Evolution of Plant MADS Box Transcription Factors: Evidence for Shifts in Selection Associated with Early Angiosperm Diversification and Concerted Gene Duplications

Hongyan Shan,† Laura Zahn,‡[1] Stephane Guindon,§|| P. Kerr Wall,‡ Hongzhi Kong,* Hong Ma,‡ Claude W. dePamphilis,‡ and Jim Leebens-Mack†*

*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Xiangshan, Beijing, People's Republic of China; †Department of Plant Biology, University of Georgia, Athens; ‡Department of Biology and the Huck Institutes of the Life Sciences, The Pennsylvania State University; §Méthodes et Algorithmes pour la Bioinformatique (MAB), LIRMM, CNRS—Université Montpellier II, France; and ||Department of Statistics, University of Auckland, Auckland, New Zealand

Phylogenomic analyses show that gene and genome duplication events have led to the diversification of transcription factor gene families throughout the evolutionary history of land plants and that gene duplications have played an important role in shaping regulatory networks influencing key phenotypic characters including floral development and flowering time. A molecular evolutionary investigation of the mode and tempo of selection acting on the angiosperm MADS box *AP1*/*SQUA*, *AP3*/*PI*, *AG*/*AGL11*, and *SEP* gene subfamilies revealed site-specific patterns of shifting evolutionary constraint throughout angiosperm history. Specific positions in the four canonical MADS box gene regions, especially K domains and C-terminal regions of all four of these MADS box gene subfamilies exhibited clade-specific shifts in selective constraint following concerted duplication events. Moreover, the frequency of site-specific shifts in constraint was correlated with gene duplications and early angiosperm diversification. We hypothesize that coevolution among interacting MADS box proteins may be responsible for simultaneous increases in the ratio of nonsynonymous to synonymous substitutions ($d_N/d_S = \omega$) early in angiosperm history and following concerted duplication events.

Introduction

Ancient polyploidy events (whole-genome duplications) are hypothesized to have played an important role in the origin and diversification of angiosperms (e.g., Soltis and Soltis 1999; De Bodt et al. 2005; Freeling and Thomas 2006). Most if not all flowering plant lineages have experienced one or more rounds of polyploidization in their evolutionary history (e.g., Masterson 1994; Cui et al. 2006; Soltis et al. 2009). Polyploidy is an important source of gene family diversification, with the fate of duplicate genes following genome-scale duplication influenced by gene function. In general, genes involved in transcription regulation, signal transduction, and development are preferentially retained following genome duplications (Seoighe and Gehring 2004; De Bodt et al. 2005; Maere et al. 2005; but see Barker et al. 2008). For example, three ancient polyploidization events evident in the *Arabidopsis* genome account for over 90% of the duplications in *Arabidopsis* transcription factors and signal transducer gene families (Maere et al. 2005).

The duplication and evolution of transcription factors can have profound effects on genetic systems and phenotypic variation (Nei 2005; Nei and Rooney 2005; Freeling and Thomas 2006). For instance, MIKC-type MADS box transcription factors play important roles in controlling floral development and flowering time, especially in specifying the identity of floral organs. The control of floral organ identities has been characterized in the classical ABC and expanded ABCE models of floral development (Coen and Meyerowitz 1991; Pelaz et al. 2000; Theissen 2001), and all but one of the A, B, C, and E function genes identified in *Arabidopsis* are MADS box genes (reviewed in Kaufmann et al. 2005). These include *APETALA1* (*AP1*, A function), *APETALA3* and *PISTILLATA* (*AP3* and *PI*, B function), *AGAMOUS* (*AG*, C function), and *SEPALLATA1*/*2*/*3*/*4* (*SEP1*/*2*/*3*/*4*, E function; formerly named *AGL2*, *AGL4*, *AGL9*, and *AGL3*, respectively). Specific combinations of proteins coded by these genes have been hypothesized to interact as tetrameric regulatory complexes to initiate and maintain the identity of specific floral organs (Theissen and Saedler 1999, 2001; Theissen 2001). The "balance gene drive" hypothesis predicts selection for dosage balance will result in the retention of interacting transcription factors (among other genes involved in integrated complexes or regulatory networks) following genome duplication, and the resulting duplicated regulatory modules can serve as the building blocks for evolutionary innovations (Freeling and Thomas 2006). Along these lines, the origin and early diversification of flowers may have been spurred by concerted gene duplication events and shifting patterns of selection acting on novel regulatory complexes.

Phylogenetic analyses of the MIKC-type MADS box genes sampled from a diverse set of seed plants have revealed well-defined subclades (e.g., Becker and Theissen 2003) including *AP1*/*SQUA*, *AP3*/*PI*, *AG*/*AGL11*, and *SEP* subfamilies. Interestingly, there seem to have been coincident duplication events across these subfamilies suggesting codiversification, perhaps due to polyploidy. Whether through polyploidy or not, gene duplications in the lineage leading to the most recent common ancestor (MRCA) of all extant angiosperms gave rise to *AP3* and *PI* lineages in the *AP3*/*PI* subfamily, *AG* and *AGL11* lineages in the *AG*/*AGL11* subfamily, and *AGL2*/*3*/*4* and *AGL9* lineages in the *SEP* subfamily. Another set of apparently simultaneous gene duplications occurred later, in the MRCA of core eudicots giving rise to eu*AP1*, eu*FUL*, and *AGL79* clades in the *AP1*/*SQUA* subfamily; eu*AP3* and *TM6* clades in the *AP3* lineage; eu*AG* and *PLE* clades

[1] Present address: The American Association for the Advancement of Science (AAAS), NW, Washington, DC.

Key words: MADS box genes, angiosperms, gene duplication, molecular evolution.

E-mail: jleebensmack@plantbio.uga.edu.

Mol. Biol. Evol. 26(10):2229–2244. 2009
doi:10.1093/molbev/msp129
Advance Access publication July 3, 2009

in the *AG* lineage; and *AGL2/4*, *AGL3*, and *FBP9* clades in the *AGL2/3/4* lineage (supplementary figs. S1–S7, Supplementary Material online; Kramer et al. 1998, 2004; Litt and Irish 2003; Kim et al. 2004; Zahn, Kong, et al. 2005, Zahn et al. 2006; Shan et al. 2007; Xu and Kong 2007). Coincident MADS box gene duplications also map back to the MRCA of all grasses (Poaceae), giving rise to *OsMADS14* and *OsMADS15* clades within the *AP1/SQUA* subfamily; *OsMADS2* and *OsMADS4* clades within the *PI* lineage; *OsMADS13*, *OsMADS3*, and *OsMADS58* clades within the *AG* lineage; *OsMADS34*, *OsMADS1*, and *OsMADS5* clades within the *AGL2/3/4* lineage; and *OsMADS7* and *OsMADS8* clades within the *AGL9* lineage (supplementary figs. S1, S4, S6, and S7, Supplementary Material online; Zahn, Kong, et al. 2005; Preston and Kellogg 2006; Whipple et al. 2007; Xu and Kong 2007).

Divergence in expression and/or function of duplicate genes has long been hypothesized to play a fundamental role in organismal evolution (Ohno 1970; Force et al. 1999). Following duplication, one paralog may be silenced due to the accumulation of deleterious mutations (nonfunctionalization), or both may be retained if ancestral functions are split between the duplicates (subfunctionalization), or one of the paralogs takes on novel function (neofunctionalization) (Force et al. 1999; Lynch and Force 2000; Moore and Purugganan 2005). Whereas changes in regulatory, noncoding sequences have been hypothesized to drive subfunctionalization and neofunctionalization (e.g., Force et al. 1999), initially neutral or adaptive amino acid substitutions can also cause differentiation in function of duplicate genes (e.g., Zhang et al. 2002; Barkman 2003; Barkman et al. 2007). Signatures of reduced constraint or positive selection associated with gene duplications can be detected in the ratio of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S). Whereas an estimated d_N/d_S (ω) value near 1.0 suggests neutrality, ω greater or less than 1.0 indicates putative positive selection or purifying selection, respectively (reviewed by Yang and Bielawski 2000). For example, analysis on the *AP3/PI* gene subfamily indicated that residues that were inferred to have been fixed by positive selection are concentrated within the K domain of AP3 and PI proteins following the *AP3–PI* duplication and in the K domain and C-terminal region of the eu*AP3* lineage following the eu*AP3–TM6* duplication (Hernández-Hernández et al. 2007). Similarly, the grass *FUL1* (i.e., *OsMADS14*) and *FUL2* (i.e., *OsMADS15*) clades mapping to the gene duplication event predating the diversification of the Poaceae may have been subject to different selective pressures as evidenced by an elevated d_N/d_S ratio in the *FUL2* clade (Preston and Kellogg 2006). Whereas these and other studies indicate that the mode and strength of selection acting on MADS box genes varies across amino acid positions and evolutionary lineages (e.g., Martinez-Castilla and Alvarez-Buylla 2003; Preston and Kellogg 2006; Gascuel and Guindon 2007; Hernández-Hernández et al. 2007; Jaramillo and Kramer 2007), a comprehensive study of shifting selective constraint across angiosperm MADS box gene subfamilies is still lacking. Such an analysis is essential for understanding the role of MADS box gene evolution in the early diversification of angiosperm flowers.

In this study, we analyzed over 900 MADS box genes to investigate variation in selective constraint across sites and branches in phylogenies for the *AP1/SQUA*, *AP3/PI*, *AG/AGL11*, and *SEP* MADS box gene subfamilies. Codon evolution was modeled as a Markov-modulated Markov process with three d_N/d_S classes (ω_1, ω_2, and ω_3), and switching among rate classes (Guindon et al. 2004; Gascuel and Guindon 2007; Chapman et al. 2008). This approach takes into account the variability of selection regimes across sites and lineages. Rate ratio parameters (ω) and switching rates were estimated from the data in a maximum likelihood (ML) framework, and posterior probabilities (PPs) of each selection class were estimated for assignment of sites on each branch to rate ratio classes (Guindon et al. 2004). We used this approach 1) to test whether shifts in selective constraint occur within MADS box gene subfamilies throughout the evolution of angiosperms; 2) to identify sites and branches in each subfamily that have experienced shifts to positive or relaxed selection; 3) to determine whether the timing of shifts in site-specific selective constraint is correlated among MADS box gene subfamilies; and 4) to determine whether shifts to positive selection or relaxed constraint are associated with concerted gene duplication events. Our results suggest that *AP1/SQUA*, *AP3/PI*, *AG/AGL11*, and *SEP* subfamilies show common patterns of shifts in constraint associated with concerted duplication events and early angiosperm diversification.

Materials and Methods

Sequence Retrieval and Alignment

The amino acid sequences and corresponding coding regions of *AP1/SQUA*-, *AP3*-, *PI*-, *AG/AGL11*-, and *SEP*-like genes were retrieved through Blast searches of publicly available databases, including GenBank (http://www.ncbi.nlm.nih.gov), the TIGR transcript assembly database (Childs et al. 2007; http://plantta.tigr.org), and the PlantTribes database (Wall et al. 2008; http://fgp.huck.psu.edu/tribe.html). Multiple query sequences were used for each subfamily to obtain comprehensive samples. The resulting data sets were then screened to remove sequences shorter than 400-bp, sequences with many ambiguous base calls and putatively allelic sequences sampled from the same species with identity higher than 95% at the DNA level and no indels. Sequence sets were further trimmed to include exemplars from the best studied species if orthologs from several congeneric species were available. After removing incomplete or redundant sequences, 171 *AP1/SQUA*-like genes, 231 *AP3*-like genes, 166 *PI*-like genes, 159 *AG/AGL11*-like genes, and 174 *SEP*-like genes were included in phylogenetic and molecular evolutionary analyses for each MADS box gene subfamily. To explore the evolutionary pattern of the extensively studied *AP3/PI* subfamily, we compiled a combined *AP3* and *PI* data set including 125 *AP3*-like genes and 103 *PI*-like genes. The outgroups of each subfamily were chosen from the closest gene subfamilies or gymnosperm homologs according to previous phylogenetic studies (Kramer et al. 1998, 2004; Litt and Irish 2003; Aoki et al. 2004; Kim et al. 2004; Stellari et al. 2004; Zahn, Kong, et al. 2005, Zahn et al. 2006; Shan

et al. 2007). Detailed information about all genes included in this study was listed in supplementary tables S1–S4, Supplementary Material online, and alignments can be found at http://jlmack.plantbio.uga.edu/Shanetal09.html.

Protein sequences for each of the six data sets (*AP1/SQUA*, *AP3*, *PI*, *AP3/PI*, *AG/AGL11*, and *SEP*) were first aligned with MUSCLE 3.6 (Edgar 2004). A preliminary tree was constructed with MADS, I and K regions of the alignment and sequences were reordered on the basis of their places on this tree. After automated global alignment, local alignments of similar sequences were checked using Blast2 sequences (Tatusova and Madden 1999) and data matrices were adjusted manually (especially the C terminus) using GeneDoc (Nicholas et al. 1997). The C-terminal region is highly variable among the MADS box gene subfamilies, but relatively conserved motifs could still be identified within the subfamilies (Kramer et al. 1998, 2004; Becker and Theissen 2003; Litt and Irish 2003; Vandenbussche et al. 2003; Zahn, Kong, et al. 2005; Shan et al. 2007). Recent studies of *AP1/SQUA*, *AG/AGL11*, and *SEP* subfamilies suggest that well-aligned amino acid sites in the C-terminal region are phylogenetically informative (Zahn, Kong, et al. 2005, Zahn et al. 2006; Shan et al. 2007). In order to objectively assess alignment quality, the column scores of each amino acid site were estimated in ClustalX 1.83 (Thompson et al. 1997) and sites with a column score greater than 12 were retained for tree reconstruction (Zahn, Kong, et al. 2005; Shan et al. 2007). In the *AG/AGL11* subfamily, the inclusion of conserved C termini of grass *OsMADS13*-like genes and *AGL11*-like genes of rice and maize caused conflict in the topologies between protein- and DNA-based trees because of the high cytosine content of their cDNA sequences (Zahn et al. 2006). Therefore, as suggested by Zahn et al. (2006), only MADS, I and K regions of these genes (*TaAGL31*, *TaAGL2*, *OsMADS13*, *ZAG2*, *ZMM1*, *ZMM25*, and *Os01g0886200*) were included in the further phylogenetic reconstruction of the *AG/AGL11* subfamily in this work. Codon-based cDNA alignments corresponding to protein alignments of each subfamily were generated using the aa2dna script (http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.html).

Phylogenetic Analysis

ML analysis of phylogenetic relationships was performed on each DNA matrix with PhyML version 2.4 (Guindon and Gascuel 2003) with the general time-reversible model and optimization of the proportion of invariable sites and the gamma distribution parameter for variation in rates across variable sites. ML searches were initiated with a BIONJ tree (Guindon and Gaschuel 2003). Recent studies of *AP3*-like genes showed that *Pachysandra AP3*-like genes were placed at different positions in the *AP3* phylogenetic trees by different authors (e.g., Kramer et al. 2006; Hernández-Hernández et al. 2007). The position of *Pachysandra AP3* homologs may influence inferences about selection following the eu*AP3–TM6* gene duplication. Therefore, to gain a better understanding of the evolutionary history of the *AP3* lineage, the topology of the *AP3* gene tree was assessed with GARLI version 0.951 (Zwickl 2006; http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html) in addition to PhyML, and the *AP3* gene tree with the highest log likelihood value was used as a starting point in PhyML for bootstrap estimation. For all data sets, bootstrap analyses were performed for 1,000 replicates.

Molecular Evolutionary Analysis

To investigate the molecular evolutionary patterns of these MADS box subfamilies, we performed likelihood analyses under a nested set of codon-substitution models (M0, M3, M3 + S1, and M3 + S2) with FITMODEL version 0.5.3 (Guindon et al. 2004). Specification of models for variation in rate ratios across codons was similar to the models of Yang and Nielsen (2002). Model M0 assumes that all the sites in a sequence alignment are subject to the same selection process; thus, d_N/d_S (ω) is constant over all the sites. As implemented in FITMODEL, under the M3 model, variation in selective constraint across sites is modeled as three rate ratio classes with $\omega_1 < \omega_2 < \omega_3$. Site-specific shifts from one rate ratio class to another across gene phylogenies were investigated using the approach of Guindon et al. (2004). Switching was modeled as a time-reversible Markov process with three additional parameters: the overall rate of interchange among rate ratio classes (δ), a coefficient for shifts between ω_1 and ω_3 (α), and a coefficient for shifts between ω_2 and ω_3 (β). The S1 model implemented in FITMODEL imposes equal switching rates among ω_1, ω_2, and ω_3 rate ratio classes ($\alpha = \beta = 1$), and the S2 model allows α and β to vary freely accounting for unequal rates of switches between selection classes (Guindon et al. 2004). The trees and alignments used in the FITMODEL analysis were obtained as described above, but the C termini of the "problematic" grass genes in the *AG–AGL11* alignment were included in the FITMODEL analysis. ML estimates were obtained for other parameters including branch length, transition–transversion rate ratio (κ), switching parameters (δ, α, and β), substitution rate ratios (ω_1, ω_2, and ω_3), and equilibrium frequencies for sites in the three rate ratio classes (p_1, p_2, and p_3).

Nested likelihood ratio tests (LRTs) were performed for the following model comparisons: no rate heterogeneity versus variation across sites (M0 vs. M3), variation across sites without versus with switching among substitution rate ratio classes (M3 vs. M3 + S1), and equal switching rates versus class-dependent switching rates across branches (M3 + S1 vs. M3 + S2). The chi-squared test was employed to estimate the significant difference. Degrees of freedom for each test were equal to the difference in the number of parameter estimates for the models being compared. Finally, FITMODEL estimated PPs for placing a site in the highest rate ratio class (ω_3) for each branch in gene trees. Estimated PPs were visualized for each codon position with BASS (Bayesian Analysis of Selected Sites) provided by J. Huelsenbeck.

Results

Shifting Constraint in the *AP1/SQUA* Subfamily

The estimated topology of the *AP1/SQUA* tree was quite similar to that of Shan et al. (2007; fig. S1). To

Table 1
Likelihood Analysis of *AP1/SQUA*, *AP3*, *PI*, *AP3/PI*, *AG/AGL11*, and *SEP*-Like Gene Sequence Data

			M0 (No Heterogeneity)	M3 (Heterogeneity Across Sites)			M3 + S1 (Shifting Across Branches)			M3 + S2 (Unequal Switching Rates)		
AP1/SQUA												
In L			−58,583.26	−56,123.11			−55,743.82			−55,663.86		
ω_1	ω_2	ω_3	0.18	0.04	0.18	0.51	0.01	0.19	0.73	0.01	0.14	0.89
p_1	p_2	p_3	1.00	0.41	0.31	0.28	0.52	0.29	0.19	0.49	0.32	0.19
R_{12}	R_{13}	R_{23}					1.64	1.64	1.64	1.19	0.13	4.99
AP3/PI												
In L			−60,660.89	−58,880.73			−57,963.22			−57,841.06		
ω_1	ω_2	ω_3	0.14	0.03	0.13	0.32	0.00	0.18	0.61	0.00	0.08	0.73
p_1	p_2	p_3	1.00	0.31	0.42	0.27	0.50	0.32	0.18	0.39	0.41	0.20
R_{12}	R_{13}	R_{23}					1.63	1.63	1.63	0.73	0.11	4.58
AP3												
In L			−76,513.28	−74,146.39			−73,553.46			−73,387.09		
ω_1	ω_2	ω_3	0.18	0.02	0.15	0.35	0.00	0.19	0.64	0.00	0.10	0.79
p_1	p_2	p_3	1.00	0.26	0.37	0.37	0.44	0.35	0.21	0.33	0.46	0.21
R_{12}	R_{13}	R_{23}					1..57	1.57	1.57	0.70	0.002	4.09
PI												
In L			−50,971.57	−49,325.53			−48,916.41			−48,792.29		
ω_1	ω_2	ω_3	0.16	0.03	0.16	0.39	0.00	0.22	0.78	0.01	0.08	0.90
p_1	p_2	p_3	1.00	0.33	0.41	0.26	0.45	0.36	0.19	0.32	0.46	0.22
R_{12}	R_{13}	R_{23}					1.58	1.58	1.58	0.54	0.11	4.09
AG/AGL11												
In L			−50,534.61	−48,591.74			−48,246.51			−48,166.77		
ω_1	ω_2	ω_3	0.10	0.02	0.12	0.31	0.00	0.17	0.55	0.01	0.11	0.70
p_1	p_2	p_3	1.00	0.49	0.28	0.23	0.57	0.30	0.14	0.50	0.36	0.14
R_{12}	R_{13}	R_{23}					1.75	1.75	1.75	1.02	0.003	6.38
SEP												
In L			−57,923.68	−55,621.92			−55,147.22			−55,032.21		
ω_1	ω_2	ω_3	0.14	0.02	0.12	0.38	0.00	0.14	0.65	0.01	0.10	0.82
p_1	p_2	p_3	1.00	0.39	0.31	0.30	0.48	0.34	0.18	0.43	0.40	0.17
R_{12}	R_{13}	R_{23}					1.61	1.61	1.61	0.99	0.003	4.79

investigate the substitution process within the *AP1/SQUA* subfamily, we performed likelihood analyses under a nested set of codon-substitution models (M0, M3, M3 + S1, and M3 + S2) (Guindon et al. 2004). Table 1 shows that log likelihoods improved significantly as parameters were added to the nested substitution models ($P \ll 0.001$; table 2). These results suggest that M3 + S2 (unequal switching rates among three rate ratio classes) is the best codon-substitution model for the *AP1/SQUA* data set. Under this model, the substitution rate ratio estimates for three classes were $\omega_1 = 0.01$, $\omega_2 = 0.14$, and $\omega_3 = 0.89$ (table 1). The switching rate between ω_2 and ω_3 ($R_{23} = 4.99$) was significantly higher than the switching rates between ω_1 and ω_2 ($R_{12} = 1.19$) and between ω_1 and ω_3 ($R_{13} = 0.13$), implying that site-specific shifts between moderate purifying selection (ω_2) and relaxed selection (ω_3) occurred more frequently than shifts involving the most highly constrained rate ratio classes (table 1). Whereas ω_3 values approaching 1.0 are described as indicating relaxed selection, we cannot discount the possibility that the ω_3 rate ratio class includes sites that have been subject to positive selection (e.g., Hernández-Hernández et al. 2007). Functionally critical (i.e., adaptive) substitutions at sites in the ω_3 class could be identified experimentally.

To characterize variation in the propensity of codons to evolve under relaxed constraint, we assessed the number of branches in the *AP1/SQUA* gene tree for which each codon was placed in the ω_3 rate ratio class with high PPs. The alignment included 236 codons, 89 of which showed evidence of relaxed selection at some point in the history of the *AP1/SQUA* gene family (PP > 0.9; summarized in fig. 1*A*). As expected, few codon positions in MADS (2/57; 3%) and K (22/87; 25%) domains showed evidence of relaxed selection on one or more branches. In contrast, relaxed constraint

Table 2
LRTs between Different Model Comparisons

	M0 versus M3 LRTs (*P* Value)	M3 versus M3 + S1 LRTs (*P* Value)	M3 + S1 versus M3 + S2 LRTs (*P* Value)
AP1/SQUA	4,920.31 (0)	758.58 (<0.001)	159.91 (<0.001)
AP3/PI	3,560.33 (0)	1,835.02 (0)	244.31 (<0.001)
AP3	4,733.78 (0)	1,185.86 (<0.001)	332.75 (<0.001)
PI	3,292.08 (0)	818.26 (<0.001)	248.23 (<0.001)
AG/AGL11	3,885.75 (0)	690.45 (<0.001)	159.48 (<0.001)
SEP	4,603.51 (0)	949.41 (<0.001)	230.01 (<0.001)

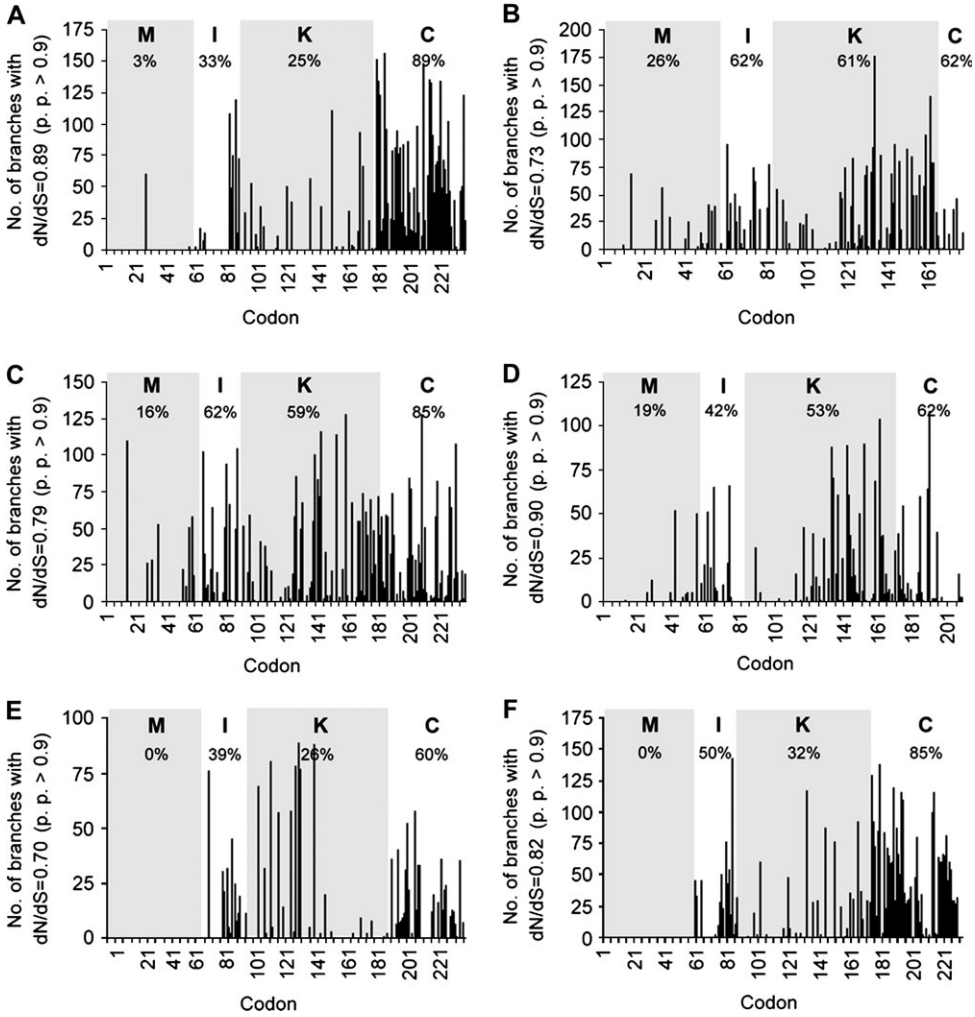


FIG. 1.—Distribution of branches with relaxed selection on each site across *AP1*/*SQUA* (*A*), *AP3*/*PI* (*B*), *AP3* (*C*), *PI* (*D*), *AG*/*AGL11* (*E*), and *SEP* (*F*) alignments, respectively. The MADS domain and the K domain are shadowed (gray). The *X*-axis represents the position of codons in the alignment, the *Y*-axis represents the number of branches exhibiting relaxed selection (PP $>$ 0.9). The number on the top of each panel denotes the percentage of sites with relaxed selection in the MADS domain, the I region, the K domain, and the C-terminal region, respectively.

was inferred on multiple branches for 10 (10/30; 33%) and 55 sites (55/62; 89%) in the I- and C-terminal regions, respectively (fig. 1*A*).

Shifts to relaxed selection at six codon positions (65, 99, 113, 194, 203, and 227; figs. 2 and 3*A*) were associated with the core eudicot duplication event, and two codons showed shifts to ω_3 following the Poaceae duplication event (at positions 205 and 207; figs. 2 and 3*A*). Among these sites, position 65 in the I region, and positions 99 and 113 in the K domain showed relaxed selection throughout the *AGL79* clade (figs. 2*A*–*C*and 3*A*). In the C-terminal region, position 194 showed both eu*AP1* and *AGL79* lineage-specific relaxation (figs. 2*D* and 3*A*); position 203 showed eu*FUL* lineage-specific relaxation (figs. 2*E* and 3*A*); position 227 showed eu*AP1* lineage-specific relaxation (figs. 2*H* and 3*A*); position 205 showed *OsMADS14* and *OsMADS15* lineage-specific relaxation (figs. 2*F* and 3*A*); and position 207 showed *OsMADS14* lineage-specific relaxation (figs. 2*G* and 3*A*). Notably, the conserved charged lysine (K) found at position 99 in the K domain of nearly all euAP1- and euFUL-like proteins of core eudicots and FUL-like proteins of basal eudicots and basal angiosperms was replaced by a hydrophobic or uncharged amino acid such as methionine (M), valine (V), leucine (L), isoleucine (I), alanine (A), serine (S), and threonine (T) within the *AGL79* clade.

The number of sites evolving under relaxed selection varied among 347 branches on the *AP1*/*SQUA* gene tree, ranging from 0 to 41 (0–17.4%; fig. 4*A*). Interestingly, branches with the largest number of sites evolving under relaxed selection were found on deep internal branches of the eu*FUL* and *AGL79* lineages, including branches leading to the *AGL79* and eu*AP1* clades, and the eu*FUL* lineage, as well as some *FUL*-like genes in the basal eudicot, Magnoliid, and monocot clades (fig. 5*A*).

Shifting Constraint in the *AP3*/*PI* Subfamily

The estimated phylogenies from *AP3*, *PI*, and *AP3*/*PI* data sets were very similar to previous studies (Kramer et al.

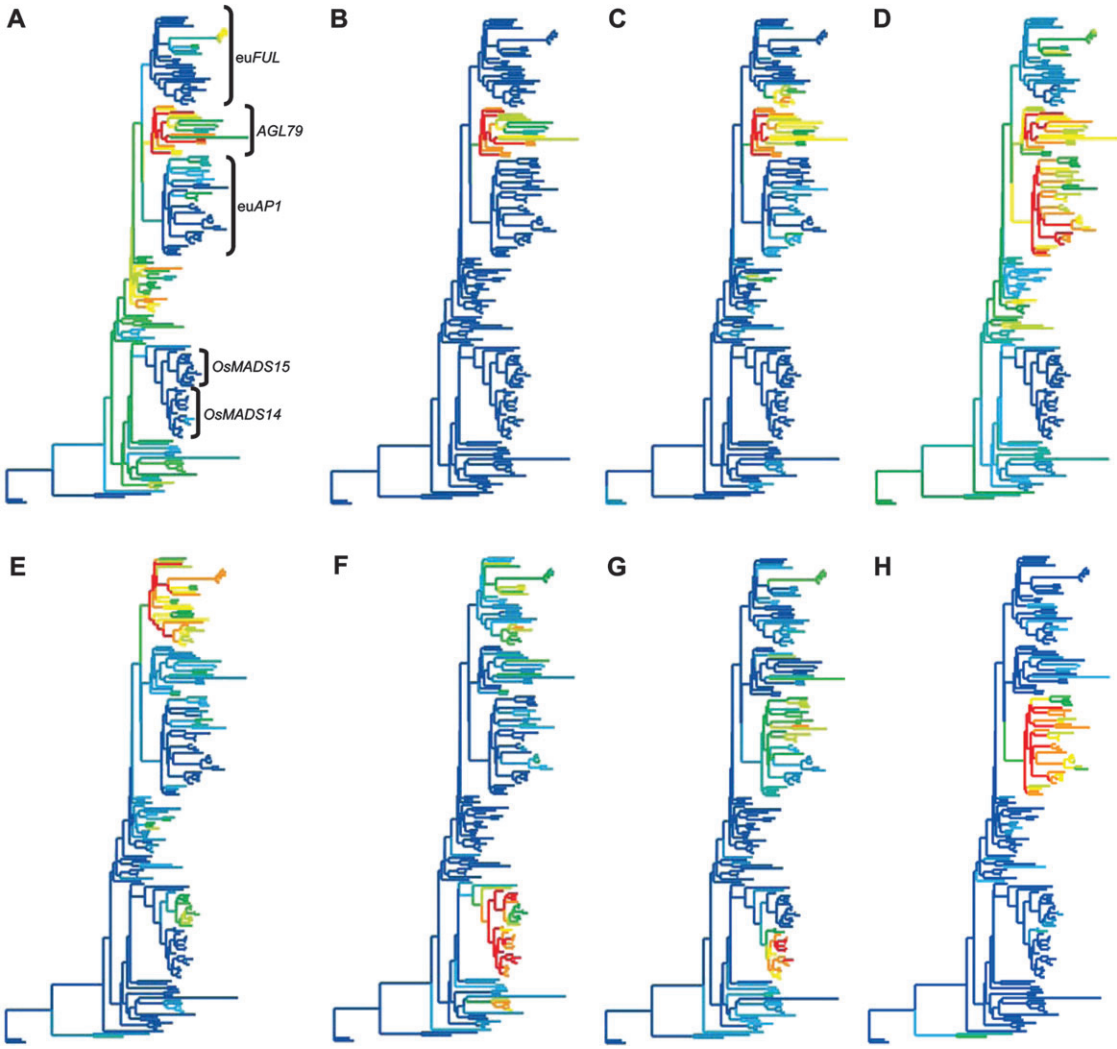


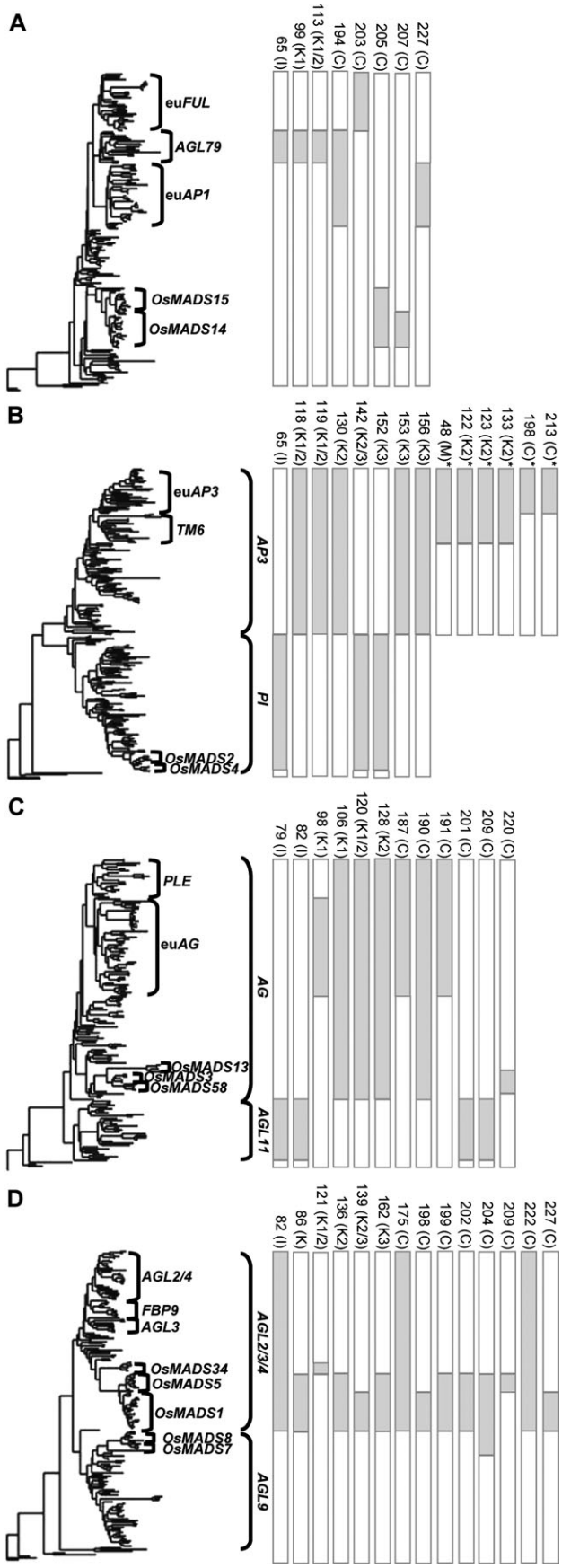
FIG. 2.—Mapping of site-specific patterns of shifting selective constraint on the *AP1/SQUA* subfamily gene tree. Branches with PP greater than 90% for selection class ω_3 are considered to have evolved under relaxed selection (red). Branches with PP for ω_3 lower than 20% are considered to be subject to purifying selection (dark blue). Branches with PP from 21% to 89% are shown with cool (light blue) to warm (orange) colors. (*A*) position 65, (*B*) position 99, (*C*) position 113, (*D*) position 194, (*E*) position 203, (*F*) position 205, (*G*) position 207, and (*H*) position 227.

1998, 2006; Aoki et al. 2004; Kim et al. 2004; Zahn et al. 2005; Hernández-Hernández et al. 2007), verifying previously inferred duplications in the *AP3/PI* gene subfamily (figs. S2–S5, Supplementary Material online, see Introduction). The primary difference between the trees generated by PhyML and GARLI was the position of *Pachysandra AP3* homologs, which grouped with the basal eudicot species *Trochodendron AP3* homologs in the PhyML tree (supplementary fig. S2, Supplementary Material online) but as sister to the core eudicot eu*AP3*-like genes in the GARLI tree (supplementary fig. S3, Supplementary Material online). We used the GARLI tree for further analyses because its log likelihood (−74,742.77) was superior to that of the PhyML tree (−74,794.17).

As was inferred for the *AP1/SQUA* subfamily, branch- and site-specific variation in evolutionary constraint within the *AP3/PI* subfamily best fit the M3 + S2 model (tables 1 and 2). The substitution rate ratios and their corresponding equilibrium frequencies in the three selection regimes under M3 + S2 suggest that heterogeneous evolution occurred during the course of the evolution of the *AP3/PI* subfamily and the paralogous *AP3* and *PI* lineages (tables 1 and 2).

The majority of codon positions in *AP3/PI*, *AP3*, and *PI* alignments were inferred to be evolving under purifying selection throughout most of their respective gene trees; the equilibrium frequency of sites in the ω_3 class was 20%, 21%, and 22% for *AP3/PI*, *AP3*, and *PI* alignments, respectively (table 1). Further, the switching rate from ω_2 to ω_3 (R_{23}) was significantly higher than switching rates from ω_1 to ω_2 (R_{12}) and from ω_1 to ω_3 (R_{13}) in all the three data sets, implying that the *AP3/PI* subfamily members underwent shifts among selection regimes similar to those inferred for the *AP1/SQUA* subfamily (table 1, fig. 5).

Unlike the *AP1/SQUA*, *AG/AGL11*, and *SEP* subfamilies (see below), analysis of the AP3/PI subfamily indicated relaxed selection on at least one branch was inferred for 15 codons in the MADS domain (fig. 1*B*). This pattern also held for the I region (16 codons)



and the K domain (51 codons) (fig. 1*B*). Only eight of 90 codon positions showing support for shifts to relaxed selection coincided with the *AP3–PI* duplication event (supplementary fig. S8, Supplementary Material online). Moreover, there were distinct patterns of shifting selection within the *AP3* and *PI* subclades. Positions 118, 119, 130, 153, and 156 in the K domain exhibited relaxed constraint within the *AP3* clade relative to the rest of the tree (fig. 3*B*, supplementary figs. S8B–S8D, S8G, and S8H, Supplementary Material online). In contrast, position 65 in the I region and positions 142 and 152 in the K domain are highly constrained in the *AP3* clade but exhibited relaxed selection in the *PI* clade (fig. 3*B*, supplementary figs. S8A, S8E, and S8F, Supplementary Material online). For example, the ancestral, positively charged lysine at position 65 (K65) was replaced by hydrophobic or uncharged amino acids, such as valine (V), isoleucine (I), leucine (L), proline (P), threonine (T), and serine (S) within the *PI* lineage.

In the *AP3* alignment, 9, 16, 51, and 46 codons showed relaxed selection in MADS, I, K, and C-terminal regions, respectively (fig. 1*C*), whereas these domains exhibited relaxed selection at 11, 11, 46, and 24 codons, respectively, in the *PI* alignment (fig. 1*D*). Six codon positions showing support for reduced constraint were associated with the eu*AP3–TM6* duplication event and included position 48 in the MADS domain, positions 122, 123, and 133 in the K domain, which showed relaxed selection on branches throughout eu*AP3* and *TM6* clades (fig. 3*B*, supplementary figs. S9A–S9D, Supplementary Material online). In contrast, shifts to relaxed selection were observed at positions 198 and 213 in the C terminus, but these were restricted to the eu*AP3* clade (fig. 3*B*, supplementary figs. S9E and S9F, Supplementary Material online). Moreover, R122 and G123 were highly conserved in most of paleoAP3-like proteins of basal eudicots and basal angiosperms but showed divergence in core eudicot euAP3 and TM6-like proteins. No codon position showed significant evidence of relaxed selection in association with the *OsMADS2–OsMADS4* duplication event within the *PI* lineage.

Figure 4*B–D* shows that the number of sites evolving under reduced selection varied across branches of *AP3*/*PI*, *AP3*, and *PI* gene trees. Moreover, shifts to relaxed selection were mainly concentrated on internal nodes (fig. 5*B–D*). Branches with the largest number of sites showing relaxed selection distributed either on deep nodes spanning the whole *AP3* gene tree or on internal branches of basal and core eudicots in the *PI* gene tree (fig. 5*C* and *D*). Among

←

FIG. 3.—Sites exhibiting shifting selective constraint following major duplication events in *AP1*/*SQUA* (*A*), *AP3*/*PI* (*B*), *AG*/*AGL11* (*C*), and *SEP* (*D*) subfamilies. Lineages emanating from these gene duplications are bracketed. Gray bars indicate that relaxed selection was inferred (PP greater than 90% for selection class ω_3) for the majority of branches within bracketed clades, whereas white bars indicate that purifying selection (ω_1 or ω_2) was inferred for the majority of lineages. Codon position within each alignment is indicated above each bar. Site numbers with asterisks in (*B*) refer to the *AP3* alignment (not the *AP3*/*PI* alignment). Domain membership is also indicated—M, MADS domain; I, I region; K1, K2, and K3, regions in the K domain (Yang et al. 2003); K1/2 and K2/3, regions between K1 and K2 or K2 and K3 subdomains; C, C-terminal region.

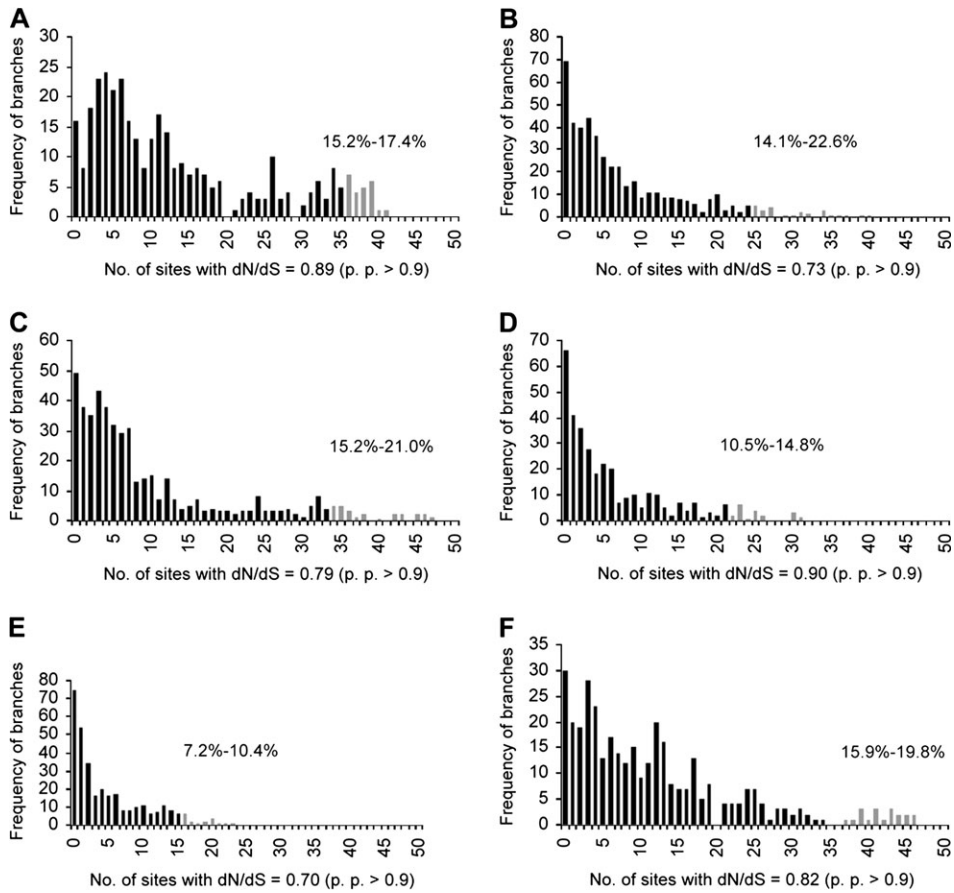


FIG. 4.—Distribution of branches with different numbers of sites under relaxed selection across *AP1*/*SQUA* (*A*), *AP3*/*PI* (*B*), *AP3* (*C*), *PI* (*D*), *AG*/*AGL11* (*E*), and *SEP* (*F*) gene trees. Gray bars indicate the 95th percentile for the number of sites on each branch exhibiting relaxed selection. The *X*-axis represents the number of sites with relaxed selection; the *Y*-axis represents the frequency of branches. The number on each panel indicates the percentage of sites with relaxed selection in the entire alignment.

these branches, only the branch leading to the eu*AP3* lineage was subject to relaxed selection at many sites following the eu*AP3–TM6* gene duplication (fig. 5*C*). The branch subtending the eu*AP3* and *TM6* clades also exhibited relaxed selection at many codon positions, suggesting that the ancestral gene of eu*AP3* and *TM6* may have been evolving under weak constraint, but experienced purifying selection within both clades following the eu*AP3–TM6* gene duplication. No such shift in selection was observed following the *AP3–PI* duplication event (fig. 5*B*) nor the *OsMADS2–OsMADS4* duplication event (fig. 5*D*).

Shifting Constraint in the *AG*/*AGL11* Subfamily

The estimated *AG*/*AGL11* gene tree was similar to that published by Zahn et al. (2006) but with improved resolution because of the increased sampling density (supplementary fig. S6, Supplementary Material online). As with the *AP1*/*SQUA* and *AP3*/*PI* subfamilies, the M3 + S2 model provided the best fit for the *AG*/*AGL11* data set (tables 1 and 2). The estimated equilibrium frequency of sites evolving under relaxed selection was slightly lower for the *AG*/*AGL11* subfamily relative to the others ($\omega_3 = 0.7$; $p_3 = 0.14$; table 1). No sites with branches showing relaxed selection were observed in the MADS domain, but 11 I-region codons, 23 K-domain codons, and 30 C-terminal codons exhibited at least one branch evolving under relaxed selection (fig. 1*E*). Long-term shifts to the ω_3 selection class were associated with concerted gene duplication events at just 12 codon positions (fig. 3*C*, supplementary fig. S10, Supplementary Material online).

As shown in figure 4*E*, the number of sites with relaxed selection varied across branches of the *AG*/*AGL11* gene tree but within a narrower range than observed for other subfamilies (from 0 to 23). Only 18 branches with 16 or more sites evolving under reduced constraint were inferred (fig. 4*E*). These branches distributed among deep nodes within the core eudicot eu*AG* lineage, *AG*-like genes of basal eudicots and magnoliids, and core eudicot *AGL11*-like genes. Only the branch leading to the eu*AG* lineage was subject to relaxed selection at many sites following the major duplication event resulting in the *PLE* and eu*AG* clades (fig. 5*E*).

Shifting Constraint in the *SEP* Subfamily

The estimated *SEP* subfamily phylogeny (supplementary fig. S7, Supplementary Material online) largely agreed with previous studies (Zahn, Kong, et al. 2005). One exception was the poorly supported placement of the *Amborella*

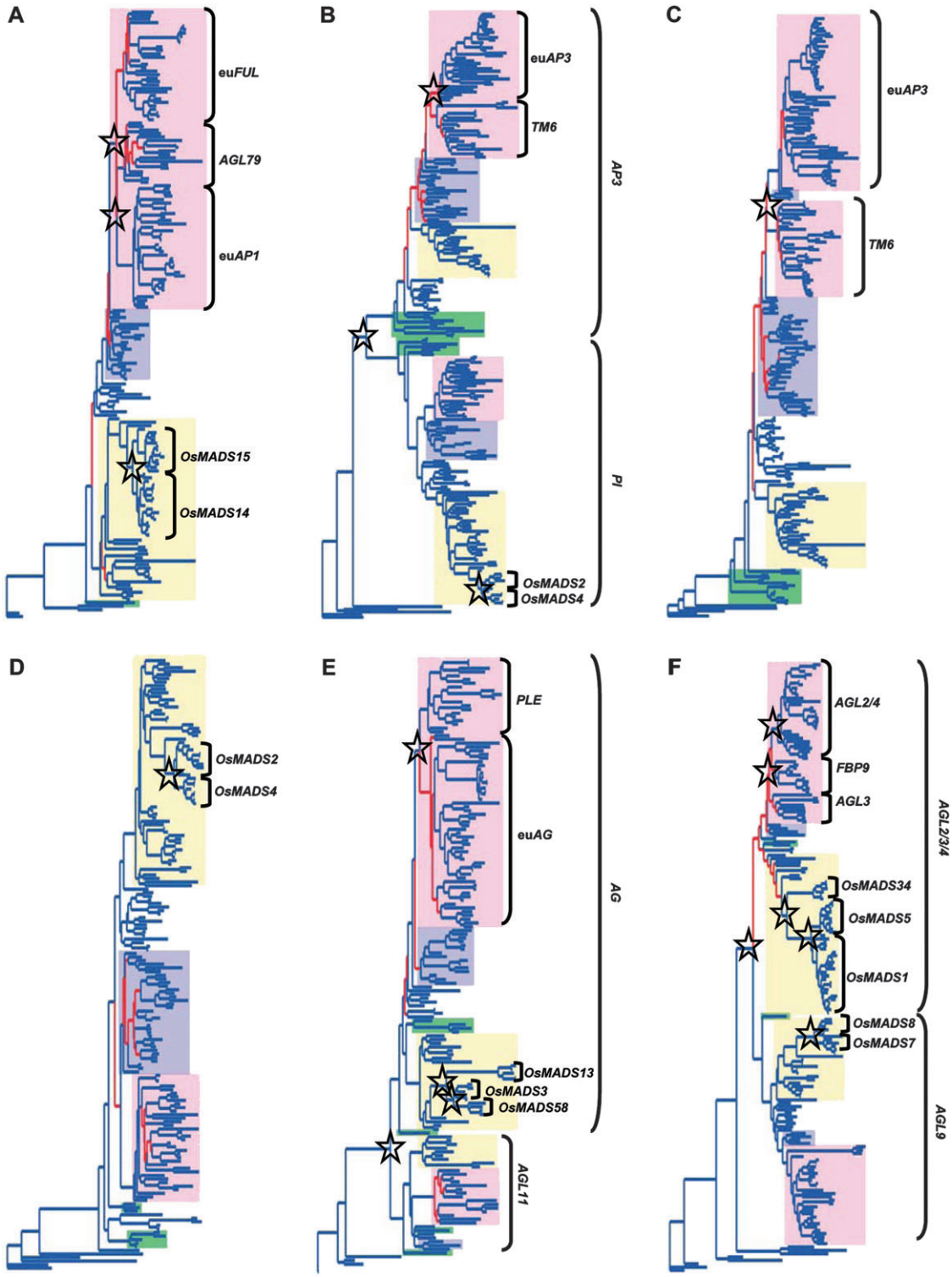


FIG. 5.—The switching pattern of selection regimes across all sites on branches of *AP1*/*SQUA* (*A*), *AP3*/*PI* (*B*), *AP3* (*C*), *PI* (*D*), *AG*/*AGL11* (*E*), and *SEP* (*F*) gene trees. Branches that evolved under relaxed selection are shown in red. Stars correspond to hypothesized concerted gene duplication events before the origin of core eudicots, grasses, and angiosperms, respectively. Brackets denote the major gene lineages. Background colors represent major angiosperm lineages: pink, core eudicots; blue, basal eudicots (ranunculids); yellow, monocots; and green, basalmost angiosperms including Amborellaceae, Nymphaeaceae, and Austrobaileyales. The unboxed ingroup genes are from the Magnoliids and Chloranthaceae.

and *Nuphar AGL2/3/4* homologs with the monocot clade rather than at the base of the *AGL2/3/4* clade (supplementary fig. S7, Supplementary Material online). This placement is likely an artifact, but we used our ML gene tree for the molecular evolutionary analyses. The results of LRTs were similar to those for other gene subfamilies (tables 1 and 2). Whereas shifts in selection were not inferred for any sites within the MADS domain, 14 I-region codons,

28 K-domain codons, and 51 C-terminal codons were inferred as evolving under relaxed selection on one or more branches (fig. 1*F*). Fourteen codon positions showing support for lasting reduced constraint following major gene duplications were observed (fig. 3*D*, supplementary fig. S11, Supplementary Material online). For instance, a shift to relaxed selection in the *AGL2/3/4* clade immediately following the *AGL2/3/4–AGL9* duplication was seen at position 82 in the I region and positions 175 and 222 in the C-terminal region (fig. 3*D*, supplementary figs. S11A, S11G, and S11M, Supplementary Material online). Relaxed selection was inferred within the *OsMADS1* and *OsMADS5* clades at positions 86, 136, and 162 in the K domain and positions 199 and 202 in the C-terminal region (fig. 3*D*, supplementary figs. S11B, S11D, S11F, S11I, and S11J, Supplementary Material online). Reduced constraint was, however, limited to the *OsMADS34* clade for position 121 in the K domain (fig. 3*D*, supplementary fig. S11C, Supplementary Material online) following the *OsMADS1/5–OsMADS34* gene duplication event.

In the *SEP* gene tree, the number of sites placed in the $\omega_3 = 0.82$ rate ratio class (PP ≥ 0.9) varied from 0 to 46 across branches (fig. 4*F*). Branches with the most codon positions evolving under reduced constraint (fig. 4*F*) were all in the *AGL2/3/4* clade and most were basal to the core eudicot triplication and Poaceae gene duplication events (fig. 5*F*). In contrast, all branches within the *AGL9* clade had fewer codon positions evolving under relaxed selection. This pattern is similar to the distribution of least conserved branches in the *AP3/PI* gene tree (fig. 5*B*).

Discussion

Molecular evolutionary analyses provide a powerful approach for identifying amino acid changes that may be associated with evolution of gene function. The results described above reveal some common themes in the evolution of MADS box genes involved in floral development. As has been predicted (Ohno 1970; Force et al. 1999), shifts to reduced selective constraint at many sites in ancestral MADS box genes were inferred following some (but not all) concerted gene duplications (fig. 3). Shifts in constraint were also inferred on deep branches within the analyzed gene trees in association with the diversification of floral form across major angiosperm lineages (fig. 5). The inferred reduction in selective constraint early in angiosperm history and following duplication events implies increased allelic diversity at these points in time and increased opportunity for novel gene interactions. We hypothesize that the shared timing of shifts in site-specific selective constraint among MADS box gene families is due in part to coevolution of interacting proteins coded by *AP1/SQUA*, *AP3/PI*, *AG/AGL11*, and *SEP* genes. Future work will test this hypothesis experimentally.

Site-Specific Shifts in Constraint Evident throughout MADS Box Gene Evolution

The MADS box genes belonging to *AP1/SQUA*, *AP3/PI*, *AG/AGL11*, and *SEP* subfamilies, play important roles in the development and evolution of angiosperm flowers. Similar to other studies, our results indicate that purifying selection has played an important role in the evolution of these MADS box gene subfamilies throughout most of angiosperm history, but there have been branch- and site-specific shifts in selection within each MADS box subfamily (e.g., Gascuel and Guindon 2007; Hernández-Hernández et al. 2007). In contrast to the patterns of positive selection of *AP3/PI*-like genes reported by Hernández-Hernández et al. (2007), however, we did not detect positive selection on any branches or at any sites within MADS box subfamilies.

The failure of FITMODEL to detect positive selection in our analyses may be due to a lack of statistical power. The widely used branch-site tests of Yang and Nielsen (2002) may offer more statistical power to detect changes in selection as it requires a constrained alternative hypothesis to compare predefined "foreground" and "background" branches (e.g., Martinez-Castilla and Alvarez-Buylla 2003; Aagaard et al. 2006; Preston and Kellogg 2006; Hernández-Hernández et al. 2007). However, variation in the mode and strength of selection on "background" branches is typically uncharacterized, and may influence inferences about selection on "foreground" branches as has been shown for the branch model (Nunney and Schuenzel 2006). Such variation should be expected in MADS box gene subfamilies with complicated histories. In addition to major concerted duplications (see Introduction), many more recent, subfamily-specific gene duplications and shifting expression patterns among duplicate genes may have prompted shifts in selective constraint. It is not clear how such shifts in selective constraint on "background" affect the widely used branch-site tests for positive selection.

A comparison of our FITMODEL results with those reported by Hernández-Hernández et al. (2007) for the CODEML branch-site tests is somewhat illuminating. Hernández-Hernández et al. (2007) found intriguing evidence for positive selection acting on sites in the functionally important K domain following the *AP3–PI* (PI86N, PI127A, and AP3115E) and eu*AP3–TM6* duplications (AP399R, AP3112C, and AP3141K; positions correspond to *Arabidopsis* AP3 [Genbank G.I. 15232493] and PI [GenBank G.I. 15241299] proteins). As described above, we did not find evidence for positive selection ($\omega_3 > 1.0$), but it is possible that sites placed in the ω_3 class have indeed experienced positive selection. More puzzling is the observation that whereas FITMODEL uncovered evidence for shifting selective constraint at sites identified in the CODEML branch-site analyses, PPs for these sites being in the ω_3 rate ratio class following the *AP3–PI* and eu*AP3–TM6* duplications were less than 0.9 (supplementary figs. S8 and S9, Supplementary Material online). It is unclear whether the discrepancies between the results of the FITMODEL and CODEML analysis are due to lack of statistical power in the FITMODEL analysis or the fact that the CODEML branch-sites test does not account for variation among background branches.

The FITMODEL analysis also implicated a number of shifts in constraint following *AP3–PI* (PI142I, PI152M, AP3118I, AP3119Q, AP3130N, AP3157Q, and AP3160I) and eu*AP3–TM6* (AP348F, AP3122R, AP3123R, AP3133K, AP3198R, and AP3213P) duplications that were

not elucidated in the CODEML analysis (supplementary figs. S8 and S9, Supplementary Material online). Most of these sites show patterns of variation in constraint that are consistent with the clade model implemented in CODEML (Bielawski and Yang 2004), but this model was not considered by Hernández-Hernández et al. (2007). Focusing on the *AP3* tree, a number of sites show an interesting shift to relaxed selection in association with the *euAP3–TM6* duplication followed by a shift back to high constraint in the lamiid *AP3* clade (supplementary fig. S9A, B, C, E, F, and I, Supplementary Material online). We do not know of any reason why this pattern would have been hypothesized in advance as is required for the CODEML clade test, and these results underscore the utility of exploratory molecular evolutionary analyses for uncovering intriguing patterns worthy of experimental investigation.

Our study indicates that sites with branches showing relaxed selection are not distributed evenly across alignments of these gene subfamilies. For instance, fewer (*AP1*/*SQUA* and *AP3*/*PI* subfamilies) and no (*AG*/*AGL11* and *SEP* subfamilies) shifts to relaxed constraint were identified in the MADS domain. Greater than 60% of sites with at least one branch showing relaxed selection mapped to the C-terminal region (fig. 1). These findings were consistent with previous analyses of variation in the level of constraint among domains in MIKC-type MADS box genes (Becker and Theissen 2003; De Bodt et al. 2003; Kaufmann et al. 2005; Nam et al. 2005).

Long-term shifts to reduced selective constraint immediately following concerted gene duplication events were identified at a few positions in the four regions, especially the K domain and the C-terminal region within each subfamily (fig. 3). The shifting levels of selective constraint elucidated in our analysis provide insights into the functional importance of specific amino acid residues. Notably, one of the sites that we inferred exhibiting switch in selective constraint, 142I in the PI lineage, has been demonstrated to be functionally important for *Arabidopsis PI* gene. When the isoleucine (I) at position 142 of PI was replaced with a proline (P), the strength of interaction between the mutated PI protein and the wild-type AP3 protein is only 40–50% of wild-type level and 35S::PI^{I142P} transgenic plants show very weak floral organ identity conversion (Yang et al. 2003). Moreover, the PI–SEP3 interaction is also affected when position 142 of PI is mutated. The strength of the PI^{I142P}–SEP3 interaction and PI^{I142F}–SEP3 interaction is about 20% and 66% of wild-type PI-SEP3, respectively (Yang and Jack 2004). In addition to 142I of PI, many amino acid substitutions within the *Arabidopsis* AP3 and PI K domains have been shown to reduce protein–protein interactions and result in defects of floral phenotypes in transgenic *Arabidopsis* lines expressing mutated B genes (Yang et al. 2003; Yang and Jack 2004). A similar experimental approach could be used to test whether site-specific changes in selective constraint that we inferred within the *AP1*/*SQUA*, *AP3*/*PI*, *AG*/*AGL11*, and *SEP* gene trees are associated with changes in DNA or protein-binding capacity. For example, position 99 in the *AP1*/*SQUA* alignment, positions 119, 130, and 156 in the *AP3*/*PI* alignment, positions 122 and 123 in the *AP3* alignment, position 128 in the *AG*/*AGL11* alignment, and positions 86, 121, and 162 in the *SEP* alignment exhibited radical amino acid substitutions in the K domain following the gene duplications (fig. 3, see Results).

Patterns of Sequence Evolution Support Functional Differences between Subfamilies

The AP1/SQUA, AP3/PI, AG/AGL11, and SEP-like proteins are hypothesized to function as master regulators in the floral development by homo- or heterodimerization in *Arabidopsis* (Fan et al. 1997; Honma and Goto 2001; Pelaz et al. 2001; Favaro et al. 2003; Yang et al. 2003; Fornara et al. 2004; de Folter et al. 2005). Similar interactions have also been documented in *Antirrhinum* (Davies et al. 1996, 1999; Egea-Cortines et al. 1999; Causier et al. 2003), tomato (Busi et al. 2003; de Martino et al. 2006; Leseberg et al. 2008), *Petunia* (Favaro et al. 2002; Kapoor et al. 2002; Ferrario et al. 2003; Immink et al. 2003; Vandenbussche et al. 2004), rice (Moon et al. 1999; Lim et al. 2000; Favaro et al. 2002; Cooper et al. 2003; Lee et al. 2003; Fornara et al. 2004), and several basal eudicot species (*Aquilegia vulgaris*, *Akebia trifoliata*, *Euptelea pleiospermum*, and *Pachysandera terminalis*) (Shan et al. 2006; Kramer et al. 2007; Liu C, Zhang J, Zhang N, Shan H, Su K, Zhang K, Meng Z, Kong H, Chen Z, submitted). These findings suggest that the interaction behaviors among AP1/SQUA, AP3/PI, AG/AGL11, and SEP proteins may be conserved over much of angiosperm evolution. At the same time, gene duplication events in multiple MADS box gene subfamilies have produced novel lineages that have been maintained through angiosperm history (Zahn, Kong, et al. 2005).

The typically conserved MADS domain is essential for dimerization and DNA binding (Huang et al. 1996; Mizukami et al. 1996). We found that the MADS domain of *AP3*/*PI* subfamily members have experienced less constraint than the corresponding domains in other subfamilies. In the MADS domain, 26%, 16%, and 19% of sites in *AP3*/*PI*, *AP3*, and *PI* alignments, respectively, exhibited shifts in selection on at least one branch, in contrast to 0–3% in the MADS domain of members of other subfamilies. It is known that AP3 and PI form obligate heterodimers in *Arabidopsis* (Riechmann et al. 1996), whereas AP1/SQUA-, AG/AGL11-, and SEP-like proteins function as homodimers or heterodimers among them (Huang et al. 1996; Mizukami et al. 1996; de Folter et al. 2005). It is reasonable to speculate that changes in one subunit of a heterodimer can promote selection for changes in the other subunit, analogous to complementary changes in DNA or RNA duplexes. Consistent with this hypothesis, inferred shifts in selection were concentrated early in eudicot history for both *AP3* and *PI* MADS box subfamilies (fig. 5*C* and *D*).

The K domain of MADS-box proteins forms an interaction surface containing three amphipathic α-helices essential for the interaction with other MADS box protein dimers (Fan et al. 1997; Yang et al. 2003). We found that the K domain in the *AP3*/*PI* subfamily included 53–61% of the sites showing shifts to relaxed selection on at least one branch, compared with only 25–32% such K-domain sites of the other three subfamilies. As described above, this pattern might be explained by the fact that eudicot AP3 and PI proteins must form heterodimers.

Among the four subfamilies, the *AG*/*AGL11* subfamily had the smallest number of branches with 16 or more sites evolving under relaxed selection, suggesting that members of this subfamily are more highly conserved than others. Indeed, it is known that functions of *AG*/*AGL11* homologs in controlling the development of stamens, carpels, and ovules tend to be conserved where functional comparisons were performed, even for gymnosperm homologs (Ma and dePamphilis 2000; Zahn et al. 2006), whereas functions of *AP1*/*SQUA* and *SEP* homologs are thought to be more variable (e.g., Uimari et al. 2004). Within the *SEP* subfamily, the fact that relaxed selection was inferred for more sites and branches within the *AGL2*/*3*/*4* clade relative to the *AGL9* clade suggests that potentially functional overlap between the *AGL2*/*3*/*4* paralogs (Flanagan and Ma 1994; Mandel and Yanofsky 1995) might result in reduced selection pressure on duplicates.

Shifts of Selection Are Associated with Functional Divergence after Speciation and Duplication

Whereas our FITMODEL analyses did not detect evidence of positive selection, shifts to nearly neutral rate ratio classes were inferred throughout MADS box subfamily gene trees. One may expect relaxed constraint at some sites following gene duplications (Ohno 1970), but we found that branches exhibiting shifts in constraint at many sites were only weakly associated with gene duplication events in the *AP3*/*PI* and other MADS box gene subfamilies. Most branches showing many site-specific shifts to reduced constraint tended to be concentrated on deeply internal branches of MADS box subfamily gene trees coincident with the origin of the basal eudicot, monocot, and magnoliid lineages (fig. 5). These changes were associated with speciation events early in angiosperm history, rather than gene duplications. Considering the extreme variability in floral form and development among ancient angiosperm lineages (Endress 2001), we hypothesize that branches showing relaxed selection on the spine of MADS box subfamily trees may represent coevolution of interacting MADS box proteins in association with the early diversification and evolution of angiosperms.

Comparative genome analysis of *Arabidopsis*, poplar, grapevine, papaya, and rice has revealed many genome-duplication events specific to major evolutionary events (Wang et al. 2005; Yu et al. 2005; Jaillon et al. 2007; Ming et al. 2008; Tang et al. 2008). The core eudicot MADS box gene duplications (*euAP3*/*TM6* and *AG*/*PLE*) and triplications (*euFUL*/*AGL79*/*euAP1* and *AGL2*/*4*/*FBP9*/*AGL3*; see fig. 5) may correspond to the gamma event characterized most recently by Tang et al. (2008). Additionally, an ancient genomewide duplication has been proposed to have occurred in the common ancestor of all or most extant angiosperms excluding *Amborella* (Cui et al. 2006). The duplication of *AP3*/*PI*, *AG*/*AGL11*, and *SEP* genes may have also occurred as part of an uncharacterized genome duplication, or alternatively independent duplications may have accumulated over 170 MYA on the branch leading to the last common ancestor of all extant angiosperms (including *Amborella*; Leebens-Mack et al. 2005 for divergence time estimates). In any event, gene and whole-genome duplications have promoted the expansion and evolution of *AP1*/*SQUA*, *AP3*/*PI*, *AG*/*AGL11*, and *SEP* subfamilies and contributed to an increased complexity of regulatory network controlling floral development (Zahn, Kong, et al. 2005; Hernández-Hernández et al. 2007; Soltis et al. 2007; Veron et al. 2007). However, our results suggest that shifts of selection pressure in these MADS box gene subfamilies were not always associated with gene or whole-genome duplications (fig. 5). However, a number of branches leading to the core eudicot lineages do show shifts in selective constraint at many sites following gene duplications and triplications at the base of core eudicots. These include shifts on the branches at the base of the *AGL79* and eu*FUL* lineage (fig. 5*A*), the eu*AP3* lineage (fig. 5*C*), the eu*AG* lineage (fig. 5*E*), as well as the *AGL3* and *FBP9* lineages (fig. 5*F*). Duplications in the *AP3* subfamily at the base of the ranunculid clade (Kramer et al. 2003) are also followed by shifts in selective constraint (fig. 5*B* and *C*).

Genetic and functional analyses of floral organ identity genes in core eudicot species, such as *Arabidopsis*, *Antirrhinum*, and *Petunia*, have indicated that nearly all these genes are important for normal development of flowers. After the gene duplication event at the base of core eudicots, the novel eu*AP3*-like genes appear to have obtained novel functions in petal identity while retaining their ancestral function in determining the development of stamens (Sommer et al. 1990; Jack et al. 1992; Vandenbussche et al. 2004). Although all *Arabidopsis SEP*-like genes were believed to have redundant function, phenotypic differences between plants with mutant *sep1*/*2*/*3* and *sep1*/*2*/*3*/*4* genes showed that *SEP4* (i.e., *AGL3*) possesses at least one slightly different function from the other three *SEP* genes (Ditta et al. 2004). Similarly, eu*FUL*-, *AGL79*-, and eu*AP1*-like genes, and eu*AG*- and *PLE*-like genes also showed functional divergence following gene duplications (Zahn et al. 2006; Shan et al. 2007; and references cited therein). Considering our results along with functional data, we hypothesize that relaxed selection in core eudicots may have permitted substitutions responsible for functional divergence and have increased the complexity of interaction of MADS box proteins following the concerted duplications in a common ancestor of all core eudicots. After a relatively short evolutionary period, any modified genetic system controlling floral development could have been fixed by purifying selection. Although we did not detect positive selection, adaptive amino acid substitutions may have also played a role in functional diversification. Further, changes in MADS box gene function can also be driven by changes of regulatory elements (Moore et al. 2005; Duarte et al. 2006).

Overall, our results indicate that shifts in selective constraint acting on MADS box genes are associated with rapid diversification early in the angiosperm history and in the core eudicots. Shifts in constraint acting on MADS box genes are more strongly associated with concerted duplications (likely due to polyploidization) early in the history of core eudicots relative to the duplications that predated diversification of the angiosperm crown group. However, functional divergence and speciation are intimately related, especially in the case of whole-genome duplications. For

example, the observed time lags between the early duplications in the *AP1*/*SQUA*, *AP3*/*PI*, *AG*/*AGL11*, and *SEP* subfamilies and shifts in selective constraint (fig. 5) are consistent with the "balance gene drive" hypothesis (Freeling and Thomas 2006) that predicts changes in the function of duplicated regulatory genes may occur after a period of purifying selection to maintain dosage balance following concerted duplications (including polyploidization).

Supplementary Material

Supplementary figures S1–S11 and supplementary tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was funded through NSF DBI-0115684 grants DBI-0638595 to C.W.D., H.M., and J.L.-M. H.S. and H.K. were also supported by the National Natural Science Foundation of China (grant numbers 30530090 and 30800065). We also thank J. Huelsenbeck for providing BASS for visualization of posterior probabilities for site-specific rate ratio. We greatly appreciate the helpful comments on the manuscript provided by Todd Barkman, Brendan Davies, Ken Wolfe, and an anonymous reviewer.

Literature Cited

Aagaard JE, Willis JH, Phillips PC. 2006. Relaxed selection among duplicate floral regulatory genes in Lamiales. J Mol Evol. 63:493–503.

Aoki S, Uehara K, Imafuku M, Hasebe M, Ito M. 2004. Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. J Plant Res. 117:229–244.

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Mol Biol Evol. 25:2445–2455.

Barkman TJ. 2003. Evidence for positive selection on the floral scent gene isoeugenol-*O*-methyltransferase. Mol Biol Evol. 20:168–172.

Barkman TJ, Martins TR, Sutton E, Stout JT. 2007. Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. Mol Biol Evol. 24:1320–1329.

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

Bielawski JP, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. J Mol Evol. 59: 121–132.

Busi MV, Bustamante C, D'Angelo C, Hidalgo-Cuevas M, Boggio SB, Valle EM, Zabaleta E. 2003. MADS-box genes expressed during tomato seed and fruit development. Plant Mol Biol. 52:801–815.

Causier B, Cook H, Davies B. 2003. An *Antirrhinum* ternary complex factor specifically interacts with C-function and SEPALLATA-like MADS-box factors. Plant Mol Biol. 52:1051–1062.

Chapman MA, Leebens-Mack JH, Burke JM. 2008. Positive selection and expression divergence following gene duplication in the sunflower *CYCLOIDEA* gene family. Mol Biol Evol. 25:1260–1273.

Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP. 2007. The TIGR Plant Transcript Assemblies database. Nucleic Acids Res. 35:D846–D851.

Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. Nature. 353: 31–37.

Cooper B, Clarke JD, Budworth P, et al. (12 co-authors). 2003. A network of rice genes associated with stress response and seed development. Proc Natl Acad Sci USA. 100:4945–4950.

Cui L, Wall PK, Leebens-Mack JH, et al. (13 co-authors). 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16:738–749.

Davies B, Egea-Cortines M, de Andrade Silva E, Saedler H, Sommer H. 1996. Multiple interactions amongst floral homeotic MADS box proteins. EMBO J. 15:4330–4343.

Davies B, Motte P, Keck E, Saedler H, Sommer H, Schwarz-Sommer Z. 1999. *PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. EMBO J. 18: 4023–4034.

De Bodt S, Maere S, Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. Trends Ecol Evol. 20:591–597.

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

de Folter S, Immink RGH, Kieffer M, et al. (12 co-authors). 2005. Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. Plant Cell. 17:1424–1433.

de Martino G, Pan I, Emmanuel E, Levy A, Irish VF. 2006. Functional analyses of two tomato *APETALA3* genes demonstrate diversification in their roles in regulating floral development. Plant Cell. 18:1833–1845.

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. 2004. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Curr Biol. 14:1935–1940.

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Mol Biol Evol. 23:469–478.

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

Egea-Cortines M, Saedler H, Sommer H. 1999. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. EMBO J. 18: 5370–5379.

Endress PK. 2001. The flowers in extant basal angiosperms and inferences on ancestral flowers. Int J Plant Sci. 162: 1111–1140.

Fan HY, Hu Y, Tudor M, Ma H. 1997. Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J. 12: 999–1010.

Favaro R, Immink RG, Ferioli V, Bernasconi B, Byzova M, Angenent GC, Kater M, Colombo L. 2002. Ovule-specific MADS-box proteins have conserved protein-protein interactions in monocot and dicot plants. Mol Genet Genomics. 268: 152–159.

Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L. 2003. MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. Plant Cell. 15:2603–2611.

Ferrario S, Immink RG, Shchennikova A, Busscher-Lange J, Angenent GC. 2003. The MADS box gene *FBP2* is required for *SEPALLATA* function in petunia. Plant Cell. 15:914–925.

Flanagan CA, Ma H. 1994. Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant *Arabidopsis* flowers. Plant Mol Biol. 26:581–595.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151: 1531–1545.

Fornara F, Parenicova L, Falasca G, Pelucchi N, Masiero S, Ciannamea S, Lopez-Dee Z, Altamura MM, Colombo L, Kater MM. 2004. Functional characterization of *OsMADS18*, a member of the *AP1/SQUA* subfamily of MADS box genes. Plant Physiol. 135:2207–2219.

Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 16:805–814.

Gascuel O, Guindon S. 2007. Modeling the variability of evolutionary process. In: Gascuel O, Steel M, editors. Reconstructing evolution: new mathematical and computational advances. New York: Oxford University Press. p. 65–107.

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

Guindon S, Rodrigo AG, Dyer KA, Huelsenbeck JP. 2004. Modeling the site-specific variation of selection patterns along lineages. Proc Natl Acad Sci USA. 101:12957–12962.

Hernández-Hernández T, Martínez-Castilla LP, Alvarez-Buylla ER. 2007. 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. 24:465–481.

Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature. 409: 525–529.

Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H. 1996. DNA binding properties of two *Arabidopsis* MADS domain proteins: binding consensus and dimer formation. Plant Cell. 8:81–94.

Immink RG, Ferrario S, Busscher-Lange J, Kooiker M, Busscher M, Angenent GC. 2003. Analysis of the petunia MADS-box transcription factor family. Mol Genet Genomics. 268:598–606.

Jack T, Brockman LL, Meyerowitz EM. 1992. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. Cell. 68: 683–697.

Jaillon O, Aury JM, Noel B, et al. (56 co-authors). 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 449:463–467.

Jaramillo MA, Kramer EM. 2007. Molecular evolution of the petal and stamen identity genes, *APETALA3* and *PISTILLATA*, after petal loss in the Piperales. Mol Phylogenet Evol. 44: 598–609.

Kapoor M, Tsuda S, Tanaka Y, Mayama T, Okuyama Y, Tsuchimoto S, Takatsuji H. 2002. Role of petunia *pMADS3* in determination of floral organ and meristem identity, as revealed by its loss of function. Plant J. 32:115–127.

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.

Kim S, Yoo MJ, Albert VA, Farris JS, Soltis PS, Soltis DE. 2004. 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. 91:2102–2118.

Kramer EM, Di Stilio VS, Schlüter P. 2003. Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. Int J Plant Sci. 164(1):1–11.

Kramer EM, Dorit RL, Irish VF. 1998. Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. Genetics. 149:765–783.

Kramer EM, Holappa L, Gould B, Jaramillo MA, Setnikov D, Santiago PM. 2007. Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. Plant Cell. 19:750–766.

Kramer EM, Jaramillo MA, Di Stilio VS. 2004. Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. Genetics. 166:1011–1023.

Kramer EM, Su HJ, Wu CC, Hu JM. 2006. A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the *APETALA3* gene lineage. BMC Evol Biol. 6:30.

Lee S, Jeon JS, An K, Moon YH, Lee S, Chung YY, An G. 2003. Alteration of floral organ identity in rice through ectopic expression of *OsMADS16*. Planta. 217:904–911.

Leebens-Mack JH, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol. 22: 1948–1963.

Leseberg CH, Eissler CL, Wang X, Johns MA, Duvall MR, Mao L. 2008. Interaction study of MADS-domain proteins in tomato. J Exp Bot. 59:2253–2265.

Lim J, Moon YH, An G, Jang SK. 2000. Two rice MADS domain proteins interact with *OsMADS1*. Plant Mol Biol. 44:513–527.

Litt A, Irish VF. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. Genetics. 165:821–833.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics. 154:459–473.

Ma H, dePamphilis C. 2000. The ABCs of floral evolution. Cell. 101:5–8.

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA. 102: 5454–5459.

Mandel MA, Yanofsky MF. 1995. The Arabidopsis *AGL8* MADS box gene is expressed in inflorescence meristems and is negatively regulated by *APETALAI*. Plant Cell. 7: 1763–1771.

Martinez-Castilla LP, Alvarez-Buylla ER. 2003. Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. Proc Natl Acad Sci USA. 100:13407–13412.

Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science. 264: 421–424.

Ming R, Hou S, Feng Y, et al. (85 co-authors). 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya Linnaeus*). Nature. 452:991–996.

Mizukami Y, Huang H, Tudor M, Hu Y, Ma H. 1996. Functional domains of the floral regulator *AGAMOUS*: characterization

of the DNA binding domain and analysis of dominant negative mutations. Plant Cell. 8:831–845.
Moon YH, Kang HG, Jung JY, Jeon JS, Sung SK, An G. 1999. Determination of the motif responsible for interaction between the rice APETALA1/AGAMOUS-LIKE9 family proteins using a yeast two-hybrid system. Plant Physiol. 120:1193–1203.
Moore RC, Grant SR, Purugganan MD. 2005. Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. Mol Biol Evol. 22:91–103.
Moore RC, Purugganan MD. 2005. The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol. 8:122–128.
Nam J, Kaufmann K, Theissen G, Nei M. 2005. A simple method for predicting the functional differentiation of duplicate genes and its application to MIKC-type MADS-box genes. Nucleic Acids Res. 33:e12.
Nei M. 2005. Selectionism and neutralism in molecular evolution. Mol Biol Evol. 22:2318–2342.
Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 39: 121–152.
Nicholas KB, Nicholas HB Jr, Deerfield DWII. 1997. Genedoc: analysis and visualization of genetic variation. Embnew News. 4:14.
Nunney L, Schuenzel EL. 2006. Detecting natural selection at the molecular level: a reexamination of some "classic" examples of adaptive evolution. J Mol Evol. 62:176–195.
Ohno S. 1970. Evolution by gene duplication. New York: Springer. p. 160.
Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C foral organ identity functions require *SEPALLATA* MADS-box genes. Nature. 405:200–203.
Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF. 2001. APETALA1 and SEPALLATA3 interact to promote flower development. Plant J. 26:385–394.
Preston JC, Kellogg EA. 2006. Reconstructing the evolutionary history of paralogous *APETALA1*/*FRUITFULL*-like genes in grasses (Poaceae). Genetics. 174:421–437.
Riechmann JL, Krizek BA, Meyerowitz EM. 1996. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc Natl Acad Sci USA. 93:4793–4798.
Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet. 20:461–464.
Shan H, Su K, Lu W, Kong H, Chen Z, Meng Z. 2006. Conservation and divergence of candidate class B genes in *Akebia trifoliata* (Lardizabalaceae). Dev Genes Evol. 216: 785–795.
Shan H, Zhang N, Liu C, Xu G, Zhang J, Chen Z, Kong H. 2007. Patterns of gene duplication and functional diversification during the evolution of the *AP1*/*SQUA* subfamily of plant MADS-box genes. Mol Phylogenet Evol. 44:26–41.
Soltis DE, Leebens-Mack JHBell CD, Paterson A, Albert VA, Zheng C, Sankoff D, Soltis PS. 2009. Polyploidy and angiosperm diversification. Am J Bot. 96:336–348.
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 Plant Sci. 12:358–367.
Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol. 14:348–352.
Sommer H, Beltran JP, Huijser P, Pape H, Lonnig WE, Saedler H, Schwarz-Sommer Z. 1990. *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. EMBO J. 9:605–613.
Stellari GM, Jaramillo MA, Kramer EM. 2004. Evolution of the *APETALA3* and *PISTILLATA* lineages of MADS-box-containing genes in the basal angiosperms. Mol Biol Evol. 21: 506–519.
Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. Science. 320:486–488.
Tatusova TA, Madden TL. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett. 174:247–250.
Theissen G. 2001. Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol. 4:75–85.
Theissen G, Saedler H. 1999. The golden decade of molecular floral development (1990–1999): a cheerful obituary. Dev Genet. 25:181–193.
Theissen G, Saedler H. 2001. Floral quartets. Nature. 409: 469–471.
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 Res. 25:4876–4882.
Uimari A, Kotilainen M, Elomaa P, Yu D, Albert VA, Teeri TH. 2004. Integration of reproductive meristem fates by a *SEPALLATA*-like MADS-box gene. Proc Natl Acad Sci USA. 101:15817–15822.
Vandenbussche M, Theissen G, Van de Peer Y, Gerats T. 2003. Structural diversification and neo-functionalization during foral MADS box gene evolution by C-terminal frameshift mutations. Nucleic Acids Res. 31:4401–4409.
Vandenbussche M, Zethof J, Royaert S, Weterings K, Gerats T. 2004. The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia hybrida* flower development. Plant Cell. 16:741–754.
Veron AS, Kaufmann K, Bornberg-Bauer E. 2007. Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. Mol Biol Evol. 24: 670–678.
Wall PK, Leebens-Mack J, Müller KF, Field D, Altman NS, dePamphilis CW. 2008. PlantTribes: a gene and gene family resource for comparative genomics in plants. Nucleic Acids Res. 36:D970–D976.
Wang X, Shi X, Hao B, Ge S, Luo J. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol. 165:937–946.
Whipple CJ, Zanis MJ, Kellogg EA, Schmidt RJ. 2007. Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicules and petals. Proc Natl Acad Sci USA. 104:1081–1086.
Xu G, Kong H. 2007. Duplication and divergence of floral MADS-box genes in grasses: evidence for the generation and modification of novel regulators. J Integr Plant Biol. 49: 927–939.
Yang Y, Fanning L, Jack T. 2003. The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. 33:47–59.
Yang Y, Jack T. 2004. Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Plant Mol Biol. 55:45–59.
Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–503.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 19:908–917.
Yu J, Wang J, Lin W, et al. (117 co-authors). 2005. The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. 3:e38.
Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE, Depamphilis CW, Ma H. 2005. The

evolution of the *SEPALLATA* subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. Genetics. 169:2209–2223.

Zahn LM, Leebens-Mack J, dePamphilis CW, Ma H, Theissen G. 2005. To B or not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. J Hered. 96:225–240.

Zahn LM, Leebens-Mack JH, Arrington JM, Hu Y, Landherr LL, dePamphilis CW, Becker A, Theissen G, Ma H. 2006. Conservation and divergence in the *AGAMOUS* subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. Evol Dev. 8:30–45.

Zhang J, Zhang YP, Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat Genet. 30:411–415.

Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation. The University of Texas at Austin.

Kenneth Wolfe, Associate Editor

Accepted June 24, 2009