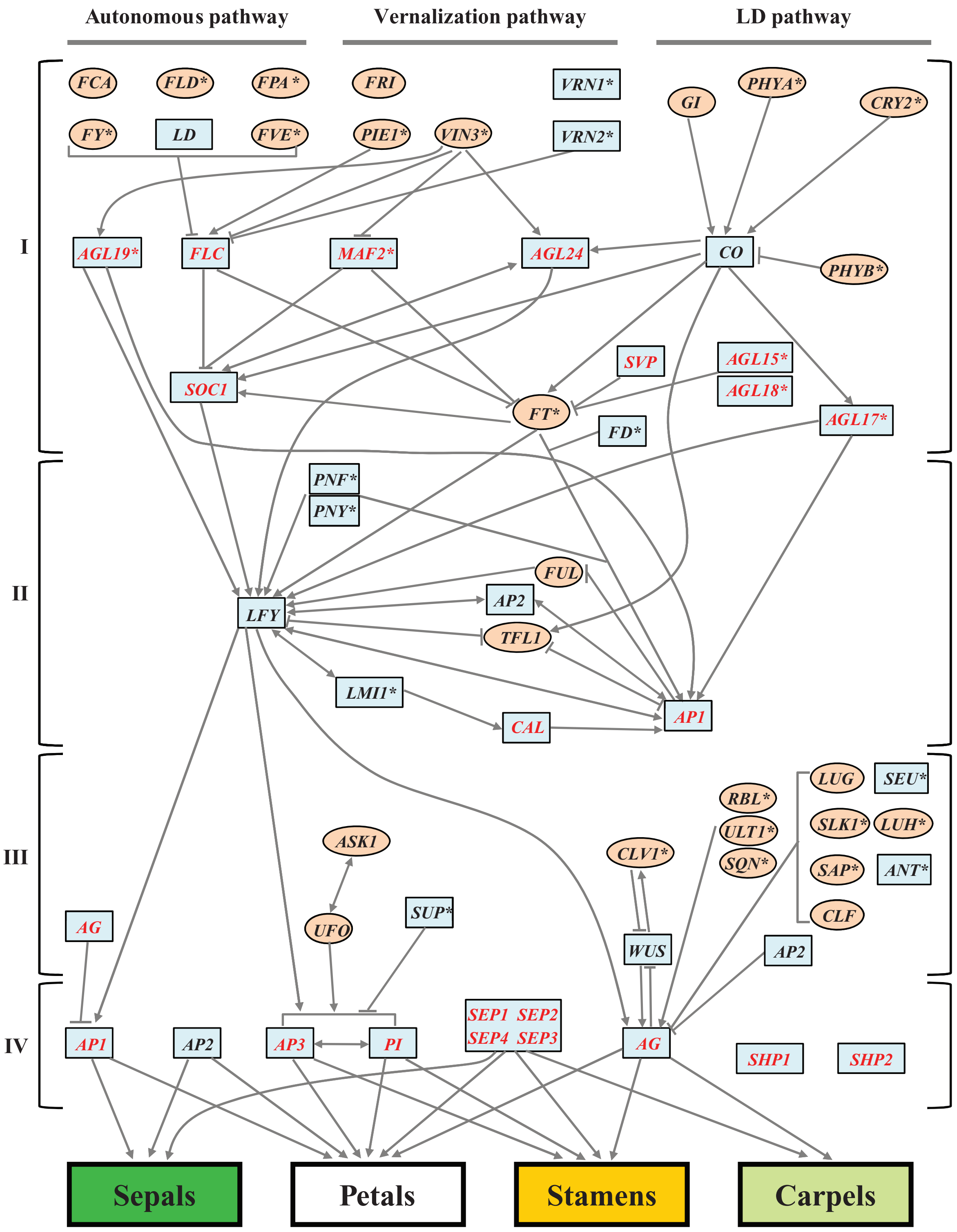

Fig. 1. Regulatory network for floral development modified from Kaufmann et al. (2005). Regulatory interactions between genes are symbolized by arrows (activation) or barred lines (inhibition, antagonistic interaction). Four categories of floral genes are denoted with roman numbers: I, flowering time genes; II, meristem identity genes; III, intermediate genes; and IV, organ identity genes. Genes encoding transcription factors are represented by light blue rectangles, MADS-box genes are highlighted with red font, and genes encoding non-transcription factors by orange ovals. The newly added genes are highlighted with asterisks.

expression of *FT* (Adamczyk et al., 2007); *MAF2*, an *FLC*-like gene, prevents flowering through a pathway independent of the *FLC* regulation under short periods of cold (Ratcliffe et al., 2003). *AGL17* and *AGL19* are floral activators that promote flowering through activating *LFY* and *AP1* in an *FT*-independent photoperiod pathway and an *FLC*-independent vernalization pathway, respectively (Schonrock, 2006; Han et al., 2008). *LUG* and *SEU* prevent ectopic *AG* expression in flowers by forming co-repressor complex (Sridhar et al., 2006), whereas *RBL*, *ULT1*, and *SQN* function redundantly in the temporal regulation of floral meristem termination by activating the expression of *AG* (Prunet et al., 2008).

According to the classification of previous studies (e.g. Soltis et al., 2002; Kaufmann et al., 2005), we divided the floral genes into four types, including 27 flowering time genes, 9 meristem identity genes, 17 intermediate genes, and 11 organ identity genes (Fig. 1). These components concern 60 different genes because *AP1*, *AP2*, and *AG* are multifunctional and play roles in different stages of floral development (Fig. 1). Among them, 35 genes encode transcription factors (TFs) and are present in all hierarchies. Notably, 60% of TFs (21/35) are members of the MADS-box gene family. They are indispensable not only for the formation of inflorescence/floral meristems and floral organs (e.g. *AG*, *AP1*, *AP3*, *CAL*, *FUL*, *PI*, *SEP1/2/3/4*, and *SHP1/2*), but also for the early regulation of flower development (e.g. *AGL15*, *AGL17*, *AGL18*, *AGL19*, *AGL24*, *FLC*, *MAF2*, *SOC1*, and *SVP*). The detailed functional information for genes and regulatory interactions among them in the new network is summarized in Table S1.

2.2 Comparison of composition and structure of floral genes

To explore the evolutionary pattern of the network, we obtained the candidate orthologs of floral genes of *thaliana* from the genome database of *lyrata*. Our results revealed that among the 60 genes surveyed, 58 genes show a 1 : 1 orthologous relationship, but two flowering time genes (*FLC* and *MAF2*) show 1 : 2 and 2 : 2 relationships in phylogenetic trees, suggestive of independent, post-speciation gene duplications. Interestingly, both *FLC* and *MAF2* are members of the *FLC*-like MADS-box gene subfamily and function as flowering repressors (Michaels & Amasino, 1999; Ratcliffe et al., 2003). Nah & Chen (2010) have reported that the *FLC* locus has experienced a tandem duplication event in *lyrata* since it separated from *thaliana*, leading to the production of two *FLC* genes, *AlFLC1* and *AlFLC2*.

Unlike *FLC*, two independent gene duplications have occurred in *MAF2* after the divergence of *thaliana* and *lyrata*, giving rise to *AtMAF2* and *AtMAF3* in *thaliana*, and *Aly496727* and *Aly497339* in *lyrata* (Fig. 2: A). By observing the gene order and organization of the genomic regions flanking *MAF2* in the two species, we found that *AtMAF2* and *AtMAF3* are arranged in a tandem array on chromosome 5 of *thaliana*, whereas *Aly496727* and *Aly497339* are distributed in scaffolds 8 and 233 of *lyrata*, respectively (Fig. 2: B). The overall microsynteny in the vicinity of *AtMAF2/AtMAF3* and *Aly496727* was highly maintained in the two species (Fig. 2: B). In contrast, in scaffold 233, *Aly497339* matches very well with *AtMAF3*, and the upstream region of *Aly497339* shows high similarity with the seventh exon of *AtMAF2* (Fig. 2: B). Coincidently, *Aly496727* has only six exons, whereas its homologs generally have seven exons (Fig. 2: B). Combining these findings and the fact that scaffold 233 is only approximately 12 kb, and contains just an intact gene *Aly497339*, we suspected that this scaffold may be a part of scaffold 8, and located between *Aly496727* and *Aly886791*. It needs to be verified through deep sequencing and fine annotation of the genomic region covering these genes. Taken together, these findings suggest that genes involved in flower development are not completely the same in the two species, although they separated by only approximately 10 million years (Hu et al., 2011).

To investigate the conservation and divergence of floral genes between the two species, we carried out exon–intron structure comparisons for the 58 genes showing a 1 : 1 relationship. To our surprise, we found that 35 genes (35/58, 60.3%) have diverged clearly in the exon–intron structure, and the mechanisms that underlie structural divergence include indels and exonizations/pseudoexonizations (Table S2). In particular, the exonizations/pseudoexonizations have led to obvious differences of *AGL15*, *RBL*, *SEP1*, and *FLD* in exon–intron boundaries and amino acid sequences, even in exon numbers. Further comparisons at cDNA and genomic levels suggested that the exonizations/pseudoexonizations of these genes were caused by different mechanisms (Fig. 3). In the case of *AGL15*, it was generated by an out-of-frame indel of 4 bp within the eighth exon (Fig. S2); *RBL*, by a substitution between “A” and “T” in the eighth exon that led to the occurrence of a new stop codon in *AtRBL* (Fig. S3); *SEP1*, by the use of a different alternative acceptor site “AG” for the sixth intron (Fig. S4); and *FLD*, by the selection of a different alternative donor site “GT” for the second intron and subsequently the introduction of a novel stop codon (TAG) to *AlFLD* (Fig. S5).

For the remaining 31 genes, the divergence of exon–intron structure was created only through indels with the length being 3 or a multiple of 3

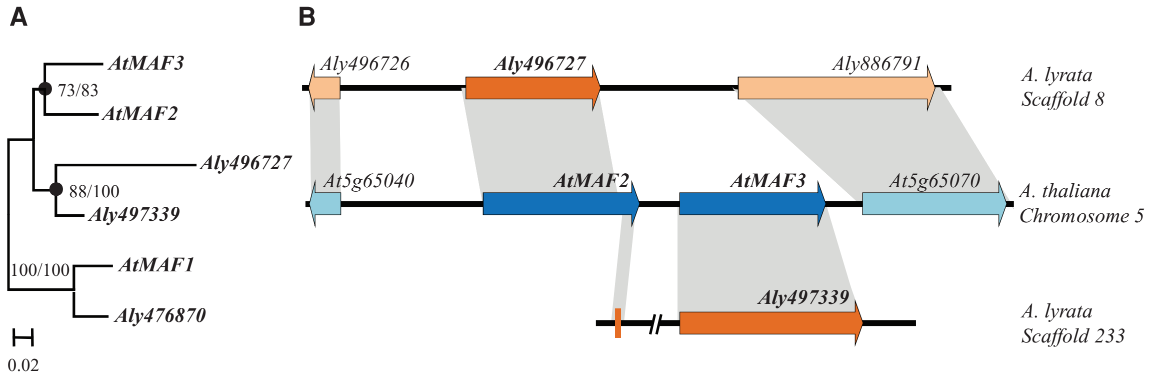

Fig. 2. Phylogenetic tree (**A**) and microsyntenic relationship (**B**) of *MAF2* genes of *thaliana* and *lyrata*. **A,** Phylogenetic tree is constructed with neighbor-joining and maximum likelihood methods based on the nucleotide sequences of introns 2–5. The numbers close to each branch indicate bootstrap values greater than 50%. The black dots indicate two independent gene duplication events in *thaliana* and *lyrata*. **B,** Annotated genes in *thaliana* and *lyrata* are represented by blue and yellow arrows, respectively. The microsyntenic relationships between orthologous genes are shown with gray shadows.

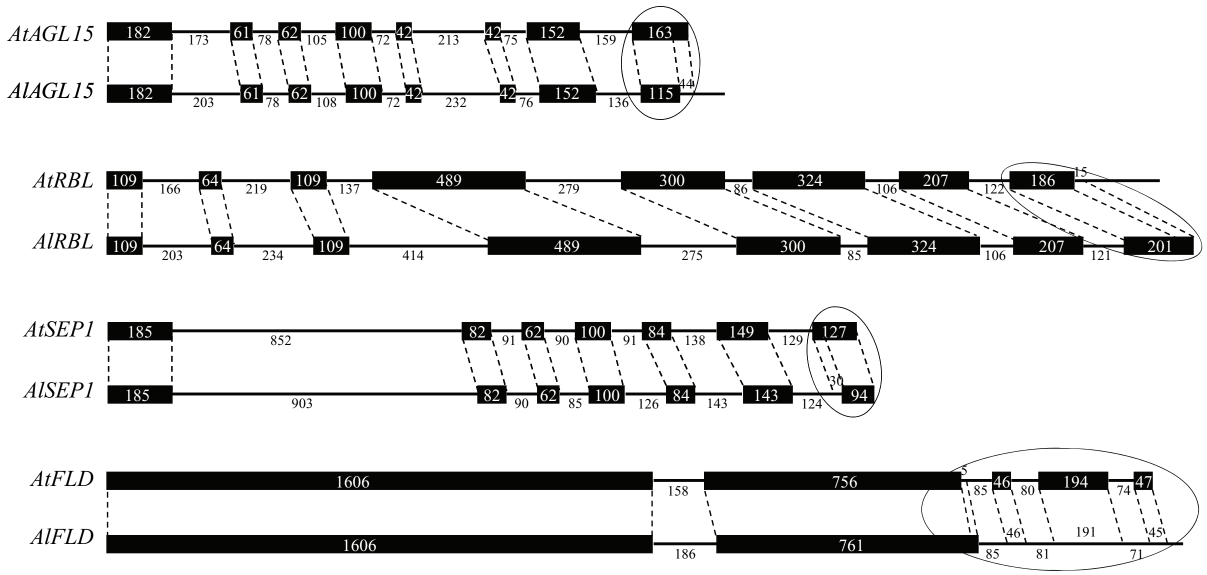

Fig. 3. Schematic representations of exon–intron structural divergence in *AGL15*, *RBL*, *SEP1*, and *FLD* of *thaliana* and *lyrata*. Exons are indicated with black boxes, and introns with lines. The numbers represent the lengths of exons and introns, which are largely drawn to scale. Regions (especially exons) that can match to each other are connected with dashed lines. Ringed regions indicate the positions where exonization/pseudoexonization took place.

(Table S2). The number of indels per gene pair ranged from 1 to 8, with an average of 2.39 (74/31). Although the length and frequency of the indels vary among genes, such events have not caused the change in the exon–intron boundaries because all the indels happened within the exon and no out-of-frame indel was detected. For 23 genes (23/58, 39.7%) not showing gene structure divergence, the sequence differences were caused by point mutations. These findings indicate that floral genes have diverged strikingly at a very small evolutionary scale. Moreover, besides point mutations, indels and exonizations/pseudoexonizations have predominantly contributed to the sequence difference.

2.3 Molecular evolutionary analyses

To investigate the patterns of molecular evolution of floral genes, we estimated d_N, d_S, and ω values of 58 genes showing a 1 : 1 relationship. As a result, all the investigated genes have ω values smaller than 0.40 (Table S2). The average ω value is 0.17, the maximum is 0.40 for *FPA*, and the minimum is 0 for *SEP1*. This suggests that these genes have overall evolved under purifying selection, but they are subject to different selective constraints during evolution. It is quite interesting that *SEP1* had the lowest ω value of 0, suggesting that no nonsynonymous substitution has been accumulated in *SEP1* orthologs. However, obvious sequence difference

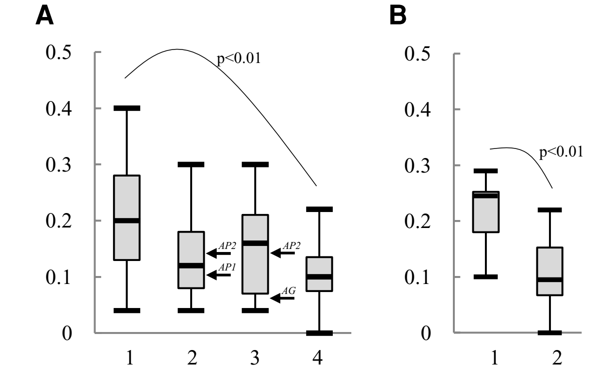

Fig. 4. Comparison of the molecular evolutionary rates of genes in the floral regulatory network in two *Arabidopsis* species. **A,** Box-plot analysis of the ω values for flowering time genes (1), meristem identity genes (2), intermediate genes (3), and organ identity genes (4). **B,** Box-plot analysis of the ω values for MADS-box genes that function as flowering time genes (1), and those that act as organ identity genes (2). The ω values of upper quartile, median, and lower quartile are indicated in each box, whereas the bars outside the box indicate the maximum and minimum ω values. *P*-values are from Mann–Whitney *U*-tests.

has been observed at the protein level due to the divergence of the exon–intron structure between *AtSEP1* and *AlSEP1* (Fig. 3). In addition, we found that five genes (*FCA*, *FRI*, *AGL19*, *CO*, and *FPA*) have ω values greater than 0.30, which were largely caused by higher d_N values, suggesting that they may evolve under relatively weak selection pressure, even positive selection.

Furthermore, we compared the ω values of genes in different hierarchies. We found that flowering time genes (average $\omega = 0.21$, 0.04–0.40) have significantly greater ω values than organ identity genes (average $\omega = 0.11$, 0–0.22) ($P < 0.01$, Mann–Whitney *U*-test), but no significant difference was observed for comparisons between other kinds of genes (Fig. 4: A). These results suggest that flowering time genes evolved under relatively relaxed purifying selection, whereas other genes in the network, especially organ identity genes, experienced strong purifying selection during evolution. In addition, the ω values of MADS-box genes regulating flowering time (average $\omega = 0.23$, 0.10–0.34) are significantly higher than those of MADS-box genes specifying floral organ identities (average $\omega = 0.11$, 0–0.22) ($P < 0.01$, Mann–Whitney *U*-test) (Fig. 4: B). This indicates that selective constraints acting on genes of the same family are closely associated with their functions in the network.

As mentioned above, the overall ω estimation could not confidently reveal the positive selection if the majority of the sites are subject to purifying selection along the sequence. Therefore, sliding window analyses were carried out to identify genes with specific regions evolving under positive selection. In total, we found 19 genes with regions showing $\omega > 1$, including 10 flowering time genes (*AGL15*, *AGL17*, *AGL19*, *AGL24*, *CO*, *FCA*, *FD*, *FPA*, *FRI*, and *LD*), three meristem identity genes (*CAL*, *PNF*, and *PNY*) and six intermediate genes (*ASK1*, *CLV1*, *LUH*, *RBL*, *SAP*, and *SLK*) (Fig. S6). It should be noted that among these genes, five flowering time genes show significant signals of positive selection in the functionally conserved regions, such as the K domains of *AGL15* and *AGL17*, the bZIP1 domain of *FD*, the SPOC domain of *FPA*, and the Frigida domain of *FRI* (Fig. S6). These results, together with ω analyses for the entire coding sequences, indicate that evolutionary heterogeneity not only exists among different types of genes in the network, but also among internal regions of some genes.

3 Discussion

It has been proposed that complicated evolution of floral genes has increased the robustness and evolvability of the regulatory network for flower development, which finally led to the innovations of flowers of angiosperms (Wagner, 2008). In the present study, we investigated the evolutionary pattern of the network by comparing 60 floral genes in two *Arabidopsis* species from different aspects. We found that the variations of these genes in gene composition, exon–intron structure, and molecular evolutionary rate have contributed to the dynamic evolution of the network. Our results shaped a general picture of the evolutionary pattern of the regulatory network for flower development and provided new ideas for the comparative genomic study in closely related species.

3.1 Multiple mechanisms contribute to evolution of floral regulatory network

Previous studies on floral genes in different accessions of *thaliana* or different *Arabidopsis* species have shown that floral organ identity genes evolved under strong purifying selection, but some flowering time genes experienced relatively relaxed purifying selection, even positive selection (Lawton-Rauh et al., 1999; Le Corre et al., 2002; Olsen et al., 2002; Moore et al., 2005; Flowers et al., 2009). However, the intermediate genes and meristem identity genes in the network were paid less attention in the past. In this study, by comparing molecular evolutionary rates of all types of floral genes, we drew a similar conclusion that, although the whole network for flower development has evolved under purifying selection, flowering time genes are generally subject to less stringent selective pressure. In particular, sliding window analyses indicated that positive selection happened in the functionally conserved domains of five flowering time genes. Among them, *FPA* and *FRI* play roles early in the autonomous pathway and the vernalization pathway, respectively. *AGL15*, *AGL17*, and *FD* function slightly later than *FPA* and *FRI*, to regulate the expression of the integrating factor *FT* or the meristem identity genes *LFY* and *AP1*. These findings indicated the contribution of positive selection to the protein divergence between *thaliana* and *lyrata*, which may have affected the functions of floral genes in the network.

Our comparative genomic studies indicated that the rapid evolution of flowering time genes could also be accomplished through alternations of gene composition or divergence of exon–intron structure in addition to variations in molecular evolutionary rates. Among the 60 genes investigated, two flowering time genes, *FLC* and *MAF2*, have experienced different evolutionary histories. After the split of *thaliana* and *lyrata*, an additional gene duplication event led to the creation of two *FLC* genes in *lyrata*, and two independent gene duplication events gave rise to *AtMAF2* and *AtMAF3* in *thaliana*, and *Aly496727* and *Aly497339* in *lyrata*. Actually, the variation of copy number in *FLC* has been shown in three *Arabidopsis*-related species, which has probably contributed to flowering time diversification in *Arabidopsis* species (Nah & Chen, 2010). In addition, population genetic studies have shown that the allelic variations at *FLC* and *MAF2* loci, which could be caused by high levels of nonsynonymous single nucleotide polymorphisms (SNPs), indels, gene fusion, insertion of transposons, and so on, were major determinants of flowering habit in *Arabidopsis* accessions (Gazzani et al., 2003; Michaels et al., 2003; Caicedo et al., 2009; Rosloski et al., 2010). Our results and previous studies suggest that *FLC* and *MAF2* genes experienced rapid and complicated divergence during the evolution of *Arabidopsis*, and their dramatic variations in sequence polymorphisms, transcription patterns, gene architectures, and copy numbers have led to the diversity of flowering time trait in different ecotypes and species of *Arabidopsis*.

It has been generally believed that orthologous genes from different species tend to be conserved in the protein-coding region and function. However, we found that approximately 60% of genes in the network have diverged in gene structure through indels and exonizations/pseudoexonizations. The change in exon–intron structure has been reported to be a major mechanism underlying the divergence of protein-coding regions of paralogous genes in MADS-box and F-box gene families (e.g. Xu & Kong, 2007; Xu et al., 2009). Moreover, it has been suggested that such change could potentially cause the functional divergence of duplicated genes (Xu & Kong, 2007; Xu et al., 2009). In addition, a similar phenomenon has also been found in orthologous genes of the F-box gene family in *thaliana* and *lyrata*, and some of the structurally diverged orthologs have dramatically differentiated in protein sequences (our unpublished data). In this study, we found that four genes (*AGL15*, *RBL*, *SEP1*, and *FLD*), two of which are flowering time genes, show obvious divergence in exon–intron boundaries and protein sequences. *AGL15* is a member of the MADS-box gene family and functions as a flowering repressor by inhibiting the expression of *FT* (Adamczyk et al., 2007). *FLD* encodes a SWIRM domain-containing protein, which upregulates *FLC* expression and extensively delays flowering (Yang & Chou, 1999). However, it is not clear whether these genes carry out different functions in *lyrata*, which needs to be further tested experimentally.

3.2 Evolutionary pattern of floral regulatory network depends on its function

Our study showed several lines of evidence revealing the evolutionary dynamics of the floral regulatory network, especially the patterns of rapid divergence of flowering time genes in the two *Arabidopsis* species. Flowering time genes have been suggested as possible targets of plant adaptive evolution because, for plants, the switch from vegetative growth to reproductive growth is an important developmental step in their life cycle. Flowering needs to occur when conditions for pollination and seed development are optimal, and consequently most plants restrict flowering to a specific time of year. For the two *Arabidopsis* species, *thaliana* usually flowers approximately 30 days after germination, but the flowering time of *lyrata* varies from 75 to 175 days without vernalization (Riihimaki et al., 2005). A recent comparison of the genomes of *thaliana* and *lyrata* showed that the overall sequence identity between them is greater than 80%, and *thaliana* has 17% fewer genes than *lyrata* (Hu et al., 2011). This suggests that the genome sequences of *thaliana* and *lyrata* have significantly diverged since they separated from each other approximately 10 Mya. Therefore, it is not surprising that we identified dramatic divergence in flowering time genes, including gene composition, gene structure, and selective constraint. These differences on the molecular level enable *thaliana* and *lyrata* to adjust reproduction in different environmental stimuli, which is an important reflection of adaptation. In contrast, floral organs, especially stamens and carpals, must keep relative stability on the molecular level, and thus in function, otherwise fitness of plants would decline sharply.

So far, the evolutionary patterns of some networks or pathways with specific functions have been studied. For example, it has been reported that, in the anthocyanin pathway, the upstream genes evolved substantially more slowly than the downstream genes, which resulted from strong purifying selection for the upstream genes (Rausher et al., 1999; Lu & Rausher, 2003; Rausher et al., 2008). In contrast, it seems that there is no clear correlation between evolutionary rate and gene position in the gibberellin pathway (Yang et al., 2009). In our case, it is obvious that the upstream flowering time genes evolved significantly faster than the downstream floral organ identity genes in the floral regulatory network. It is possible that, for different genetic pathways or networks, the specific nature of selection on the component genes depends largely on the function of the pathway and the network, as indicated by Cork & Purugganan (2004).

In this study, we revealed a general picture of the evolutionary pattern of the floral regulatory network and found that gene composition and changes in the exon–intron structure of orthologs are two important contributors to the rapid evolution of the network in addition to the molecular evolutionary rate. However, our study is preliminary. The downstream organ-building genes were not included in the current study, which will be helpful for elucidating the molecular mechanism underlying the difference in flower size of *thaliana* and *lyrata*. It also remains unclear whether variations in the gene copy number and gene structure of orthologs happened prevalently in other closely related plant species. Therefore, in future comparative genomic studies, more genes and more plants should be included to completely appreciate the conservation and evolvability of the regulatory network for flower development.

Acknowledgements We are very grateful to Dahai GAO, Wenwu WU, and two anonymous reviewers for their constructive comments on the manuscript. This research was supported by the National Natural Science Foundation of China (Grant No. 30970210), and the Chinese Academy of Sciences Knowledge Innovation Program (KSCX2-YW-R-135).

References

Adamczyk BJ, Lehti-Shiu MD, Fernandez DE. 2007. The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in *Arabidopsis*. The Plant Journal 50: 1007–1019.

Beaulieu J, Jean M, Belzile F. 2007. Linkage maps for *Arabidopsis lyrata* subsp. *lyrata* and *Arabidopsis lyrata* subsp. *petraea* combining anonymous and *Arabidopsis thaliana*-derived markers. Genome 50: 142–150.

Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Molecular Phylogenetics and Evolution 29: 464–489.

Caicedo AL, Richards C, Ehrenreich IM, Purugganan MD. 2009. Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the *Arabidopsis thaliana MAF2–5* flowering time gene cluster. Molecular Biology and Evolution 26: 699–711.

Cork JM, Purugganan MD. 2004. The evolution of molecular genetic pathways and networks. BioEssays 26: 479–484.

Diaz-Riquelme J, Lijavetzky D, Martinez-Zapater JM, Carmona MJ. 2009. Genome-wide analysis of MIKC^{C}-type MADS-box genes in grapevine. Plant Physiology 149: 354–369.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the *Arabidopsis thaliana* flowering time gene network. Molecular Biology and Evolution 26: 2475–2486.

Gazzani S, Gendall AR, Lister C, Dean C. 2003. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiology 132: 1107–1114.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704.

Han P, García-Ponce B, Fonseca-Salazar G, Alvarez-Buylla ER, Yu H. 2008. *AGAMOUS-LIKE 17*, a novel flowering promoter, acts in a *FT*-independent photoperiod pathway. The Plant Journal 55: 253–265.

Higgins JA, Bailey PC, Laurie DA. 2010. Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. PLoS One 5: e10065.

Himi S, Sano R, Nishiyama T, Tanahashi T, Kato M, Ueda K, Hasebe M. 2001. Evolution of MADS-box gene induction by *FLO*/*LFY* genes. Journal of Molecular Evolution 53: 387–393.

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister J, Ossowski S, Ottilar R, Salamov A, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah M, Bergelson J, Carrington J, Gaut B, Schmutz J, Mayer K, Van de Peer Y, Grigoriev I, Nordborg M, Weigel D, Guo Y. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nature Genetics 43: 476–481.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167–170.

Irish VF, Litt A. 2005. Flower development and evolution: Gene duplication, diversification and redeployment. Current Opinion in Genetics and Development 15: 454–460.

Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS domain proteins: Structural modularity, protein interactions and network evolution in land plants. Gene 347: 183–198.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Lawton-Rauh AL, Buckler ESt, Purugganan MD. 1999. Patterns of molecular evolution among paralogous floral homeotic genes. Molecular Biology and Evolution 16: 1037–1045.

Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: Extensive nonsynonymous variation is consistent with local selection for flowering time. Molecular Biology and Evolution 19: 1261–1271.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. Molecular Biology and Evolution 20: 1844–1853.

Michaels SD, Amasino RM. 1999. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell 11: 949–956.

Michaels SD, He YH, Scortecci KC, Amasino RM. 2003. Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. Proceedings of the National Academy of Sciences USA 100: 10102–10107.

Moore RC, Grant SR, Purugganan MD. 2005. Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. Molecular Biology and Evolution 22: 91–103.

Nah G, Chen ZJ. 2010. Tandem duplication of the *FLC* locus and the origin of a new gene in *Arabidopsis* related species and their functional implications in allopolyploids. New Phytologist 186: 228–238.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418–426.

Nicholas K, Nicholas H Jr. 1997. GeneDoc: A tool for editing and annotating multiple sequence alignments. EMBNEW NEWS 4: 1–4.

Olsen KM, Womack A, Garrett AR, Suddith JI, Purugganan MD. 2002. Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics 160: 1641–1650.

Prunet N, Morel P, Thierry AM, Eshed Y, Bowman JL, Negrutiu I, Trehin C. 2008. *REBELOTE*, *SQUINT*, and *ULTRAPETALA1* function redundantly in the temporal regulation of floral meristem termination in *Arabidopsis thaliana*. The Plant Cell 20: 901–919.

Rao NN, Prasad K, Kumar PR, Vijayraghavan U. 2008. Distinct regulatory role for *RFL*, the rice *LFY* homolog, in determining flowering time and plant architecture. Proceedings of the National Academy of Sciences USA 105: 3646–3651.

Ratcliffe OJ, Kumimoto RW, Wong BJ, Riechmann JL. 2003. Analysis of the *Arabidopsis MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. The Plant Cell 15: 1159–1169.

Rausher MD, Lu Y, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. Journal of Molecular Evolution 67: 137–144.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Molecular Biology and Evolution 16: 266–274.

Riihimaki M, Podolsky R, Kuittinen H, Koelewijn H, Savolainen O. 2005. Studying genetics of adaptive variation in model organisms: Flowering time variation in *Arabidopsis lyrata*. Genetica 123: 63–74.

Rosloski SM, Jali SS, Balasubramanian S, Weigel D, Grbic V. 2010. Natural diversity in flowering responses of *Arabidopsis thaliana* caused by variation in a tandem gene array. Genetics 186: 263–276.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497.

Schonrock N. 2006. Polycomb-group proteins repress the floral activator *AGL19* in the *FLC*-independent vernalization pathway. Genes and Development 20: 1667–1678.

Shan H, Zahn L, Guindon S, Wall PK, Kong H, Ma H, dePamphilis CW, Leebens-Mack J. 2009. Evolution of plant MADS-box transcription factors: Evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. Molecular Biology and Evolution 26: 2229–2244.

Soltis DE, Ma H, Frohlich MW, Soltis PS, Albert VA, Oppenheimer DG, Altman NS, dePamphilis C, Leebens-Mack J. 2007. The floral genome: An evolutionary history of gene duplication and shifting patterns of gene expression. Trends in Plant Science 12: 358–367.

Soltis DE, Soltis PS, Albert VA, Oppenheimer DG, dePamphilis CW, Ma H, Frohlich MW, Theissen G. 2002. Missing links: The genetic architecture of flower and floral diversification. Trends in Plant Science 7: 22–31.

Soltis PS, Soltis DE. 2004. The origin and diversification of angiosperms. American Journal of Botany 91: 1614–1626.

Sridhar VV, Surendrarao A, Liu Z. 2006. *APETALA1* and *SEPALLATA3* interact with *SEUSS* to mediate transcription repression during flower development. Development 133: 3159–3166.

Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K. 2009. Variations in Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. Proceedings of the National Academy of Sciences USA 106: 4555–4560.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596–1599.

Theissen G. 2001. Development of floral organ identity: Stories from the MADS house. Current Opinion in Plant Biology 4: 75–85.

Theissen G, Saedler H. 2001. Plant biology: Floral quartets. Nature 409: 469–471.

Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter K, Saedler H. 2000. A short history of MADS-box genes in plants. Plant Molecular Biology 42: 115–149.

Theissen G, Kim JT, Saedler H. 1996. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. Journal of Molecular Evolution 43: 484–516.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The Clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876–4882.

Wagner A. 2008. Gene duplications, robustness and evolutionary innovations. BioEssays 30: 367–373.

Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: Calibrating the family tree. Proceedings of the Royal Society of London, Series B: Biological Sciences 268: 2211–2220.

Xu G, Kong H. 2007. Duplication and divergence of floral MADS-box genes in grasses: Evidence for the generation and modification of novel regulators. Journal of Integrative Plant Biology 49: 927–939.

Xu G, Ma H, Nei M, Kong H. 2009. Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. Proceedings of the National Academy of Sciences USA 106: 835–840.

Yang CH, Chou ML. 1999. *FLD* interacts with *CO* to affect both flowering time and floral initiation in *Arabidopsis thaliana*. Plant and Cell Physiology 40: 647–650.

Yang Y, Zhang F, Ge S. 2009. Evolutionary rate patterns of the gibberellin pathway genes. BMC Evolutionary Biology 9: 206.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. The Plant Cell 12: 2473–2483.

Zhao D, Yu Q, Chen C, Ma H. 2001. Genetic control of reproductive meristems. In: McManus M, Veit B eds. Annual plant reviews: Meristematic tissues in plant growth and development. Sheffield: Sheffield Academic Press. 89–142.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Genes included in this study and their functions in *Arabidopsis thaliana*.

Table S2. Comparisons of orthologous genes of two *Arabidopsis* species in patterns of molecular evolution and exon–intron structural divergence.

Fig. S1. Phylogenetic trees of *MAF2* genes from *thaliana* and *lyrata*. **A,** Neighbor-joining tree based on protein matrix. **B,** Maximum likelihood tree based on protein matrix. **C,** Neighbor-joining tree based on coding DNA matrix. **D,** Maximum likelihood tree based on coding DNA matrix. Numbers close to each branch indicate bootstrap values >50%.

Fig. S2. Comparison of the exon–intron structure of *AtAGL15* and *AlAGL15*. **A,** Schematic representations of exon–intron structural divergence. **B,** Alignment for the eighth exon of *AtAGL15* with its counterpart and the flanking downstream intergenic region of *AlAGL15*. Bold uppercase letters denote exon sequence; lowercase letters denote intron sequence. Vertical lines indicate identical nucleotides between the two genes. Amino acid sequences are given above and below the exons. The nucleotides highlighted in red are the stop codons.

Fig. S3. Comparison of the exon–intron structure of *AtRBL* and *AlRBL*. **A,** Schematic representations of divergence in exon–intron structure. **B,** Alignment for the eighth exon and the flanking downstream intergenic region of *AtRBL* with the eighth exon of *AlRBL*.

Fig. S4. Comparison of the exon–intron structure of *AtSEP1* and *AlSEP1*. **A,** Schematic representations of divergence in exon–intron structure. **B,** Alignment for part of the sixth intron and the seventh exon of *AtSEP1* and the corresponding region of *AlSEP1*.

Fig. S5. Comparison of the exon–intron structure of *AtFLD* and *AlFLD*. **A,** Schematic representations of divergence of exon–intron structure. **B,** Alignment for the sequence from the end of the second exon to the fifth exon of *AtFLD* and that of part of the second exon and flanking downstream intergenic region of *AlFLD*.

Fig. S6. Fluctuation in ω values along each gene indicated in sliding window analyses with window length = 45 bp and step size = 15 bp. The corresponding protein structure for each gene is depicted under panels.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.