Contrasting Evolutionary Forces in the *Arabidopsis thaliana* Floral Developmental Pathway

Kenneth M. Olsen, Andrew Womack, Ashley R. Garrett, Jane I. Suddith and Michael D. Purugganan[1]

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695

Manuscript received October 11, 2001
Accepted for publication January 25, 2002

ABSTRACT

The floral developmental pathway in *Arabidopsis thaliana* is composed of several interacting regulatory genes, including the inflorescence architecture gene *TERMINAL FLOWER1* (*TFL1*), the floral meristem identity genes *LEAFY* (*LFY*), *APETALA1* (*AP1*), and *CAULIFLOWER* (*CAL*), and the floral organ identity genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*). Molecular population genetic analyses of these different genes indicate that the coding regions of *AP3* and *PI*, as well as *AP1* and *CAL*, share similar levels and patterns of nucleotide diversity. In contrast, the coding regions of *TFL1* and *LFY* display a significant reduction in nucleotide variation, suggesting that these sequences have been subjected to a recent adaptive sweep. Moreover, the promoter of *TFL1*, unlike its coding region, displays high levels of diversity organized into two distinct haplogroups that appear to be maintained by selection. These results suggest that patterns of molecular evoution differ among regulatory genes in this developmental pathway, with the earlier acting genes exhibiting evidence of adaptive evolution.

GENES that control morphogenesis invariably function as interacting components of complex developmental networks (Arnone and Davidson 1997; Shubin *et al.* 1997; Davidson 1999). Morphological evolution is believed to arise from the diversification of these interacting developmental genes (Shubin *et al.* 1997; Doebley and Lukens 1998; Purugganan 1998), including both the *trans*-acting regulatory genes that function within developmental networks and *cis*-acting promoter sequences that control gene expression patterns (Doebley and Lukens 1998). Despite the central importance of developmental gene evolution for the diversification of morphology, little is known about how developmental gene networks evolve. Previous studies on individual developmental genes indicate that they are subject to a variety of microevolutionary forces (Wang *et al.* 1999; Ludwig *et al.* 2000; Purugganan 2000), and some genes have been shown to be targets for adaptive evolution (Wang *et al.* 1999). However, the evolutionary dynamics across interacting regulatory genetic components remain unclear.

Analysis of patterns of nucleotide variation at different genes within a developmental pathway may allow us to assess whether the genetic components of a developmental network are subject to equivalent evolutionary forces, or whether they differ in their modes of evolution. An evolutionary analysis across genes also allows an examination of how the selective pressures acting on a gene relate to the gene's position and functional role within the developmental pathway. Moreover, this type of integrated analysis may further allow us to examine how the structure of developmental gene networks constrains the types of evolutionary change observed in nature.

Flower development provides an excellent system for studying the evolution of morphogenesis in plants (Lawton-Rauh *et al.* 2000; Cronk 2001). Geneticists have managed to identify and isolate many of the genes that control flower development in the weed *Arabidopsis thaliana,* which has served as a model plant genetic system (Weigel 1995; Yanofsky 1995; Jack 2001). Moreover, in many cases, geneticists have elucidated the interactions that occur among these genes to control floral development (see Figure 1; Yanofsky 1995; Liljegren *et al.* 1999; Ng and Yanofsky 2000; Jack 2001).

Among the regulatory genes first expressed in floral development is the inflorescence architecture gene *TERMINAL FLOWER1* (*TFL1*), whose product is similar to RAF kinase inhibitor proteins and which appears to be required for the maintenance of inflorescence meristem identity (Bradley *et al.* 1997; Ratcliffe *et al.* 1998). In *A. thaliana,* mutations at *TFL1* result in the transformation of the apical inflorescence meristem into a floral meristem, leading to the formation of a single terminal flower instead of an indeterminate inflorescence (Shannon and Meeks-Wagner 1991; Ratcliffe *et al.* 1998, 1999).

Three floral meristem identity genes, *LEAFY (LFY)*, *APETALA1 (AP1)*, and *CAULIFLOWER (CAL)*, act downstream of *TFL1* and specify the formation of flowers

[1]*Corresponding author:* 3513 Gardner Hall, Department of Genetics, North Carolina State University, Raleigh, NC 27695.
E-mail: michael_purugganan@ncsu.edu

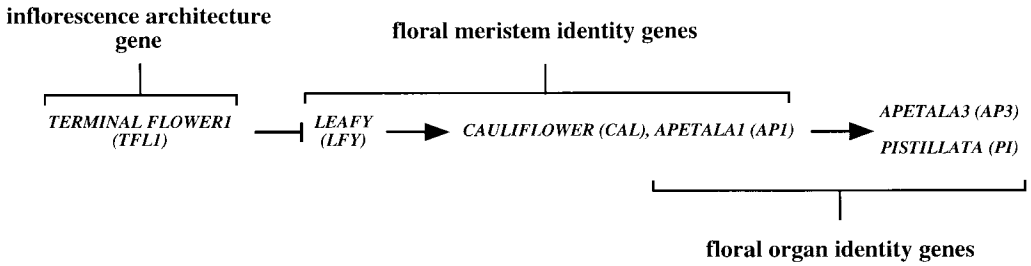

FIGURE 1.—Schematic of the genetic interactions in the *A. thaliana* floral developmental pathway.

(see Figure 1). *LFY* encodes a putative DNA-binding transcriptional activator; mutations in this gene result in the partial transformation of flowers to inflorescence shoots (WEIGEL *et al.* 1992; WEIGEL and MEYEROWITZ 1993). Genetic studies indicate that *TFL1* acts in part by repressing the expression of *LEAFY* in the inflorescence meristems (see Figure 1; LILJEGREN *et al.* 1999). Thus, downregulation of *TFL1* leads to *LFY* expression and is one of the first steps in the genetic cascade that leads to flower formation. *LFY* in turn activates the expression of *AP1*, which encodes a MADS-box DNA-binding transcriptional factor (MANDEL *et al.* 1992; GUSTAFSON-BROWN *et al.* 1994; LILJEGREN *et al.* 1999). The third floral meristem identity gene, *CAL*, is a relatively recent duplicate of *AP1* (KEMPIN *et al.* 1995). *CAL* and *AP1* share redundant functions in the establishment of floral meristem identity, and double mutants of these two genes result in a massive proliferation of inflorescences resulting from an ontogenetic arrest in the ability to form floral meristems (BOWMAN *et al.* 1993; KEMPIN *et al.* 1995).

Floral organ identity genes control the identity of the organs in the four floral whorls—the sepals, petals, stamens, and carpels (YANOFSKY 1995)—and act downstream of the meristem identity loci. The *AP1* locus has floral organ identity functions in addition to its meristem identity functions described above; mutations at *AP1* lead to loss of floral sepals and petals (BOWMAN *et al.* 1993). Other floral organ identity genes include the MADS-box loci *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), both of which are required for petal and stamen development (GOTO and MEYEROWITZ 1994; JACK *et al.* 1992). Mutations at these loci result in the homeotic transformation of petals and stamens to sepals and carpel-like structures, respectively. *AP3* and *PI* are ancient gene duplicates (PURUGGANAN 1997; KRAMER *et al.* 1998). Expression of these two genes involves coregulation as well as regulation by the floral meristem identity genes *LFY* and *AP1* (see Figure 1; SAMACH *et al.* 1997).

The broad functional categories used to classify these floral developmental genes—inflorescence architecture, meristem identity and organ identity—are not absolute, and some genes may be expected to have roles characteristic of more than one functional class, as is evident in *AP1*. Nevertheless, these categories provide a framework for formulating hypotheses about how evolution may act among the different genetic components of the floral developmental pathway. One such hypothesis involves the action of selection on genes that control evolutionarily conserved *vs.* variable morphological traits. Like all members of the Brassicaceae, *A. thaliana* shows strong conservation in the number, positioning, and identities of floral organs. In contrast, adaptive divergence is observed in the numbers and positioning of reproductive shoots, both within *A. thaliana* and among its close relatives (ENDRESS 1992; SHU *et al.* 2000). This difference in evolutionary constraint should be reflected in the evolution of the genes underlying these morphological traits. In particular, we would expect that evidence of adaptive evolution should be more prevalent in the inflorescence architecture and meristem identity loci, which control the evolutionarily more labile reproductive meristem traits, than in the organ identity genes.

In previous studies, we reported on the molecular population genetics of the floral meristem identity gene *CAL* (PURUGGANAN and SUDDITH 1998) and the floral organ identity genes *AP3* and *PI* in *A. thaliana* (PURUGGANAN and SUDDITH 1999). We have now characterized the molecular population genetics of three additional *A. thaliana* floral developmental genes: the floral meristem identity genes *AP1* and *LFY* and the inflorescence architecture gene *TFL1*. Our results indicate that these six genes possess distinctive patterns of sequence diversity and that the two early acting genes in the floral developmental pathway—the floral meristem identity gene *LFY* and the inflorescence architecture gene *TFL1*—show evidence of adaptive evolution.

MATERIALS AND METHODS

Isolation and sequencing of alleles: The *A. thaliana* ecotypes were obtained from single-seed propagated material provided by the *Arabidopsis* Biological Resource Center (ABRC; see Table 1). The Kent, Bretagne, Lisse, and Corsacalla seed stocks were from the population collection of P. H. Williams maintained at ABRC. *A. lyrata* seed was provided by C. H. Langley and O. Savolainen. The majority of the accessions used in this study were the same as those in previous work (PURUGGANAN and SUDDITH 1998, 1999).

Miniprep DNA was isolated from young leaves as previously described (AUSUBEL 1992). PCR was performed with the error-correcting *Pwo* polymerase (Roche, Indianapolis) to minimize nucleotide misincorporation. The error rate for this polymer-

TABLE 1

Ecotypes/field strains

Ecotype/field strain	Locality	ABRC seed accession no.
Basel-1	Basel, Switzerland	CS996
Blanes/Gerona-1	Blanes/Gerona, Spain	CS970
Bretagne-2[a]	Bretagne, France	CS6096
Bretagne-4[a]	Bretagne, France	CS6094
Burghaun/Rhon-0	Burghaun/Rhon, Germany	CS1006
Burghaun/Rhon-2	Burghaun/Rhon, Germany	CS1008
Cape Verde-0	Cape Verde Islands	CS902
Chisdra-1	Chisdra, Russia	CS1074
Coimbra-1	Coimbra, Portugal	CS1084
Columbia	Landsberg, Germany	CS20
Corsacalla-1[a]	Corsacalla, Italy	CS6042
Graz-3	Graz, Austria	CS1202
Jelinka-1	Vranov u Brno, Czechoslovakia	CS1248
Kashmir-1	Kashmir, India	CS903
Kent-2[a]	Kent, Great Britain	CS6054
Kent-4[a]	Kent, Great Britain	CS6048
Kent-5[a]	Kent, Great Britain	CS6058
Landsberg erecta	Landsberg, Germany	—
Limburg-3	Limburg, Germany	CS1316
Limburg-8	Limburg, Germany	CS1332
Lisse-1[a]	Lisse, Netherlands	CS6090
Lisse-2[a]	Lisse, Netherlands	CS6092
Ws	Wassileskija, Ukraine	CS915

[a] Field strains from the Williams collection (ABRC). The name is not an official ecotype designation.

ase formulation, based on multiple amplification and resequencing of known genes, is less than one in 7–10 kb (our unpublished observations). We estimate that the nonsampling variance of nucleotide diversity due to PCR misincorporation, $V_{PCR}(\pi)$, is negligible [$V_{PCR}(\pi)/V(\pi) \sim 0.14$] and does not significantly affect the frequency distribution of polymorphisms. In several cases, multiple sequences from independently amplified products were obtained to recheck potential PCR-induced errors. The *APETALA1* gene was amplified in three segments in *A. thaliana* and *A. lyrata*: the first segment primers AP1-PR1F (5′-AGGCTTATGCAATATATGCCTTAAGC-3′) and AP1-X1R (5′-TTGTCTATTGATCTTGTTCTCTATCC-3′); the second segment primers AP1-1F (5′-ATGGGAAGGGGTAGGG TTCA-3′) and AP1-2R (5′-AAGGTTGCAGTTGTAAACGGG-3′); and the third segment primers AP1-X2F (5′-ATGGAGAAGA TACTTGAACG-3′) and AP1-X2R (5′-CTCAGGTGCAATAAG CTGTCT-3′). The *LEAFY* gene was amplified in two segments: the first segment primers LFY1F (5′-CAGACTCAGAGTGCTG ATATTTCT-3′) and LFY1R (5′-GTTCCTCAGATAACCCTGT CCA-3′) and the second segment primers LFY2F (5′-TGGA CAGGGTTATCTGAGGAAC-3′) and LFY2R (5′-ATCTTAGTA CTTTTGAGTTTGACC-3′). Finally, the *TERMINAL FLOWER1* gene was amplified with the primers TFL16F (5′-CCTACTCT GAGCAATAATTGTATCC-3′) and TFL1R (5′-GCAGTTTAT GACAATCATGAAACTA-3′). Amplification conditions followed the *Pwo* polymerase manufacturer's protocols (Roche), with annealing temperatures adjusted for each primer pair.

Amplified DNA products were cloned into pCR2.1 using the TOPO TA cloning kit (Invitrogen, San Diego). DNA sequencing for both genes was conducted with an ABI377 automated sequencer using a series of nested internal sense and antisense primers. All sequence polymorphisms were visually rechecked from chromatograms, with special attention to low-frequency polymorphisms (Hamblin and Aquadro 1997). The DNA sequences are available from GenBank (accession nos. AF466771–AF466817).

Data analysis: Sequences used in this study were visually aligned. Phylogenetic analyses were conducted using the heuristic search algorithm (maximum parsimony criterion) in PAUP 3.1 (Swofford 1992), with the *A. lyrata* orthologs specified as the outgroup. Node support was assessed with 500 bootstrap replicates of the data. Nucleotide diversity was analyzed using the DnaSP program (Rozas and Rozas 1999). Levels of nucleotide diversity were estimated as mean pairwise differences (π). The Tajima (1989) and Fu and Li (1993) tests for selection were conducted without specifying an outgroup. For the *TFL1* promoter, Tajima's *D* was estimated using nucleotide and indel polymorphism data, with the latter coded as single characters. Indels in an interrupted microsatellite near the *TFL1* translation start were excluded from the analysis. Significance of Tajima's *D* was determined in simulations with 10,000 runs using the number of segregating sites ($s = 46$) and the recombination parameter ($R = 2.3$) estimated from the data. Contingency tests for independence of mutational categories, referred to as the McDonald-Kreitman test (McDonald and Kreitman 1991), were conducted using Fisher's exact test to evaluate significance. The Hudson-Kreitman-Aguade (HKA) two-locus and multilocus tests (Hudson *et al.* 1987) were conducted using DnaSP (Rozas and Rozas 1999) and the HKA program available from J. Hey.

RESULTS AND DISCUSSION

Nucleotide variation in the coding regions of the *A. thaliana* floral developmental genes: A total of 15 *APETALA1*, 15 *LEAFY*, and 14 *TERMINAL FLOWER1* alleles were isolated from a collection of *A. thaliana* ecotypes, sampled primarily from Europe. Allele sequences from the entire coding region and a portion (~1 kb) of the promoter and 5′-untranslated region (5′-UTR) were obtained for each gene. Approximately 4.7 kb was sequenced for each *AP1* allele, spanning exons 1–8 and including 1.2 kb of the 5′-untranslated region and the promoter (Figure 2). Approximately 3.9 kb of sequence was obtained from *LFY* alleles. The *LFY* sequences include the entire coding region (from exons 1–3 and intervening introns) and 1.1 kb of the 5′-UTR and promoter (Figure 3). About 1.8 kb of the *TFL1* allele sequence was obtained, including exons 1–4 and 0.7 kb of promoter and 5′-UTR (Figure 4). The *AP1*, *LFY*, and *TFL1* genes encode proteins of 255, 424, and 177 amino acids in length, respectively.

These three *A. thaliana* floral developmental genes show different levels of coding region nucleotide polymorphism. In *AP1*, a total of 91 polymorphic nucleotide sites are in the 3.5-kb coding region of this gene (Figure 2). Of these, 10 exon polymorphisms are replacements and 8 are synonymous changes; 73 polymorphisms are found in introns. There are also 8 insertion/deletion (indel) polymorphisms, all in introns, which range in size from 1 to 5 bp. The *LEAFY* alleles have a total of

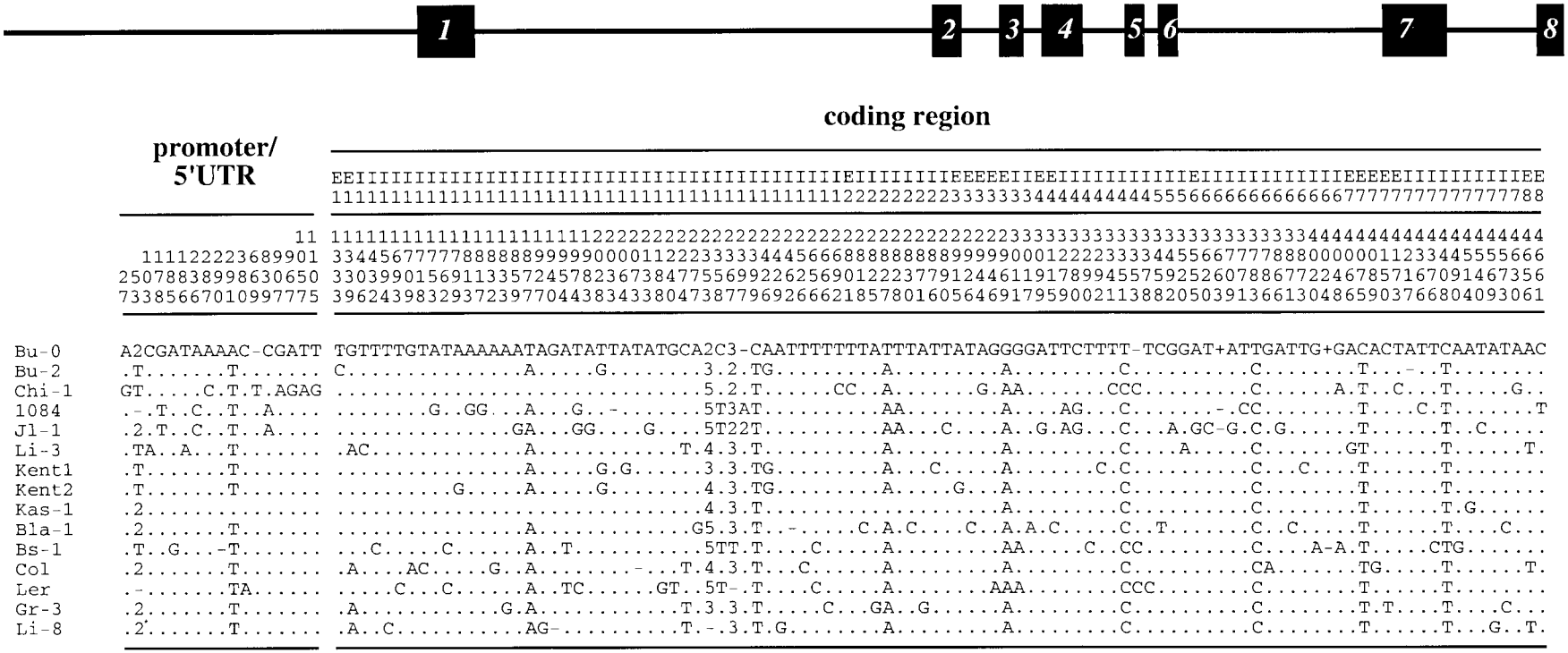

FIGURE 2.—Sequence of *APETALA1* alleles from different *A. thaliana* ecotypes. All allele sequences are compared to the reference allele from the Burghaun/Rhon-0 Bu-0) ecotype. The position of the polymorphic sites and their locations in introns and exons are indicated at the top. Numbers within indels indicate a repeated nucleotide; the repeated nucleotide at position 2353 is T. Insertions (+) at positions 3663 and 4024 are CTAT and ATA, respectively.

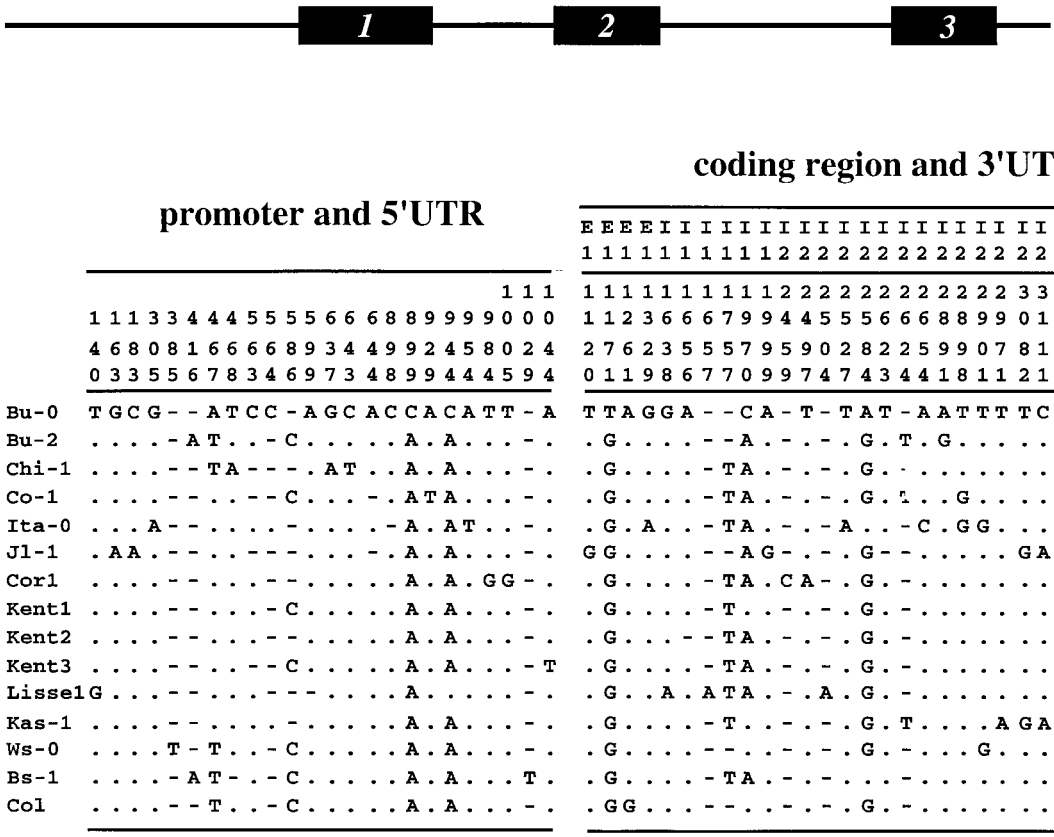

FIGURE 3.—Sequence of *LEAFY* alleles from different *A. thaliana* ecotypes. All allele sequences are compared to the reference allele from the Burghaun/Rhon-0 (Bu-0) ecotype. The position of the polymorphic sites and their locations in introns and exons are indicated at the top. Insertions (+) at positions 599 and 3871 are AAAA and ATGGCTAATTTGTT, respectively. Polymorphisms not shown: position 576, multiple A repeats; position 1082, multiple GA repeats.

20 nucleotide polymorphisms within the 2.9-kb coding region (Figure 3). There are 4 replacement and 2 synonymous polymorphisms in exon sequences, while 14 segregating sites are found in introns. Eight indel polymorphisms occur in *LEAFY*; like *AP1*, all of these are found in intron regions. Finally, the sampled alleles from the *TFL1* locus have 6 polymorphic nucleotide sites, including 1 site with two mutations, in the 1.1-kb region that spans the exons and introns. Of these polymorphisms, 3 are replacements, 1 is a synonymous polymorphism, and 3 are found in intron sequences. Three indel polymorphisms are observed in the *TFL1* coding region, all in introns.

Estimates of silent nucleotide variation within the coding region of the *APETALA1* locus are comparable to the three other floral homeotic genes previously studied (PURUGGANAN and SUDDITH 1998, 1999; see Table 2 and Figure 5). The estimate of *AP1* species-wide nucleotide diversity for coding region silent sites, π, is 0.0047. This level of variation is slightly lower than that of *CAL* (π = 0.0069), which is a recent duplicate of *AP1*; both possess redundant meristem identity functions in flower development. The level of silent site coding region variation for *LFY*, however, is 0.0019 (see Table 2 and Figure 5), which is less than half that of *AP1*. An even lower level of silent variation is observed in the *TFL1* coding region, with an estimate of π for silent sites at 0.0007.

Excess of low-frequency and replacement polymorphisms at the *AP1* transcriptional unit: The frequency distribution of polymorphisms provides information on the relative roles of neutral drift *vs.* selection at a locus. The skewness of frequency distributions for nucleotide polymorphisms can be evaluated with both the TAJIMA (1989) and FU and LI (1993) tests for selection. Since *A. thaliana* may have experienced a recent population expansion, these two tests should be interpreted with caution when inferring selection. However, they may still provide information on deviations in molecular diversity patterns from predictions of the neutral-equilibrium model. When the nucleotide diversity in the *AP1* coding region is examined, the distribution of polymorphic sites is significantly skewed toward rare alleles. Seventy-three of the 92 nucleotide polymorphisms occur in only a single allele (singleton mutations). The Tajima test statistic, *D*, is -2.101 for *AP1* ($P < 0.05$); the negative value of this statistic indicates that sampled alleles show an excess of rare polymorphisms over that expected in a population at mutation-drift equilibrium (Table 2). The Fu and Li test statistic, D^*, is also significantly negative for this gene ($D^* = -3.031$, $P < 0.02$), again indicat-

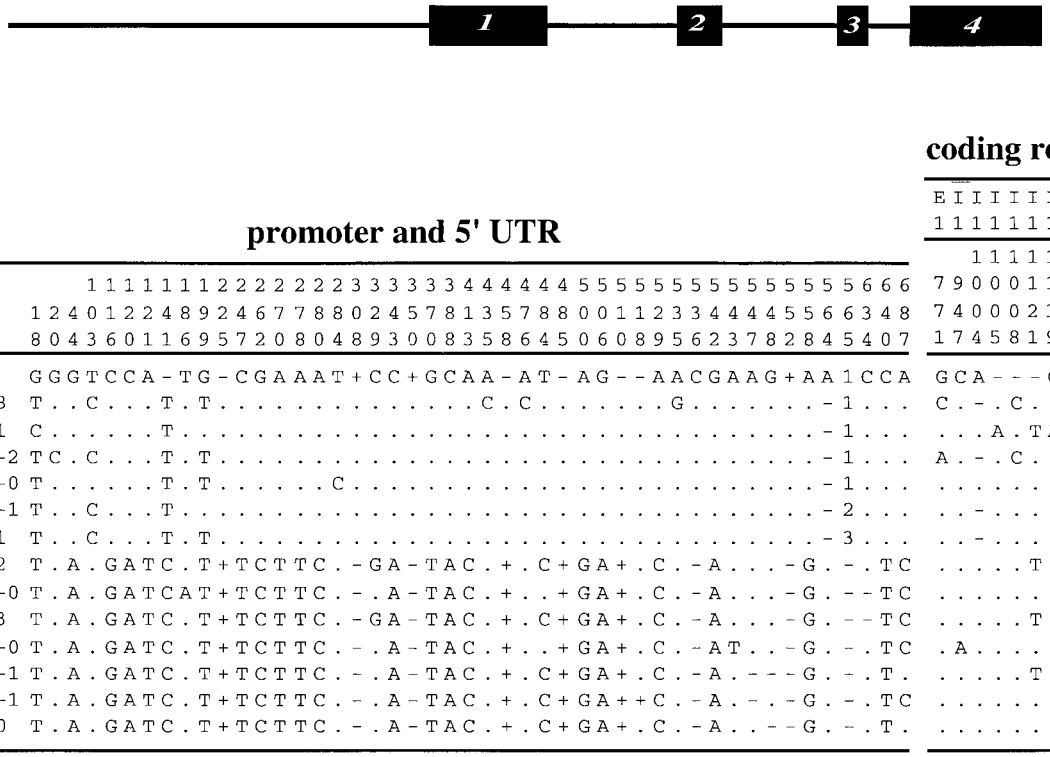

FIGURE 4.—Sequence of *TERMINAL FLOWER1* alleles from different *A. thaliana* ecotypes. All allele sequences are compared to the reference allele from the Columbia (Col-0) ecotype. The position of the polymorphic sites and their locations in introns and exons are indicated at the top. Insertions (+) are as follows: position 225, TATATAGG; position 308, TTGAAACCATG; position 350, TGTTGCTGAAGAAGTCAA; position 458, AGCTT; position 485, AGTATTAGAAA; position 510, TACG; position 518, CA; position 552, AA; position 565, multiple GA repeats.

ing an excess of singletons. This excess of low-frequency polymorphisms for *AP1* is similar to that observed for several other Arabidopsis nuclear genes, including the floral organ identity genes *AP3* and *PI*, as well as the *AP1* paralogue *CAL* (PURUGGANAN and SUDDITH 1998, 1999; see Table 2). This pattern of variation may reflect the inbreeding associated with this selfing plant and/or population disequilibrium caused by the rapid post-Pleistocene range expansion in this species (PURUGGANAN and SUDDITH 1999; KUITTINEN and AGUADE 2000; AGUADE 2001).

Selective sweeps at the *LFY* and *TFL1* transcriptional units: Like *AP1*, the *TFL1* locus also possesses an excess of low-frequency polymorphisms in its coding region. The Tajima test statistic, *D*, is −2.032 for this gene ($P < 0.05$; see Table 2); the Fu and Li test statistic *D** is also significantly negative ($D^* = -2.600$, $P < 0.02$). Negative values for Tajima and Fu and Li test statistics are also observed for *LFY*; this excess of rare alleles, however, is not significantly different from the distribution expected under a neutral-equilibrium model ($D = -1.572$, $P > 0.05$; $D^* = -1.516$, $P > 0.05$; Table 3).

The excess of low-frequency polymorphisms observed at *TFL1* (and to some extent *LFY)* is associated with a reduction in overall levels of nucleotide variation relative to other genes in the floral developmental pathway.

TABLE 2

Features of coding region sequence variation at *A. thaliana* floral developmental genes

Gene	Map position[a]	Length (kb)	*n*	π (silent)	π (nonsynonymous)	Tajima *D*	Fu and Li *D**
AP3	3 (80)	1.7	19	0.0077	0.0040	−2.151 ($P < 0.05$)*	−3.373 ($P < 0.02$)*
PI	5 (28)	2.0	16	0.0061	0.0030	−2.018 ($P < 0.05$)*	−2.876 ($P < 0.02$)*
CAL	1 (46)	2.0	17	0.0069	0.0055	−1.661 ($P < 0.08$)	−2.372 ($P < 0.06$)
AP1	1 (99)	3.5	15	0.0047	0.0022	−2.101 ($P < 0.05$)*	−3.032 ($P < 0.02$)*
LFY	5 (117)	2.9	15	0.0019	0.0002	−1.572 ($P < 0.10$)	−1.516 ($P < 0.10$)
TFL1	5 (2)	1.1	14	0.0007	0.0014	−2.032 ($P < 0.05$)*	−2.600 ($P < 0.02$)*

*Significant at the 0.05 level.

[a] Position given as chromosome number, map position in parentheses.

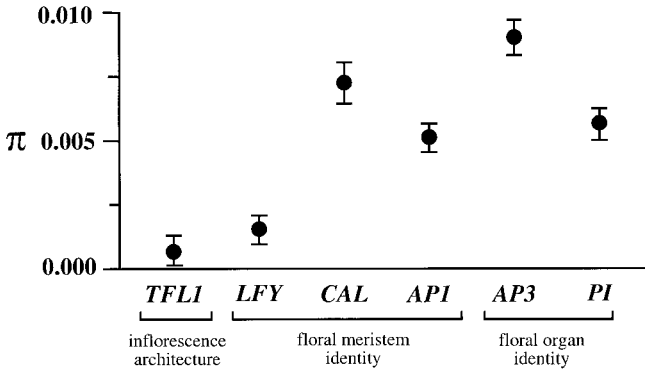

FIGURE 5.—Comparison of silent site nucleotide diversity levels in the coding regions of genes in the *A. thaliana* floral developmental pathway. The genes are arranged from earlier acting (left) to later acting (right) genes.

Whereas the mean silent-site nucleotide diversity for the coding regions of these two genes is 0.0013, the mean value for the four other genes examined here is 0.0063. Compared to 15 previously examined Arabidopsis nuclear loci, *TFL1* and *LFY* have a 5- to 10-fold reduction in polymorphism (KAWABE *et al.* 1997; MIYASHITA *et al.* 1998; KAWABE and MIYASHITA 1999; KAWABE *et al.* 2000; KUITTINEN and AGUADE 2000; SAVOLAINEN *et al.* 2000; AGUADE 2001; MIYASHITA 2001).

A multilocus HKA test comparing intraspecific polymorphism with interspecific divergence in the coding regions of all six floral developmental genes is significant ($\chi^2 = 11.935$, $P < 0.035$), with *LFY* and *TFL1* contributing most to the deviation from neutral expectations (Figure 6). This pattern suggests that the evolutionary dynamics of the coding regions of these two regulatory genes differ significantly from the other genes of the floral developmental pathway. Specifically, the patterns of evolution for the coding regions of *TFL1* and *LFY* suggest a recent selective sweep, either at these loci or at closely linked genes. Of 15 genes in *A. thaliana* analyzed thus far, only one other locus—*CHALCONE ISOMERASE*—shows evidence of a recent selective sweep (KUITTINEN and AGUADE 2000).

Patterns of regulatory protein evolution: An excess of intraspecific replacement polymorphisms has previously been documented for the floral developmental genes *AP3*, *PI*, and *CAL* (see Table 3), as well as several other nuclear loci (KAWABE *et al.* 1997; PURUGGANAN and SUDDITH 1998, 1999). While not observed at either *AP1* or *LFY* (Table 3), this pattern is found at *TFL1*. This inflorescence architecture gene has three replacement polymorphisms but only a single synonymous polymorphism within *A. thaliana*, compared to three replacements and 13 synonymous changes fixed between *A. thaliana* and *A. lyrata* (Table 3). Nonetheless, this observed elevation in within-species replacement polymorphism is not statistically significant (McDonald-Kreitman test, $P = 0.0609$).

Patterns of replacement and synonymous variation at the six floral genes examined here suggest that many of the nonsynonymous polymorphisms may be slightly deleterious. A hierarchical Bayesian model of protein evolution indicates negative selection intensities against amino acid replacements for all of these floral developmental genes except *LFY*, a pattern consistent with that found in the majority of *A. thaliana* nuclear loci (BUSTAMANTE *et al.* 2002). This prevalence of deleterious mutations may reflect the effect of inbreeding in this predominantly selfing species (BUSTAMANTE *et al.* 2002).

Promoter variation in floral meristem identity and inflorescence architecture genes: Comparison of promoter/5′-UTR regions of *AP1*, *LFY*, and *TFL1* using a multilocus HKA test does not indicate statistically significant differences in patterns of evolution among these three genes ($\chi^2 = 1.635$, $P > 0.4$). However, examination of patterns and levels of nucleotide variation within each of these genes suggests that they do differ in their evolutionary dynamics. For two of the genes, *AP1* and *LFY*, nucleotide variation in the promoter/5′-UTR is very similar to that observed in the coding region of the gene. Individual HKA tests for both of these loci indicate no significant difference between the promoter/5′-UTR region and coding region in levels of within-species polymorphism *vs.* between-species diver-

TABLE 3

McDonald-Kreitman tables for *A. thaliana* floral developmental genes

Gene	Polymorphisms		Fixed differences[a]		McDonald-Kreitman test
	Replacement	Synonymous	Replacement	Synonymous	
AP3	19	9	6	14	$P = 0.018$*
PI	12	4	9	18	$P = 0.012$*
CAL	15	5	14	14	$P = 0.134$
AP1	10	8	8	11	$P = 0.517$
LFY	2	4	18	37	$P = 1.000$
TFL1	3	1	3	13	$P = 0.061$

* Significant at the 0.05 level.
[a] Compared with *A. lyrata*.

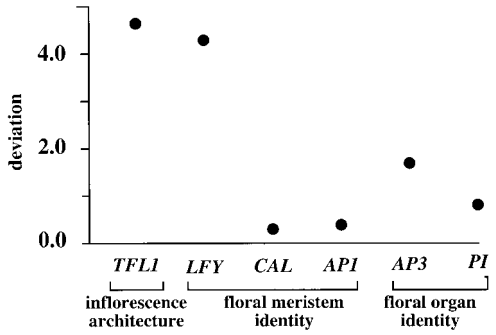

FIGURE 6.—Contribution of each gene in the *A. thaliana* floral developmental pathway to the χ^2 statistic for the multilocus HKA test of selection on coding region sequences. The total χ^2 is significantly greater than expected ($P < 0.035$) in the multilocus HKA.

gence (Table 4). These patterns suggest that for *API* and *LFY*, evolutionary forces have acted comparably in the two portions of the gene.

In contrast to *API* and *LFY*, *TFL1* shows a dramatic difference in nucleotide variation between its promoter/5′-UTR and coding region. Whereas nucleotide diversity in the coding region is extremely low (π = 0.0007; see above), that of the promoter/5′-UTR is very high (π = 0.0187; Table 4). The HKA test confirms that these two portions of the gene differ in their patterns of evolution ($\chi^2 = 6.243$, $P < 0.012$). Moreover, whereas Tajima's *D* and Fu and Li's *D** estimates are negative for the promoter/5′-UTR and coding regions of *API* and *LFY* as well as for the coding region of *TFL1*, these two polymorphism measures are positive for the *TFL1* promoter/5′-UTR (see Tables 2 and 4). Thus, the promoter/5′-UTR and coding region of *TFL1* have differed markedly in their patterns of evolution.

Nucleotide variation in the *TFL1* promoter characterizes two distinct classes of alleles (haplogroups), which are distinguished by 20 nucleotide polymorphisms and 10 indels (Figure 4). This pattern of allelic dimorphism is not unprecedented in *A. thaliana*, having been reported in several loci including *FAH* (AGUADE 2001) and *ChiB* (KAWABE and MIYASHITA 1999). The origins and evolutionary forces maintaining such allelic dimorphism are still unclear (KAWABE and MIYASHITA 1999; AGUADE 2001). What distinguishes the *TFL1* promoter sequence dimorphism from other genes, however, is (i) the close proximity of the dimorphic region to sequences that appear to have undergone a recent selective sweep and (ii) the significantly positive value of Tajima's *D*. Both of these observations are consistent with selection to maintain *TFL1* promoter dimorphism in the face of selective sweeps in the adjacent coding region of this gene.

Variation of *TFL1* is expected to affect the number of floral meristems established by the shoot apical meristem. Indeed, in a sample of 78 ecotypes, there is a difference in the number of main axis flowers produced in plants that have *TFL1* promoter haplogroups A and B—14.29 ± 3.93 and 16.08 ± 3.86, respectively (our unpublished observations). This difference is only marginally nonsignificant after correcting for the effects of population structure ($P < 0.064$).

The molecular population genetics of the floral developmental pathway: We have examined the molecular population genetics of six regulatory genes that are found at different positions in the floral developmental pathway. These loci have various roles in floral development, including the maintenance of inflorescence meristem identity (*TFL1*), the specification of floral meristem identity (*LFY*, *API*, and *CAL*), and the development of sepals, petals, and stamens (*API*, *AP3*, and *PI*). On the basis of the broad functional classes, one gene is an inflorescence architecure gene (*TFL1*), two are categorized as floral meristem identity genes (*LFY* and *CAL*), while two are floral organ identity loci (*AP3* and *PI*). *API* is both a meristem specification and a floral patterning gene.

Two of these loci show evidence of having experienced a recent adaptive sweep. These are the inflorescence architecture gene *TFL1* and the floral meristem gene *LFY*, both of which are early acting genes that regulate the identities of inflorescence and floral meristems. Coding regions of both of these loci show reduced levels of polymorphism compared to other floral developmental loci, consistent with recent positive selection. It is unclear whether these loci were the actual targets of selection; it is possible that the selective target may have been a closely linked gene (BARTON 2000). However, the chromosomal locations of these genes do not suggest that they would be tightly linked to other

TABLE 4

Promoter/5′-UTR sequence variation at Arabidopsis floral developmental genes

Gene	Length (kb)	π	Tajima *D*	Fu and Li *D**	HKA test[a]
API	1.2	0.0013	−1.824 ($P < 0.05$)*	−1.744 ($P > 0.10$)	$\chi^2 = 0.112$ ($P > 0.74$)
LFY	1.1	0.0024	−1.825 ($P < 0.05$)*	−2.391 ($P < 0.05$)*	$\chi^2 = 1.337$ ($P > 0.25$)
TFL1	0.7	0.0187	+1.531 ($P < 0.05$)[b]	+0.354 ($P > 0.10$)	$\chi^2 = 6.243$ ($P < 0.01$)*

* Significant at the 0.05 level.
[a] Comparison of promoter/5′-UTR with coding region.
[b] Significance determined by coalescence simulations.

loci; comparison of physical and genetic map distances in the regions flanking both of these loci indicates that they are not regions of low recombination (K. M. Olsen and M. D. Purugganan, unpublished observation). Moreover, the data on *TFL1* indicate that the selective sweep was confined to the coding region and did not include the promoter sequences, which suggests that recombination was sufficient to restrict the physical impact of the selective sweep.

In addition to selective sweeps, other selective forces are also evident in these floral genes. At *TFL1*, the two promoter haplogroups occur at roughly equal frequency in wild ecotypes and haplogroup differentiation does not extend to the 5′ proximal gene *rpS28* (our unpublished observations). Moreover, the *TFL1* promoter haplogroup types are weakly associated with the developmental decision to form flowers. The allele structure of *TFL1* appears to have been shaped by two contrasting selective forces: an adaptive sweep in the coding region and selection to maintain variation in the promoter/5′-UTR. At the meristem identity gene *CAL*, genetic analysis indicates that naturally occurring alleles differ in their ability to regulate floral meristem identity specification; however, unambiguous evidence for positive or diversifying selection has not been observed for this locus (Purugganan and Suddith 1998). In contrast to *LFY*, *TFL1*, and *CAL*, the three genes with exclusively organ identity functions (*AP1*, *AP3*, and *PI*) show no evidence of either positive or balancing selection (see also Purugganan and Suddith 1999).

The selective forces inferred for these genes are consistent with our predictions based on patterns of morphological evolution. Organ identity genes (*AP1*, *AP3*, and *PI*), which control evolutionarily conserved floral organ traits, show no evidence of adaptive evolution. The significant excess of rare polymorphisms in these three genes, as reflected in the negative Tajima's *D* estimate for these loci, appear to result from the persistence of deleterious mutations in Arabidopsis nuclear genes (Bustamante *et al.* 2002). In contrast, the inflorescence architecture and meristem identity genes (*TFL1* and *LFY*, respectively), which control traits associated with inflorescence morphology in *A. thaliana* and its allies, bear molecular signatures suggesting adaptive selection. Both *TFL1* and *LFY* are positioned at critical points in the developmental pathway, integrating signals from multiple genes and regulating developmental decisions on the initiation of the floral developmental program (Bradley *et al.* 1997; Liljegren *et al.* 1999). This suggests that the adaptive evolution of inflorescence morphology may arise through selection on the timing and spatial patterning of flower and inflorescence branch formation. It will be interesting to determine how selection has operated even earlier than *TFL1* in the floral developmental pathway, including those genes that specify natural variation in flowering time (Johanson *et al.* 2000).

These results suggest that detailed understanding of the organization of developmental gene pathways and the precise functional roles and interactions of their component loci may provide information on the patterns of diversification of genes that control morphogenesis. Evolutionary analyses of enzymatic pathways, such as those of the glycolytic pathways in Drosophila (Eanes 1999; Verrelli and Eanes 2001) and the anthocyanin biosynthesis pathway in Ipomoea (Rausher *et al.* 1999), have already provided insights into the position- and function-dependent action of selective forces. In the study of Drosophila glycolytic enzymes, evolutionary analysis was informed by metabolic control theory, which provides a theoretical basis for advancing hypotheses on the evolution of pathway components (Eanes 1999). A similar theoretical framework is possible for developmental gene networks (Mendoza *et al.* 1999; Hasty *et al.* 2001), although quantitative models permitting precise evolutionary hypotheses have yet to be advanced. Nevertheless, the ability to correlate patterns of adaptive evolution with the functional roles of regulatory genes in a developmental gene network allows greater understanding of the mechanisms of morphological evolution.

The authors thank K. A. Shepard for insightful discussions and members of the Purugganan laboratory for a critical reading of this manuscript. This work was funded with a grant from the National Science Foundation to M.D.P., T. F. C. Mackay, and J. Schmitt and an Alfred P. Sloan Foundation Young Investigator Award in Molecular Evolution to M.D.P.

LITERATURE CITED

Aguade, M., 2001 Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH* and *F3H* genes in *Arabidopsis thaliana*. Mol. Biol. Evol. **18:** 1–9.

Arnone, M. I., and E. H. Davidson, 1997 The hardwiring of development: organization and function of genomic regulatory systems. Development **124:** 1851–1864.

Ausubel, F., 1992 *Short Protocols in Molecular Biology*. John Wiley & Sons, New York.

Barton, N. H., 2000 Genetic hitchhiking. Philos. Trans. R. Soc. London B Biol. Sci. **355:** 1553–1562.

Bowman, J. L., J. Alvarez, D. Weigel, E. M. Meyerowitz and D. R. Smyth, 1993 Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development **119:** 721–743.

Bradley, D., O. Ratcliffe, C. Vincent, R. Carpenter and E. S. Coen, 1997 Inflorescence commitment and architecture in Arabidopsis. Science **275:** 80–83.

Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan *et al.*, 2002 The cost of inbreeding in *Arabidopsis*. Nature (in press).

Cronk, Q. C. B., 2001 Plant evolution and development in a post-genomic context. Nat. Rev. Genet. **2:** 607–619.

Davidson, E. H., 1999 A view from the genome: spatial control of transcription in sea urchin development. Curr. Opin. Genet. Dev. **9:** 530–541.

Doebley, J., and L. Lukens, 1998 Transcriptional regulators and the evolution of plant form. Plant Cell **10:** 1075–1082.

Eanes, W. F., 1999 Analysis of selection on enzyme polymorphisms. Annu. Rev. Ecol. Syst. **30:** 301–326.

Endress, P. K., 1992 Evolution and floral diversity—the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum*. Int. J. Plant Sci. **153:** S106–S122.

Fu, Y.-X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Goto, K., and E. M. Meyerowitz, 1994 Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes Dev. **13:** 1548–1560.

Gustafson-Brown, C., B. Savidge and M. F. Yanofsky, 1994 Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. Cell **76:** 131–143.

Hamblin, M. T., and C. F. Aquadro, 1997 Contrasting patterns of nucleotide sequence variation at the *Gld* locus in different populations of *Drosophila melanogaster*. Genetics **145:** 1053–1062.

Hasty, J., D. McMillen, F. Isaacs and J. J. Collins, 2001 Computational studies of gene regulatory networks: *in numero* molecular biology. Nat. Rev. Genet. **2:** 268–279.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Jack, T., 2001 Relearning our ABCs: new twists on an old model. Trends Plant. Sci. **6:** 310–316.

Jack, T., L. L. Brockmann and E. M. Meyerowitz, 1992 The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. Cell **68:** 683–697.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino *et al.*, 2000 Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science **290:** 344–347.

Kawabe, A., and N. T. Miyashita, 1999 DNA variation in the basic *ChiB* region of the wild plant *Arabidopsis thaliana*. Genetics **153:** 1445–1453.

Kawabe, A., H. Innan, R. Terauchi and N. T. Miyashita, 1997 Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol. Biol. Evol. **14:** 1303–1315.

Kawabe, A., K. Yamane and N. T. Miyashita, 2000 DNA polymorphism at the cytosolic *PGI* locus of the wild plant *Arabidopsis thaliana*. Genetics **156:** 1339–1347.

Kempin, S. A., B. Savidge and M. F. Yanofsky, 1995 Molecular basis of the cauliflower phenotype in *Arabidopsis*. Science **267:** 522–525.

Kramer, E., R. Dorit and V. Irish, 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.

Kuittinen, H., and M. Aguade, 2000 Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. Genetics **155:** 863–872.

Lawton-Rauh, A., E. Alvarez-Buylla and M. D. Purugganan, 2000 Molecular evolution of flower development. Trends Ecol. Evol. **15:** 144–149.

Liljegren, S., C. Gustafson-Brown, A. Pinyopich, G. S. Ditta and M. F. Yanofsky, 1999 Interactions among the *APETALA1*, *LEAFY* and *TERMINAL FLOWER1* specify meristem fate. Plant Cell **11:** 1007–1018.

Ludwig, M. Z., C. Bergman, N. H. Patel and M. Kreitman, 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403:** 564–567.

Mandel, M. A., C. Gustafson-Brown, B. Savidge and M. F. Yanofsky, 1992 Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360:** 273–277.

McDonald, J., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Mendoza, L., D. Thieffry and E. R. Alvarez-Buylla, 1999 Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. Bioinformatics **15:** 593–606.

Miyashita, N., 2001 DNA variation in the 5′ upstream region of the *Adh* locus of the wild plants *Arabidopsis thaliana* and *Arabis gemmifera*. Mol. Biol. Evol. **18:** 164–171.

Miyashita, N. T., A. Kawabe, H. Innan and R. Terauchi, 1998 Intra- and interspecific DNA variation and codon bias of the alcohol dehydrogenase (*Adh*) locus in *Arabis* and *Arabidopsis* species. Mol. Biol. Evol. **15:** 1420–1429.

Ng, M., and M. F. Yanofsky, 2000 Three ways to learn the ABCs. Curr. Opin. Plant Biol. **3:** 47–52.

Purugganan, M. D., 1997 The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. J. Mol. Evol. **45:** 392–396.

Purugganan, M. D., 1998 The molecular evolution of development. Bioessays **20:** 700–711.

Purugganan, M. D., 2000 The molecular population genetics of regulatory genes. Mol. Ecol. **9:** 1451–1461.

Purugganan, M. D., and J. Suddith, 1998 Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: nonneutral evolution and wild variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95:** 8130–8134.

Purugganan, M. D., and J. I. Suddith, 1999 Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics **151:** 839–848.

Ratcliffe, O. J., I. Amaya, C. A. Vincent, S. Rothstein, R. Carpenter *et al.*, 1998 A common mechanism controls the life cycle and architecture of plants. Development **125:** 1609–1615.

Ratcliffe, O. J., D. J. Bradley and E. S. Coen, 1999 Separation of shoot and floral identity in *Arabidopsis*. Development **126:** 1109–1120.

Rausher, M. D., R. E. Miller and P. Tiffin, 1999 Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol. Biol. Evol. **16:** 266–274.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Samach, A., S. E. Kohalmi, P. Motte, R. Datla and G. W. Haughn, 1997 Divergence of function and regulation of class B floral organ identity genes. Plant Cell **9:** 559–570.

Savolainen, O., C. H. Langley, B. P. Lazzaro and H. Freville, 2000 Contrasting patterns of nucleotide polymorphisms at the *Adh* locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol. Biol. Evol. **17:** 645–655.

Shannon, S., and D. R. Meekswagner, 1991 A mutation in the *Arabidopsis TFL1* gene affects inflorescence meristem development. Plant Cell **3:** 877–892.

Shu, G. P., W. Amaral, L. C. Hileman and D. A. Baum, 2000 *LEAFY* and the evolution of rosette flowering in violet cress. Am. J. Bot. **87:** 634–641.

Shubin, N., C. Tabin and S. Carroll, 1997 Fossils, genes and the evolution of animal limbs. Nature **388:** 639–648.

Swofford, D., 1992 *Phylogenetic Analysis Using Parsimony 3.1*. Illinois Natural History Survey, Champaign, IL.

Tajima, F., 1989 Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Verrelli, B. C., and W. F. Eanes, 2001 Clinal variation for amino acid polymorphisms at the *Pgm* locus in *Drosophila melanogaster*. Genetics **157:** 1649–1663.

Wang, R. L., A. Stec, J. Hey, L. Lukens and J. Doebley, 1999 The limits of selection during maize domestication. Nature **398:** 236–239.

Weigel, D., 1995 The genetics of flower development: from floral induction to ovule morphogenesis. Annu. Rev. Genet. **29:** 19–39.

Weigel, D., and E. M. Meyerowitz, 1993 Activation of floral hometic genes in *Arabidopsis*. Science **261:** 1723–1726.

Weigel, D., J. Alvarez, D. R. Smyth, M. F. Yanofsky and E. M. Meyerowitz, 1992 *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell **69:** 843–859.

Yanofsky, M. F., 1995 Floral meristems to floral organs—genes controlling early events in *Arabidopsis* flower development. Annu. Rev. Plant Physiol. Plant Mol. Biol. **46:** 167–188.

Communicating editor: M. K. Uyenoyama