J Mol Evol (1997) 45:392–396

JOURNAL OF MOLECULAR EVOLUTION

The MADS-Box Floral Homeotic Gene Lineages Predate the Origin of Seed Plants: Phylogenetic and Molecular Clock Estimates

Michael D. Purugganan

Department of Genetics, P.O. Box 7614, North Carolina State University, Raleigh, NC 27695-7614, USA

Received: 31 January 1997 / Accepted: 9 April 1997

Abstract. Flower development in angiosperms is controlled in part by floral homeotic genes, many of which are members of the plant MADS-box regulatory gene family. The evolutionary history of these developmental genes was reconstructed using 74 loci from 15 dicot, three monocot, and one conifer species. Molecular clock estimates suggest that the different floral homeotic gene lineages began to diverge from one another about 450–500 mya, around the time of the origin of land plants themselves.

Key words: Floral homeotic genes — Flowers — Development — Evolution — MADS-Box

Introduction

In recent years, geneticists have managed to identify and isolate a series of genes that control flower development from the model systems *Arabidopsis thaliana*, *Antirrhinum majus*, or *Petunia hybrida* (Bowman et al. 1993; Coen 1991; Coen and Meyerowitz 1991; Weigel 1995; Yanofsky 1995). These genes, which include the *Arabidopsis AGAMOUS*, *APETALA1* and *PISTILLATA* loci, are referred to as floral homeotic genes because mutations in these loci result in the transformation of one floral organ type to another organ. Molecular analysis has demonstrated that most floral homeotic genes isolated to date belong to the MADS-box regulatory gene family. Members of this family, which are found in all eukaryotic groups, are known to encode sequence-specific DNA-binding transcriptional activators which carry out a variety of developmental functions (Ma et al. 1991; Pollock and Treismann 1991; Davies and Schwarz-Sommer 1994). In plants, genes of this family encode a protein of 240–260 amino acids in length, which includes the conserved 57-amino acid MADS-box.

Previous molecular evolutionary analyses of this regulatory gene family revealed that most plant MADS-box genes are organized into monophyletic gene groups now referred to as floral homeotic gene groups (Purugganan et al. 1995). The origins of these floral homeotic gene lineages are of great interest: Were these lineages established at the time of the origin of flowering plants or do they predate the arrival of the angiosperms? Preliminary molecular clock estimates tentatively suggested that these floral homeotic gene lineages were around prior to the origin of the angiosperms, although determination of a more precise date of origin was hampered by uncertainties surrounding the limited calibration time points previously available (Purugganan et al. 1995).

In an effort to obtain a more precise estimate for the diversification of floral homeotic gene lineages, a more thorough molecular clock analysis of the plant MADS-box regulatory gene family was undertaken. All sequences used in this analysis are available from Genbank and EMBL DNA data bases (accession number list available from the author). An alignment of plant MADS-box protein sequences was made using PILEUP of the UWGCG package with visual refinement. Sequence distance calculations were carried out using the Molecular Evolutionary Genetic Analyses package (Kumar et al. 1994). Levels of nucleotide substitutions were estimated using the method proposed by Tajima and Nei (1984). For the phylogenetic analyses, a 507-bp region encompassing the MADS-box, the I-region, and K-box was uti-

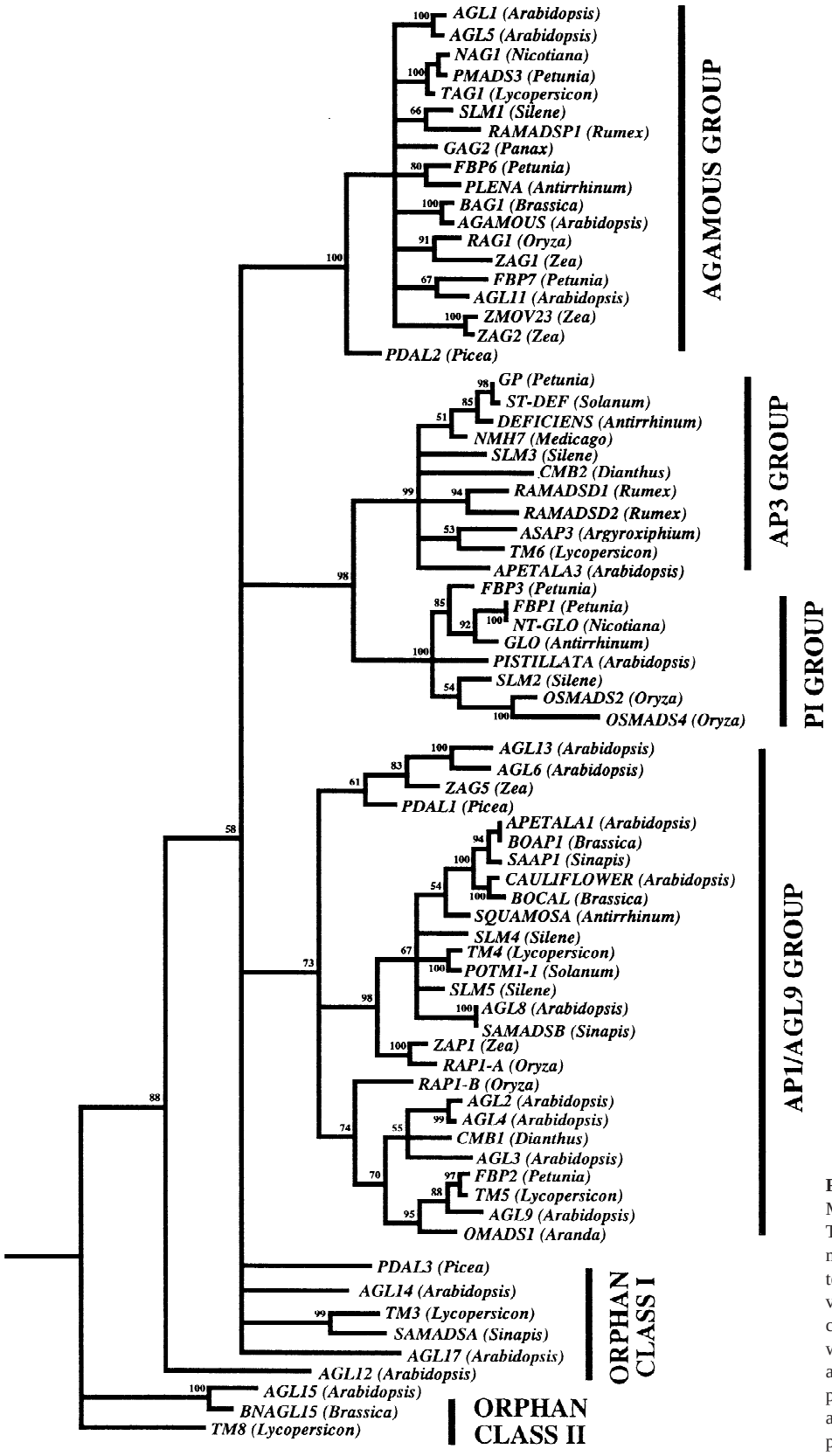

Fig. 1. Phylogeny of the plant MADS-box regulatory gene family. The phylogeny was estimated by maximum parsimony, with a consistency index of 0.27. The bootstrap values from 100 replicates are indicated on most major nodes. All nodes with less than 50% bootstrap support are collapsed on the tree. The monophyletic floral homeotic gene groups are delineated, as well as the two orphan gene classes.

lized. Phylogenetic analyses using maximum parsimony methods were undertaken using the PAUP program (Swofford 1993); only the first and second codon positions were utilized.

The phylogenetic history of the plant MADS-box regulatory gene family was reconstructed using 74 genes from 15 dicot, three monocot, and one conifer species (Fig. 1). Of the 74 genes used in this analysis, all but four

Table 1. Estimated nucleotide substitution rates at various time points

Comparison	Substitution rate	
	MIK region	MADS-Box
Rice/maize (65 mya)		
1	5.34 ± 1.16	4.28 ± 1.79
2	12.12 ± 1.91*	2.78 ± 1.40
Monocot/dicot (200 mya)		
1	6.23 ± 0.56	3.25 ± 0.68
2	7.11 ± 0.27	3.09 ± 0.60
3	4.28 ± 0.34	1.97 ± 0.15
4	6.35 ± 0.76	2.08 ± 0.64
Conifer/angiosperm (285 mya)		
1	5.30 ± 1.20	1.78 ± 0.65
2	4.56 ± 0.54	1.67 ± 0.35

Nucleotide substitution rates are given $\times 10^{10}$ substitutions/site/year at the 1st and 2nd codon positions. The comparisons are as follows: (i) rice/maize: 1, *RAP1*-A versus *ZAP1*, 2, *RAG1* versus *ZAG1;* (ii) monocot/dicot: 1, *ZAP1, RAP1-A* versus dicot *AP1* orthologues, 2, *ZAG5* versus *AGL6, AGL13,* 3, *OMADS1* versus *AGL9, TM5, FBP2,* 4, *RAG1, ZAG1, ZMOV23, ZAG2* versus dicot *AGAMOUS* orthologues; (iii) conifer/angiosperm: 1, *PDAL1* versus *ZAG5, AGL6, AGL13,* 2, *PDAL2* versus all angiosperm AGAMOUS group genes

* Significantly different at the $p > 0.05$ level

Table 2. Estimated nucleotide substitution rates for MADS-box floral homeotic gene groups

Group	Substitution rate	
	MIK Region	MADS-Box
AP1/AGL9	5.57 ± 1.02	2.87 ± 1.02
AGAMOUS	7.68 ± 3.95	2.18 ± 0.56
GLOBAL MEAN	5.60 ± 1.02	2.36 ± 0.61

Nucleotide substitution rates are given $\times 10^{10}$ substitutions/site/year at the 1st and 2nd codon positions. The global mean was calculated as the mean of substitution rates from all the calibration timepoints (see Table 1)

are members of a large monophyletic assemblage that include the major floral homeotic gene groups (the AGAMOUS, APETALA3, PISTILLATA, and APETALA1/AGL9 groups) previously identified (Purugganan et al. 1995; Doyle 1994). The existence of these floral homeotic gene groups, which contain loci with genetically defined floral homeotic functions in *Arabidopsis thaliana, Antirrhinum majus, Petunia hybrida,* or *Zea mays,* remains strongly supported in this analysis.

The topology of the plant MADS-box phylogenetic tree reveals a late boundary for the diversification of floral homeotic gene groups at 285 mya, the time of the last common ancestor of the seed plants (Tandre et al. 1995). The phylogenetic analyses, however, suggests a protracted period prior to the conifer/angiosperm split, in which the direct ancestral lineages to the extant floral homeotic gene groups were in existence. Molecular clock estimates allows us to further assess the age of these gene groups and when they began to diverge from one another, thus delineating important dates in the evolution of this regulatory gene family.

The molecular clock analysis utilized eight different calibration time points in both the AGAMOUS and AP1/AGL9 groups (Table 1). These molecular clock calibrations relied on a broad range of time points: the divergence of rice and maize (65 mya) (Crepet and Feldman 1991), monocots and dicots (200 mya) (Wolfe et al. 1989) and conifers and angiosperms (285 mya) (Savard et al. 1994). The monocot-dicot and conifer-angiosperm divergence dates are corroborated by independent molecular clock estimates at multiple loci (Wolfe et al. 1989; Savard et al. 1994). Only the first two codon positions were used, as the third position sites are saturated between distantly related loci.

Local molecular clocks were also calibrated separately for the AGAMOUS and AP1/AGL9 groups by taking the mean of various estimates for pairwise substitution rates of genes within these two clades. A combined substitution rate for the AGAMOUS and AP1/AGL9 groups (to be referred to from now on as the global molecular clock) was also estimated by taking a mean of pairwise substitutions for all genes in these two groups for which a calibration was possible. The exception to rate homeogeneity within this regulatory gene family are the members of the AP3/PI groups; it appears from maximum likelihood ratio tests that members of these two groups are evolving at about 20–40% faster than all other plant MADS-box genes (M.D. Purugganan and A. Rambaut, unpublished results). Because of this apparent rate heterogeneity and the lack of an unambiguous local AP3/PI group calibration, these two clades were excluded from all subsequent analyses.

From the pairwise sequence divergence data, the estimated mean nucleotide substitution rates for the MADS-box genes are calculated (Table 2). The substitution rate for the MIK region is 5.6×10^{-10} substitutions/site/year, while the MADS-box alone is evolving at about one-third this rate. This mean nucleotide substitution rate presumably reflects the mean substitution rates throughout the evolutionary time period being considered. This appears to be a valid assumption, since lineage-specific rate heterogeneity through time, such as uncorrelated bursts of adaptive nucleotide substitutions, might be expected to manifest itself in differences in relative rates between lineages (Gillespie 1991). For the plant MADS-box gene family, however, most genes appear to be evolving at about the same rate (Purugganan et al. 1995). There is also no significant difference in substitution rates estimated within the AG and AP1/AGL9 groups, indicating that the substitution rate is not dependent on the specific MADS-box gene used. Moreover, the substitution rates estimated from the grass, monocot/dicot and conifer/angiosperm divergence dates are, with one exception (*RAG1* versus *ZAG1*), not significantly

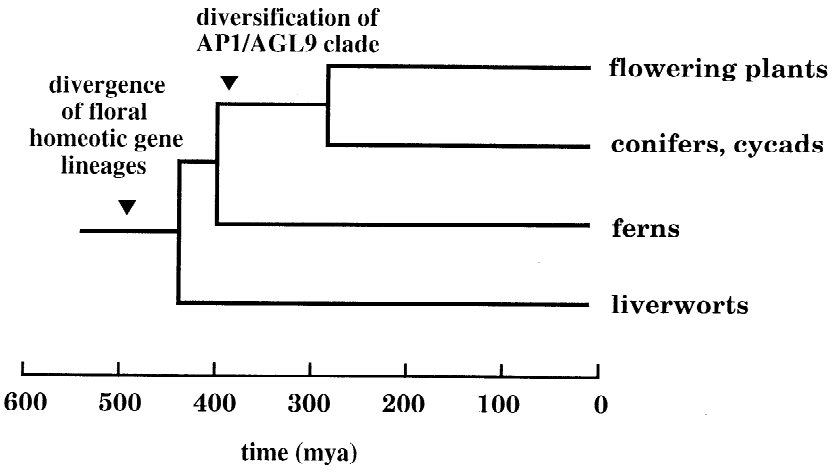

Fig. 2. Evolutionary events in the diversification of the plant MADS-box regulatory gene family. The possible times of divergence for homeotic gene lineages estimated from molecular clock data are mapped on the phylogeny. The phylogeny of the major groups of land plants is based on Savard et al. (1994).

different from one another; this suggests that the rate of evolution is also homogenous through time.

The mean pairwise substitution values for the AG and AP1/AGL9 group at the MIK region is 0.543 ± 0.05. This leads to an estimate of 486 mya ± 45 for the last common ancestor between these floral homeotic gene groups. Using only the MADS-box sequence gives a date of last common ancestry for the floral homeotic gene groups in agreement with the estimate from the MIK region (478 mya ± 24). The date for the last common ancestor of the AP1/AGL9 subgroups is estimated, based on MIK region substitution rates, at 373 mya ± 40 mya.

These estimates suggest that the last common ancestor of the floral homeotic gene lineages apparently existed sometime during the Ordovician (Figure 2). It is about this time the first land plants began to appear, as evidenced by spore microfossils that date back to the Ordovician and early Silurian (Graham 1993; Stewart and Rothwell 1993). These microfossils precede the appearance of the megafossils of *Cooksonia*, an early vascular land plant. If the estimates for the divergence of the floral homeotic gene group is correct, then the diversification that led to the establishment of these developmentally important gene lineages may have occurred sometime during the first appearance of terrestrial plants themselves. Even later diversification events, such as the elaboration of the AP1/AGL9 group into different subgroups, may have transpired at around the time of the split of the ferns from the seed plants 395 mya (Stewart and Rothwell 1993), but certainly no later than the divergence of conifers and flowering plants.

If the initial major expansion of the plant MADS-box gene family leading to the floral homeotic gene groups did occur at the time of the plant colonization of the Ordovician landscape, the establishment of these early lineages could be correlated with the rise of developmental dependency between sporophyte and gametophyte generations that may have been the pivotal innovation in plant evolution during this period (Graham 1993). Subsequent expansions of the family, such as the establishment of the subgroups within the AP1/AGL9 clade, may have arisen as more elaborate and specific developmental processes became necessary to sculpt the reproductive organs of the seed plants, and maybe even vascular plants in general. Recent characterization of several MADS-box genes from the ferns *Ceratopteris* and *Ophioglossum* show that multiple MADS-box genes are also present in these taxa, and some genes are apparently related to *Arabidopsis* loci (Munster et al. 1997). Isolation of these genes in bryophyte and other basal plant species is also ongoing and should shed light on the origins and early evolution of these key plant regulatory genes.

Acknowledgments. I would like to thank R.J. Schmidt, M.F. Yanofsky, J. Thorne, and A. Rambaut for helpful discussions. This work was funded in part by the North Carolina Agricultural Research Service.

References

Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth D (1993) Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development 119:721–743

Coen E (1991) The role of homeotic genes in flower development and evolution. Ann Rev Plant Plysiol Plant Mol Biol 42:241–279

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

Crepet WL, Feldman GD (1991) The earliest remains of grasses in the fossil record. Am J Bot 78:1010–1014

Davies B, Schwarz-Sommer Zs (1994) Control of floral organ identity by homeotic MADS-box transcription factors. Cell Diff 20:235–258

Doyle J (1994) Evolution of a plant homeotic multigene family: towards connecting molecular systematics and molecular developmental genetics. Syst Biol 43:307–328

Gillespie JH (1991) The causes of molecular evolution. Oxford University Press, Oxford

Graham L (1993) Origin of land plants. John Wiley, New York

Kumar S, Tamura K, Nei M (1994) Molecular Evolutionary Genetic Analysis Package 1.1. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, State College

Ma H, Yanofsky MF, Meyerowitz EM (1991) *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev 5:484–495

Munster T, Pahnke J, Di Rosa A, Kim J, Martin W, Saedler H, Theissen G (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. Proc Natl Acad Sci USA 94:2415–2420

Pollock R, Treismann R (1991) Human *SRF*-related proteins—DNA-binding properties and potential regulatory targets. Genes Dev 5: 2327–2341

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development—diversification of the plant MADS-box regulatory gene family. Genetics 140:345–356

Savard L, Li P, Strauss SH, Chase MW, Michaud M, Bousquet J (1994) Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. Proc Natl Acad Sci USA 91:5163–5167

Stewart WN, Rothwell GW (1993) Paleobotany and the evolution of plants, 2nd ed. Cambridge University Press, Cambridge

Swofford D (1993) Phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285

Tandre K, Albert VA, Sundas A, Engstrom P (1995) Conifer homologues to genes that control floral development in angiosperms. Plant Mol Biol 27:69–78

Weigel D (1995) The genetics of flower development—from floral induction to ovule morphogenesis. Ann Rev Genetics 29:19–39

Wolfe K, Gouy M, Yang Y, Sharp P, Li WH (1989) Date of the monocot dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:5201–5205

Yanofsky MF (1995) Floral meristems to floral organs—genes controlling early events in *Arabidopsis* flower development. Ann Rev Plant Phys Plant Mol Biol 46:167–188