Population Genomics of the *Arabidopsis thaliana* Flowering Time Gene Network

Jonathan M. Flowers,[1] Yoshie Hanzawa,*†[1] Megan C. Hall,* Richard C. Moore,‡ and Michael D. Purugganan**

*Department of Biology and Center for Genomics and Systems Biology, New York University; †Department of Crop Sciences and Institute for Genome Biology, University of Illinois at Urbana-Champaign, 1201 W Gregory Drive, Urbana, Illinois 61801; and ‡Department of Botany, Miami University, Oxford

The time to flowering is a key component of the life-history strategy of the model plant *Arabidopsis thaliana* that varies quantitatively among genotypes. A significant problem for evolutionary and ecological genetics is to understand how natural selection may operate on this ecologically significant trait. Here, we conduct a population genomic study of resequencing data from 52 genes in the flowering time network. McDonald–Kreitman tests of neutrality suggested a strong excess of amino acid polymorphism when pooling across loci. This excess of replacement polymorphism across the flowering time network and a skewed derived frequency spectrum toward rare alleles for both replacement and noncoding polymorphisms relative to synonymous changes is consistent with a large class of deleterious polymorphisms segregating in these genes. Assuming selective neutrality of synonymous changes, we estimate that approximately 30% of amino acid polymorphisms are deleterious. Evidence of adaptive substitution is less prominent in our analysis. The photoperiod regulatory gene, *CO*, and a gibberellic acid transcription factor, *AtMYB33*, show evidence of adaptive fixation of amino acid mutations. A test for extended haplotypes revealed no examples of flowering time alleles with haplotypes comparable in length to those associated with the null fri_{Col} allele reported previously. This suggests that the *FRI* gene likely has a uniquely intense or recent history of selection among the flowering time genes considered here. Although there is some evidence for adaptive evolution in these life-history genes, it appears that slightly deleterious polymorphisms are a major component of natural molecular variation in the flowering time network of *A. thaliana*.

Introduction

A significant challenge for evolutionary and developmental genetics is to identify the genes that control fitness-related traits and to characterize the molecular polymorphisms that contribute to phenotypic variation in natural populations (Ehrenreich and Purugganan 2006). Substantial progress has been made in characterizing the relationship between genotype and phenotype (Alonso-Blanco et al. 2005; Mitchell-Olds and Schmitt 2006), yet despite considerable effort, the evolutionary mechanisms that operate on the genes that control these traits remain poorly understood. A promising way forward has been to examine the genetic architecture of morphological traits and to establish the contribution of different mechanisms including adaptation, constraint, and other factors that govern natural trait variation. In a few cases, ecological genetic and other experimental approaches have determined the evolutionary mechanisms that control the fate of functional polymorphisms (e.g., Subramaniam and Rausher 2000; Scarcelli et al. 2007; Coberly and Rausher 2008). These studies have yielded important insight into the contributions of epistasis, pleiotropy, and adaptation in the evolution of phenotypic traits.

In most cases, little is known about the molecular basis of natural variation in a trait or how natural selection has operated on the genes that control the phenotype. A complementary approach has been to use population genetic methods to infer the incidence of natural selection in the networks that underlie ecologically significant traits. Examples include studies of the immune system (e.g., Schlenke and Begun 2003; Bakker et al. 2006) and pigmentation pathways (e.g., Rausher et al. 2008), which have frequently found evidence for a variety of selective processes operating on standing genetic variation. For example, in a study of 27 *R* genes of the nucleotide-binding site—leucine-rich repeat family of disease resistance loci, 14 were shown to fall in the 5% tails of the genomewide empirical distribution for one or more population genetic statistics (Bakker et al. 2006). Many of these loci possess high levels of polymorphism and skews in the site frequency spectrum toward high-frequency alleles consistent with diversifying selection for pathogen resistance. These studies of the incidence of natural selection at the molecular level have provided important examples of the diversity of evolutionary mechanisms that shape phenotypic variation in adaptive traits.

The transition to flowering is a major quantitative trait that defines the developmental shift from vegetative to reproductive development in flowering plants (Koornneef et al. 2004; Engelmann and Purugganan 2006). In Arabidopsis, this transition is induced by both internal signals, such as those associated with the plant hormone gibberellic acid, and external cues that include photoperiod, day length, vernalization, ambient temperature, and nutrient status (Engelmann and Purugganan 2006). This transition is a critical point in the life history of flowering plants because it synchronizes the development of reproductive structures with environmental conditions suitable for reproduction. Because *Arabidopsis thaliana* populations are found in highly variable environments, it may be expected that flowering time variation in this species is shaped in part by selection on standing variation. Moreover, the evolution of flowering time may also be expected to extend to adaptive differences in life-history traits between *A. thaliana*, an annual, and its perennial sister species *Arabidopsis lyrata*.

The flowering time genetic network in *A. thaliana* represents one of the best-studied functional systems in plants (Mouradov et al. 2002; Komeda 2004; Bäurle and Dean 2006). This large genetic network consists of genes representing a variety of different light sensing, hormone signaling, and developmental pathways (Mouradov et al. 2002;

[1] These authors contributed equally to the paper.

Key words: population genomics, gene network, SNPs, selection, deleterious mutations, adaptive evolution.

E-mail: mp132@nyu.edu.

Mol. Biol. Evol. 26(11):2475–2486. 2009
doi:10.1093/molbev/msp161
Advance Access publication July 22, 2009

© The Author 2009. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

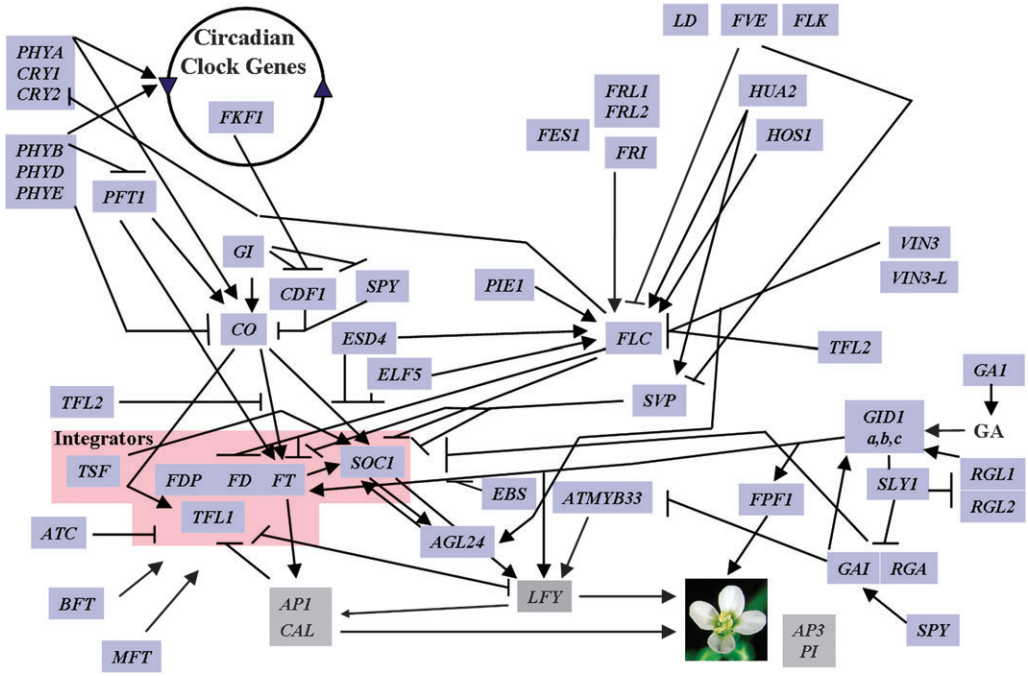

FIG. 1.—The flowering time gene network in *Arabidopsis thaliana*. The interactions of the genes used in this study are inferred based on a compilation of previously published genetic studies (adapted from Ehrenreich et al. forthcoming).

Komeda 2004; Bäurle and Dean 2006) that produce various signals that are interpreted by a set of floral integrator genes. Although many of these genes were discovered by forward genetic screens, there have recently been concerted efforts to determine the genetic basis of natural flowering time variation in *A. thaliana* populations. Quantitative trait locus mapping analyses (Clarke et al. 1995; Kuittinen et al. 1997; Ungerer et al. 2002, 2003; Weinig et al. 2002) and association studies (Zhao et al. 2007; Ehrenreich et al. forthcoming) have identified chromosomal regions that harbor loci associated with flowering time. These map-based techniques have led to identification of causal polymorphisms that contribute to flowering time variation in the genes *FRI* (Johanson et al. 2000), *FLC* (Shindo et al. 2005), *CRY2* (El-Assal et al. 2001), *PHYC* (Balasubramanian et al. 2006), *FRL1* and *2* (Schläppi 2006), *HUA2* (Doyle et al. 2005; Wang et al. 2007), and *FLM* (Werner et al. 2005).

A number of studies have examined the evolution of flowering time genes (Le Corre et al. 2002; Olsen et al. 2002, 2004; Caicedo et al. 2004; Stinchcombe et al. 2004; Toomajian et al. 2006; Filiault et al. 2008) and in some cases have found evidence of adaptive evolution (Le Corre 2005; Toomajian et al. 2006). However, there has been no large-scale study of the evolution of the flowering time network and therefore tests of hypotheses concerning the operation of selection in this system of genes have not been addressed (Olsen et al. 2004). Here, we report the analysis of resequencing data from 52 genes that represent a large fraction of genes found to control flowering time in forward genetic screens (Mouradov et al. 2002; Komeda 2004; Bäurle and Dean 2006). The selected genes include members of the vernalization, gibberillic acid, photoperiod, and integrator pathways and a diversity of protein functions including DNA-binding transcription factors, photoreceptor genes, signaling loci, and hormone biosynthetic genes (see fig. 1). Analysis of polymorphism and divergence with *A. lyrata* revealed limited evidence of adaptive evolution across the flowering time network. Instead, there appears to be widespread purifying selection in this system of genes with a large fraction of amino acid and noncoding polymorphisms inferred to be slightly deleterious.

Materials and Methods

DNA Sequencing

Genomic DNA was isolated from seedlings (14 days old after germination) using Qiagen DNeasy Plant maxi kits (Qiagen, CA, United States). Twenty-four *A. thaliana* accessions (see supplementary table S1, Supplementary Material online) were chosen randomly from the 96 accessions sequenced by Nordborg et al. (2005) and include a broad range of early- to late-flowering time phenotypes. Under controlled long-day conditions, the range of mean flowering times in these accessions are 38.3 days to 77.1 days, with a mean of 46.01 ± 8.73 days.

The primers used for polymerase chain reaction (PCR) and direct sequencing were designed based on the Col-0 genomic sequence obtained from the Arabidopsis Information Service (TAIR) site (Swarbreck et al. 2008) and spaced approximately 500 bp apart using Primer3 (http://frodo.wi.mit.edu/) (see supplementary table S2, Supplementary Material online). Neighboring products were designed to overlap each other by approximately 50–100 bp. The sequencing was designed to span the entire gene including approximately 1 kb of the promoter region and approximately 500 bp of the 3′ flanking region. In cases where

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

a gene is alternatively spliced, the representative gene model defined by TAIR was used. Forty-seven flowering time genes described in the literature were chosen for analysis, including representatives of the photoperiod, vernalization, gibberellic acid and floral integrator pathways (Mouradov et al. 2002; Bäurle and Dean 2006). Primer synthesis, quality tests for the primers, PCR amplification, purification of the PCR products, and direct sequencing reactions were performed by Cogenics (NC, United States).

Sequences were assembled using a bioinformatics pipeline developed by Chris Smith (NC State University, United States) that incorporates Phred/Phrap (CodonCode, Dedham, MA) and Biolign v. 2.0.9 (Tom Hall, Ibis Therapeutics, Carlsbad, CA). For regions that had low-quality sequence data, new primer sets were designed and the sequencing procedure described above was repeated up to three times. Assembled sequences were aligned against the Col-0 genomic sequences. All polymorphisms observed in the aligned sequences were manually confirmed by referring to the original trace files. *Arabidopsis thaliana* sequences for *FRI* (Stinchcombe et al. 2004), *FLC* (Caicedo et al. 2004), *CRY2* (Olsen et al. 2004), *LFY*, and *TFL1* (Olsen et al. 2002) were from previous studies. The accessions sequenced in these studies partially overlapped those used in the present work. All sequences in this study were deposited in Genbank with accession numbers GQ176870–GQ177973.

Assembly of *A. lyrata* Sequences

Local assemblies for *A. lyrata* were based on sequence traces retrieved from the NCBI trace archive database (http://www.ncbi.nlm.nih.gov/Traces/home/). Traces were obtained by querying the *A. lyrata* sequence trace repository with a Blast protocol (Altschul et al. 1990). All hits with *e*-values less than 0.01 were retained, subjected to a vector removal step using crossmatch (implemented at http://egassembler.hgc.jp/), and used in a reverse Blast query against the Col-0 genome sequence. Traces were retained for assembly if they met a reciprocal best match criterion. Local assemblies were generated with CAP3 (Huang and Madan 1999) using the traces and PHRED quality scores as input. Initial clipping of low-quality ends was performed using CAP3 default settings. Contigs were then manually inspected for assembly artifacts and trimmed to remove low-quality ends. Base calls with CAP3 contig quality scores of less than 50 were marked as ambiguous (Huang XQ, personal communication).

Multiple sequence alignments of *A. thaliana* accessions were generated with Biolign. A separate set of alignments that included the *A. lyrata* sequence was generated with MUSCLE v. 3.6 (Edgar 2004). All alignments including the *A. lyrata* sequence were visually inspected and minor adjustments were made manually. Large noncoding regions that could not be aligned were removed prior to analysis. Gene annotations are based on TAIR seven genome release (Swarbreck et al. 2008).

Data Analysis

Population genetic summary statistics were estimated for flowering time genes and a set of 1,102 background fragments from *A. thaliana* originally reported by Nordborg et al. (2005) and Toomajian et al. (2006) using the libsequence package (Thornton 2003) or scripts written in Perl. In these analyses, the same set of 25 accessions was used for the flowering time genes and background loci, except for genes *FRI*, *FLC*, *CRY2*, *LFY*, and *TFL1* as noted above. A second set of 572 background alignments with an *A. lyrata* outgroup sequence (Foxe et al. 2008, Wright SI, personal communication) was incorporated in comparisons of the derived nonsynonymous and synonymous site frequency spectra (SFS) and fixation data with the flowering time genes. These alignments contained only coding regions, totaled 195 kb in length, and were also filtered to include only the set of 25 accessions sequenced in the present study. In addition, accessions with premature stops or frameshift mutations were removed from these alignments prior to analysis and heterozygous sites were resolved randomly. This was done to allow for direct comparison of background and flowering time genes, to which we had applied the same criteria. A separate analysis of the derived SFS for noncoding DNA in the flowering time data set was conducted after removing all sites within 10 bp of indels. This was done to limit the mis-assignment of the derived allele due to uncertainty in alignments of noncoding DNA between *A. thaliana* and *A. lyrata* in sites flanking gaps (DeRose-Wilson and Gaut 2007).

A pairwise haplotype sharing (PHS) test for extended haplotypes (Toomajian et al. 2006) was conducted using tag single nucleotide polymorphisms (SNPs) in flowering time genes that were genotyped in an association mapping study (Ehrenreich et al. forthcoming) and the 1,102 background fragments from Toomajian et al. (2006). The accessions included 92 of the 96 accessions (excluding NFA-8, Ren-11, CS22491, and Fab-2) sequenced by Toomajian et al. (2006), which facilitated estimation of PHS scores for flowering time alleles for comparison with *FRI* null alleles and the background empirical distribution. The PHS score (see eqs. 1 and 2 in Toomajian et al. 2006) aims to measure the average length of shared haplotypes flanking a focal allele found in at least two accessions, while correcting for population substructure. The test was implemented in a few different ways (e.g., by allowing or disallowing the shared haplotype flanking a focal allele to be disrupted by mutations in the focal fragment or flowering time gene). This was done to check the robustness of extended haplotypes to minor changes in the details of the implementation of the test and to check for possible biases introduced by combining data from sequenced regions of different length (i.e., flowering time genes are longer than the background fragments), although the PHS statistic is such that it may be robust to this type of effect (see Toomajian et al. 2006, eq. 1). Additional method details were intended to replicate as closely as possible those outlined in Toomajian et al. (2006). Polymorphic sites with fewer than 60 alleles (due to ambiguities) were not considered and only the allele with the highest PHS score was retained from a set of correlated sites (i.e., sites in linkage disequilibrium, $r^2 > 0.5$) to reduce redundancy in the distribution of scores. The genetic map for PHS calculations was constructed using marker data reported in Toomajian et al. (2006) and the test implemented using custom Perl scripts.

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

McDonald–Kreitman (MK) tests of neutrality (McDonald and Kreitman 1991) were conducted using the MK test program (Thornton 2003) and significance assessed with a Fisher's Exact Test (FET) implemented in the R statistical package version 2.6.1 (www.R-project.org). Tests that involved pooling mutations across loci were conducted using the bootstrap procedure of Shapiro et al. (2007). Estimates of nonsynonymous (K_a) and synonymous (K_s) rates of evolution were obtained using the method of Comeron (1995) with gestimator (Thornton 2003). Estimates of the average strength of selection ($2N_e$s) of amino acid mutations were obtained using a Bayesian implementation of Poisson random field with MKPRF (Bustamante et al. 2002). In this analysis, the entries from the MK table for each gene are modeled as Poisson random variables following Sawyer and Hartl (1992). Model parameters θ_S, θ_R, τ, and $2N_e$s were estimated from the data using a Bayesian method assuming independence among sites and neutrality of synonymous mutations following methodology and equations in Bustamante et al. (2002). In the present analysis, background and flowering time genes were specified as separate classes. The posterior distribution of $2N_e$s was constructed using Markov Chain Monte Carlo (MCMC) sampling and reported estimates are the mean of the posterior distribution for each gene. The analysis was implemented by the Cornell Computational Biology Service Unit at http://cbsuapps.tc.cornell.edu/mkprf.aspx. Alignments with fewer than four nonsynonymous mutations (polymorphisms or fixations) were filtered prior to analysis to eliminate genes that are uninformative about selection (Bustamante et al. 2005). MCMC sampling of the posterior distribution was performed with a burn-in of 1,000 steps followed by retaining 5,000 draws from each of 10 chains. All other settings were the MKPRF defaults. Genes were considered to deviate from neutrality if the 95% posterior quantiles of $2N_e$s did not include zero with values greater than zero interpreted as evidence of positive selection and values less than zero as evidence of purifying selection of slightly deleterious polymorphism (Bustamante et al. 2002).

Results and Discussion

Nucleotide Variation in the Flowering Time Gene Network

We investigated the molecular population genetics of 52 *A. thaliana* flowering time genes that represent a majority of known flowering time loci in the photoperiod, vernalization, gibberellic acid, and floral integrator pathways (see figs. 1 and 2, supplementary fig. S1, Supplementary Material online). For 47 of the genes, we resequenced a region that spans approximately 1 kb of noncoding DNA upstream of the annotated gene, the 5′ untranslated region (UTR), all exons, introns, and the 3′ UTR in 24 accessions. A total of approximately 86.5 kb of sequence was obtained for each accession. The sequences of the other five genes used in this study were from previous studies (Olsen et al. 2002, 2004; Caicedo et al. 2004; Stinchcombe et al. 2004) and included the coding region and at least a portion of the promoter and 5′ UTR.

Population summary statistics estimated from flowering time gene regions and from 1,102 background fragments from previous studies (Nordborg et al. 2005; Toomajian et al. 2006) revealed few examples where flowering time loci fall in the tails of the background empirical distributions for either nucleotide diversity or Tajima's *D* (Tajima 1989). The number of polymorphic sites per flowering time gene ranged from 18 to 174 per gene, with per site nucleotide diversity (π, Nei 1987) estimates ranging from 0.0009 to 0.0114 (see fig. 2 and supplementary table S3, Supplementary Material online). These values and diversity estimates based on the number of segregating sites, θ_W (Watterson 1975), are well within the range observed for the background fragments. The average π of 0.0037 for the flowering time set is comparable to average estimates of genomewide nucleotide diversity based on background coding regions (Nordborg et al. 2005). Similarly, the mean Tajima's *D* of −0.77 for flowering time genes is comparable to the background average of −0.67 for the 25 accessions used in this study. A single observation of Tajima's *D* of −2.30 from *FRL2* falls outside the 95th percentile of the empirical distribution for this statistic (see supplementary table S3, Supplementary Material online). Tests for differences between the empirical distribution and flowering time genes with a Mann–Whitney *U* test (MWU) were not significant for any of these population test statistics ($P > 0.05$).

Nonfunctional alleles caused by deletions in coding regions have been previously identified in the *FRI* locus, and such null alleles are a potentially significant form of adaptive variation in the flowering time network (Johanson et al. 2000; Le Corre et al. 2002; Caicedo et al. 2004; Stinchcombe et al. 2004; Toomajian et al. 2006). We observed premature stop codons that may lead to loss-of-function alleles in three additional flowering time genes—*FRL1*, *FRL2*, and *PHYD* (see supplementary fig. S1, Supplementary Material online). *FRL1* has a GAG→TAG nonsense mutation at codon position 279 (Glu279) in accessions Ws-0, Wei-0, Ct-1, and Mrk-0. This premature stop is 192 codon positions upstream of the annotated stop in Col-0 and is the same nonsense mutation previously reported in the nonfunctional *FRL1* allele in Landsberg *erecta* (*Ler*, Schläppi 2006). In *FRL2*, a CAA→TAA change in Cvi-0 at codon position 436 (Pro436) introduces a premature stop codon, 38 codon positions upstream of the Col-0 stop. In *PHYD*, probable null alleles associated with premature stop codons have been reported in accessions Ita-0 and Gre-0 (Mathews and McBreen 2008). In our accessions, we observed five independent frameshifting deletions in Sorbo (coding positions 385–386), Ei-2 (1135–1147), Ga-0 (739–739), Mt-0 (2529–2529), and a 14-bp deletion in Ag-0 (88–101) that has been reported previously in Ws (Aukerman et al. 1997). The deletions in four of these accessions (Ag-0, Sorbo, Ei-2, and Ga-0) shift the reading frame upstream of the functionally important phytochrome domain and all five accessions have premature stops relative to Col-0. None of the genes in our analysis showed evidence of being a pseudogene in *A. lyrata*.

Recent positive selection is expected to generate linkage disequilibrium over extended physical distances due to a rapid increase in the frequency of a selected haplotype.

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

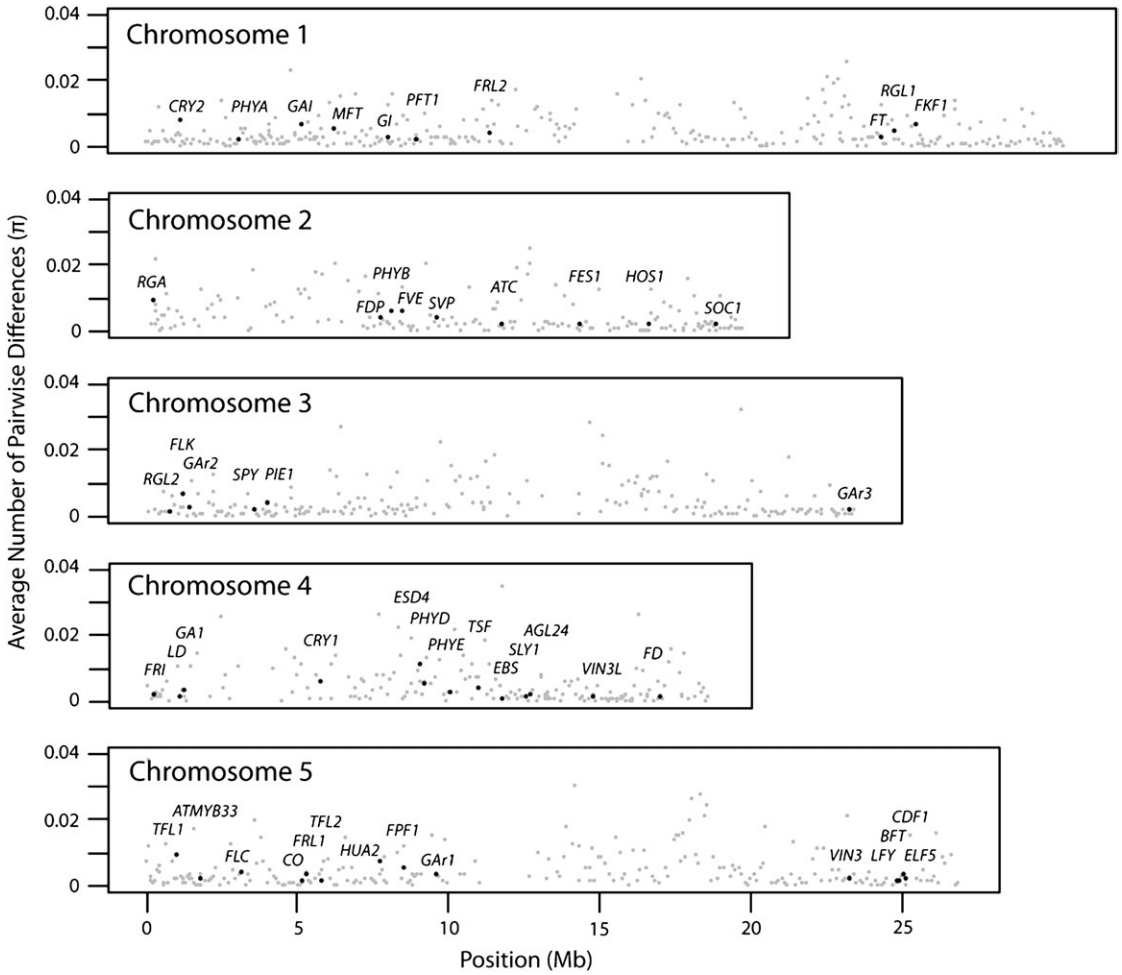

FIG. 2.—Nucleotide diversity of flowering time genes in *Arabidopsis thaliana*. Flowering time genes are marked with black dots and background loci with gray dots.

A recent survey of genomewide haplotype structure in *A. thaliana* identified long-range haplotypes associated with fri_{Col} and fri_{Ler} gene deletions, suggesting that the early flowering phenotype has been under recent directional selection (Toomajian et al. 2006). We wanted to determine whether the *FRI* deletions were unique or whether other variants in flowering time genes were also associated with putative partial selective sweeps. To accomplish this, we calculated the PHS scores (Toomajian et al. 2006) for flowering time and background SNPs from a set of 92 accessions (see Materials and Methods). In our analysis, the extended haplotype associated with the fri_{Col} deletion is a clear outlier both with respect to background and flowering time SNPs (see fig. 3, supplementary table S5, Supplementary Material online). This suggests a uniquely intense or recent history of directional selection at *FRI* compared with other flowering time genes. Extended haplotypes also flank alleles at *GA1* (A at position 1241741) and *LD* (T at position 1123428). These alleles are in the 99th percentile of the empirical length distribution of PHS (see fig. 3, supplementary table S5, Supplementary Material online) and are notable because they may be hitchhiking with variation at *FRI*. Both genes are found approximately 1 Mb downstream of *FRI* on the proximal arm of chromosome 4 (see fig. 2) in a region previously identified by Toomajian et al. (2006) as harboring a number of SNPs with significantly long haplotype blocks. A number of additional flowering time alleles show unusually long haplotypes (supplementary table S5, Supplementary Material online). However, a few of these alleles (*FD*, 17006355*T; *CO*, 5173514*A; *FPF1*, 8542350*T; and *GAr1*, 9628010*G) must be considered with some caution because they were insignificant when minor adjustments were made to the implementation of the PHS method (data not shown). Although extended haplotypes for these and other SNPs in the tails of the empirical distribution (supplementary table S5, Supplementary Material online) are consistent with expectations for partial selective sweeps, this approach is not a formal test of statistical significance for selection and only indicate that these loci are in extreme regions of the genomewide haplotype length distribution.

Selection and Flowering Time Protein Evolution

The MK test is based on the neutral expectation that the ratio of nonsynonymous to synonymous mutations will be equivalent for intraspecific polymorphism and

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

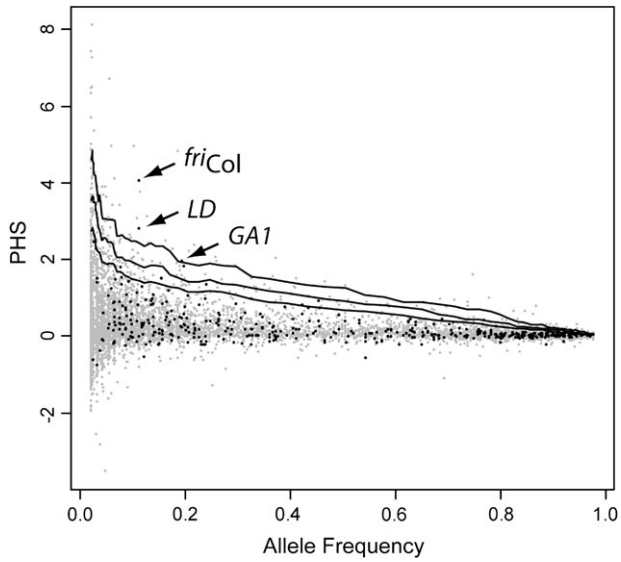

FIG. 3.—Degree of PHS of flowering time compared with background SNPs. PHS is plotted against allele frequency for flowering time (black) and background alleles (gray). The percentile lines (95th, 97.5th, and 99th) are based on sliding windows of 1,000 data points with step size of 100. Flowering time alleles above the 99th percentile are labeled. To reduce redundancy among sites that are in linkage disequilibrium ($r^2 > 0.5$), the single allele with the highest PHS score was retained from pairs of correlated SNPs. Genotypes for fri_{Ler} and fri_{Col} are from Toomajian et al. (2006). The genes *ESD4*, *PHYD*, *CRY2*, *LFY*, *TFL1*, and *FLC* were not included in the analysis.

between-species divergence (McDonald and Kreitman 1991). After removing sequences with premature stops or frameshifting mutations in *FRI*, *PHYD*, *FRL1*, and *FRL2* alignments, MK tests on four genes (*PHYD*, $P = 2.83 \times 10^{-5}$; *PIE1*, $P < 0.05$; *RGL1*, $P < 0.05$, and *TFL1*, $P < 0.05$) (see supplementary table S4, Supplementary Material online) deviated significantly from neutrality in the direction of an excess of replacement polymorphism, although only *PHYD* remained significant after Bonferroni correction. Evidence of seven independent frameshifting mutations in *PHYD* together with the excess of replacement polymorphisms in the MK test (this study, Mathews and McBreen 2008) may reflect a recent relaxation of selective constraints on this gene and possibly the initial stages of pseudogenization. An alternative explanation is that putatively null frameshift alleles and amino acid polymorphisms are adaptive in some environments in a fashion analogous to *FRI*. These possibilities cannot be clearly distinguished with the data presented here, although it is important to note that *PHYD* polymorphisms have also been shown to be significantly associated with flowering time variation in both natural accessions and a recombinant inbred intercross mapping population (Ehrenreich et al. forthcoming).

The combination of an excess of amino acid polymorphism and probable loss-of-function alleles at *PHYD* is unusual in *A. thaliana* (but see Le Corre et al. 2002; Shindo et al. 2005). However, excesses of nonsynonymous polymorphism relative to divergence are common in Arabidopsis genes and are typically attributed to weak selection against deleterious amino acid mutations (see below, Weinreich and Rand 2000; Bustamante et al. 2002; Wright and Andolfatto 2008). Because slightly deleterious mutations are expected to contribute more to polymorphism than divergence (Kimura 1983), they could obscure the signature of positive selection. Therefore, we repeated the MK tests by excluding singleton nonsynonymous and synonymous polymorphisms (Fay et al. 2001). After removal of singletons, *FLK* is significant in the direction of an excess of replacement fixations between *A. thaliana* and *A. lyrata* ($P = 0.026$, see supplementary table S4, Supplementary Material online), and *PHYD* remains significant ($P < 0.001$) in the direction of an excess of replacement polymorphism within *A. thaliana* (see supplementary table S4, Supplementary Material online).

Estimates of the average selection coefficient on amino acid changes were performed on background and flowering time genes using a Poisson Random Field framework (see fig. 4). The mean of the posterior probability distribution of the scaled selection coefficient, $2N_e s$, was significantly less than zero for genes *PIE1*, *CRY2*, *PHYD*, and *RGL1*. This is consistent with the nominally significant excesses of replacement polymorphism in individual MK tests for three of the four genes (see above). In addition, we observed values of $2N_e s$ greater than zero for genes *CO* and *AtMYB33* (see fig. 4), which suggests positive selection for protein divergence between species. However, we observed no global difference in the distribution of selection coefficients on background and flowering time genes with a minimum of four nonsynonymous mutations (MWU test, $P > 0.05$).

Analysis of polymorphism and divergence was also conducted by pooling nonsynonymous (A) and synonymous (S) mutations across loci. For this analysis, we compared 572 coding region background loci (Foxe et al. 2008, Wright SI, personal communication) with the 52 flowering time genes. At both background and flowering time loci, we observed a strong excess of replacement polymorphism within *A. thaliana* compared with between-species divergence. The flowering time loci have an A/S ratio of 0.698 for within-species polymorphism and 0.575 for divergence (FET, $P = 0.01198$), whereas for background genes, the A/S ratios for polymorphism and divergence are 0.872 and 0.665 (FET, $P = 1.6 \times 10^{-9}$). The lower ratios for flowering time genes for both polymorphism and divergence could indicate a greater level of constraint on protein sequences at these genes compared with genome background loci. However, no difference in K_a or K_a/K_s was observed when comparing the flowering time and background distributions (MWU, $P > 0.05$).

Pooling data across recombining loci can, however, introduce biases in the MK test (McDonald-Kreitman 1991; Shapiro et al. 2007). To circumvent this, we employed a bootstrap procedure to assess statistical significance (Shapiro et al. 2007). Application of this method to the flowering time data again suggests a large excess of replacement polymorphism ($P < 0.0005$) with the observed fixation index (FI = A/S for divergence/A/S for polymorphism) of 0.840 being lower than the lowest FI of 0.885 observed in a set of 10,000 bootstrap replicates. When genes that individually showed a significant excess of replacement polymorphism in MK tests are removed, the FI increases to 0.998 but remains significant ($P < 0.05$).

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

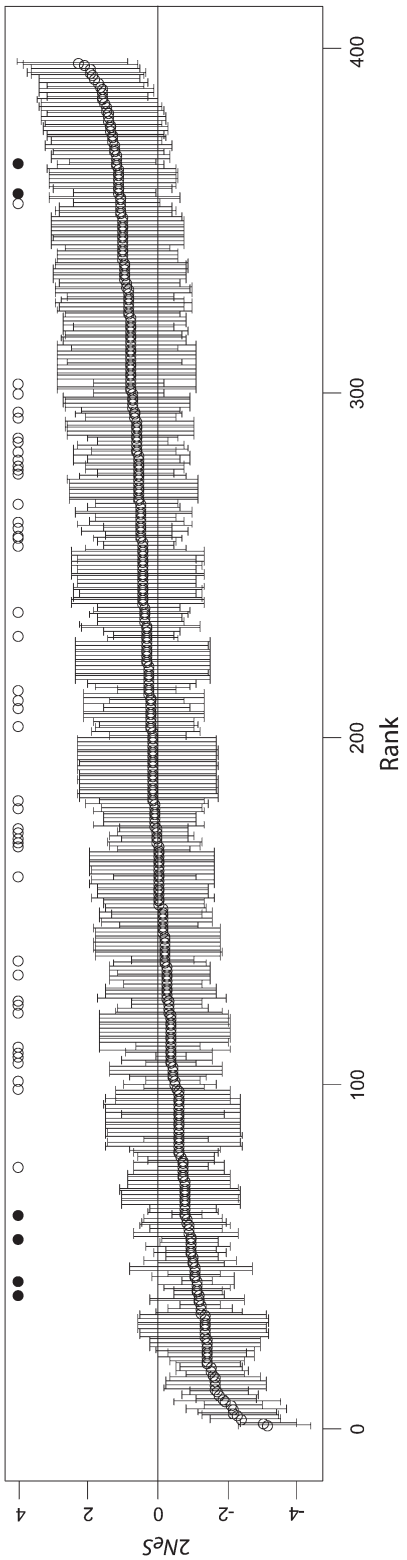

FIG. 4.—Rank order of N_es estimates for flowering time and background loci for the MKPRF analysis. Flowering time genes are denoted by circles at the top of the figure. Error bars are 95% credibility intervals of the Bayesian posterior distribution. Open circles represent genes where the credibility interval includes zero. Closed circles represent genes whose interval does not include zero.

This indicates that the excess of replacement polymorphisms in the flowering time gene network is also found in loci other than those that are significant in individual tests. The low FI values and accompanying excesses of replacement polymorphism within *A. thaliana* in this and other studies (Bustamante et al. 2002) suggests that there is a large class of amino acid polymorphisms that segregate at detectable frequencies but are unlikely to become fixed and contribute to between-species divergence.

Segregating Deleterious Polymorphisms in the Flowering Time Gene Network

Excesses of replacement polymorphism in MK tests have been interpreted as evidence that a significant amount of the standing variation in *A. thaliana* populations may be deleterious (Weinreich and Rand 2000; Bustamante et al. 2002). Moreover, slightly deleterious polymorphisms are expected to segregate at lower frequency due to purifying selection. To examine this, we constructed the unfolded site frequency spectrum by polarizing mutations using parsimony with *A. lyrata* as the outgroup. Derived nonsynonymous polymorphisms on average segregate at lower frequencies than derived synonymous polymorphisms in both background and flowering time genes, and the A/S ratio is larger for low-frequency polymorphisms compared with those at higher frequencies (see fig. 5). For flowering time genes, the A/S ratio of 1.03 (198 nonsynonymous/193 synonymous polymorphism) for derived polymorphisms segregating at less than 10% frequency is significantly different from the ratio of 0.5 observed for higher frequency SNPs (163 nonsynonymous/325 synonymous polymorphisms) (FET, $P < 3.1 \times 10^{-7}$). This observation is inconsistent with a neutral distribution of alleles (Fay et al. 2001).

The fraction of deleterious replacement polymorphisms in flowering time genes was estimated by calculating the excess of amino acid polymorphisms segregating at low frequency (see fig. 5) compared with a putative class of largely neutral polymorphisms segregating at moderate- to high-frequency (Fay et al. 2001; Shapiro et al. 2007). Given the A/S ratio for high-frequency (>10%) polymorphisms (163/325 = 0.502, see fig. 5) and the A/S ratio for all polymorphisms (361/518 = 0.697), we estimated the fraction of slightly deleterious amino acid polymorphisms in the flowering time network to be (0.697–0.502)/0.697 = 28% and the background gene fragments to be 24%. These results are largely insensitive to the frequency cutoff used to specify the high-frequency alleles enriched for neutral polymorphisms. Changing the cutoff to 15%, as suggested by Charlesworth and Eyre-Walker (2008), 20%, or 25% increases the estimates by 2–3% for flowering time and background loci (data not shown).

A comparable analysis of the site frequency spectrum can be conducted on noncoding DNA using the ratio of noncoding to synonymous (N/S) SNP counts. When derived polymorphisms segregating at less than 10% frequency are compared with higher frequency SNPs, we observed a significantly greater N/S ratio for low-frequency polymorphisms in noncoding DNA (fig. 6). For example, the N/S ratio for intergenic DNA is 0.82

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

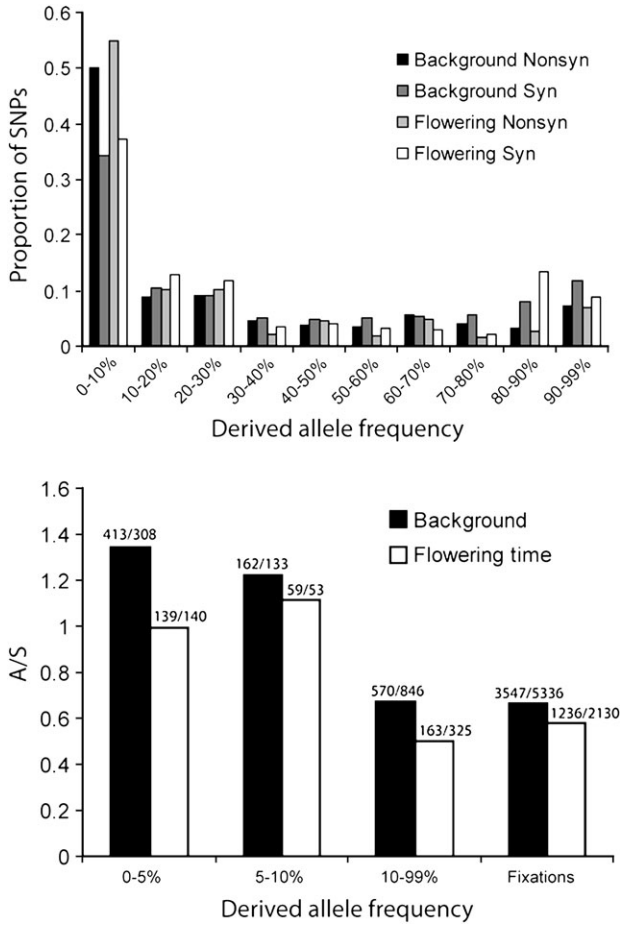

FIG. 5.—Site frequency spectra for flowering time gene polymorphisms. (*A*) Derived site frequency spectra for background and flowering time loci. (*B*) A/S ratios for polymorphism and divergence for background and flowering time genes. Numbers above bars represent SNP counts.

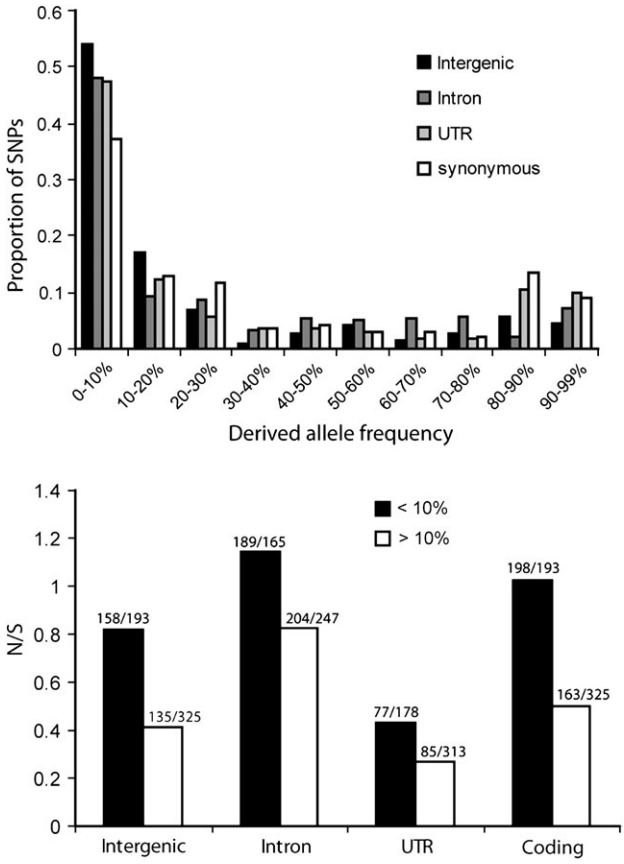

FIG. 6.—Site frequency spectra for noncoding regions of flowering time genes. (*A*) Derived site frequency spectra for intergenic regions (i.e., nontranscribed DNA), introns, 5′ and 3′ UTRs, and synonymous SNPs. (*B*) Ratios of noncoding to synonymous (N/S) SNP counts across the derived site frequency spectrum for different classes of noncoding DNA. The A/S ratio for coding regions from figure 5 is shown for comparison. Noncoding to synonymous ratios are presented for derived alleles with low (0–10%) and high (10–99%) frequency. UTRs are defined as all transcribed noncoding sequence up- or downstream of the first or last translated exon (i.e., introns and exons in the UTRs are combined). Numbers above bars represent mutational counts.

for low-frequency polymorphisms and 0.42 for high-frequency polymorphisms ($P < 0.00005$, FET). Similar results were also observed for introns ($P < 0.05$) and UTRs ($P < 0.05$). This skew in the allele frequency distribution of noncoding SNPs toward rare variants is comparable to that observed for replacement polymorphism and similarly suggests a large class of deleterious polymorphisms in noncoding regions of the flowering time gene network.

Finally, examination of the A/S ratio for divergence with the A/S ratio for high-frequency polymorphism revealed no evidence of adaptive substitution when pooling across loci. Although the observed A/S ratio for fixations between *A. thaliana* and *A. lyrata* is slightly higher relative to A/S for the high-frequency "neutral" fraction as expected for positive selection (FI = 1.16, fig. 5), this difference is not significant for flowering time genes or background loci using either an FET or the bootstrapping approach (FET for flowering time, $P > 0.16$; FET for background loci $P > 0.8$). Therefore, in contrast to comparable studies in Drosophila (Fay et al. 2002; Shapiro et al. 2007; Wright and Andolfatto 2008), we found no evidence of adaptive fixations between *A. thaliana* and *A. lyrata* for flowering time genes using this methodology.

Patterns of Protein Evolution Differ between Flowering Time Pathways

We have shown that there is an excess of low-frequency amino acid polymorphisms across the flowering time network. This system of genes, however, is comprised of several interacting pathways that play distinct roles in modulating and/or integrating environmental and hormonal signals (Mouradov et al. 2002), and genes in these pathways may thus experience various patterns of molecular evolution due to any of several factors, such as variation in levels of pleiotropic constraints. To examine possible differences, we partitioned genes into the photoperiod, vernalization, gibberellic acid, and floral integrator pathways and evaluated whether proteins are evolving differently among them.

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

Table 1
Polymorphism and Divergence Summary Statistics for Individual Pathways in the Flowering Time Network

	π_A	π_S	π_A/π_S	K_a	K_s	K_a/K_s
Photoperiod	0.0012 (0.0002)	0.0099 (0.0029)	0.1898 (0.0485)	0.0181 (0.0043)	0.1204 (0.0075)	0.1455 (0.0318)
Vernalization	0.0014 (0.0003)	0.0051 (0.0009)	0.3762 (0.0849)	0.0312 (0.0036)	0.1135 (0.0114)	0.2838 (0.028)
Integrator	0.0006 (0.0002)	0.0029 (0.001)	0.6378 (0.3006)	0.0205 (0.0045)	0.1098 (0.0077)	0.1988 (0.0516)
Gibberillic Acid	0.0006 (0.0002)	0.0083 (0.0027)	0.166 (0.063)	0.0167 (0.0022)	0.1529 (0.0164)	0.1178 (0.0209)
All flowering time	0.0009 (0.0001)	0.0063 (0.001)	0.3512 (0.0819)	0.0221 (0.002)	0.1239 (0.0061)	0.1918 (0.0192)
Background[a]	0.0027 (0.0003)	0.0126 (0.0012)	0.319 (0.0305)	0.0336 (0.002)	0.1618 (0.0071)	0.2347 (0.0134)

Values represent means ± standard error (in parentheses).
[a] Based on 336 background fragments with a minimum length of 100 codons.

Within *A. thaliana*, we examined levels of nonsynonymous (π_A) and synonymous (π_S) nucleotide diversity at each of the functional pathways (see table 1). Levels of π_A are significantly different among pathways using a Kruskal–Wallis (KW) test ($\chi^2 = 8.3048$, df = 3, $P = 0.04$), with π_A for photoperiod, gibberellic acid, and integrator genes approximately 2-fold lower than those of the vernalization pathway. However, posthoc MWU tests between pathways are not significant for any of the six pairwise comparisons after applying a Bonferroni correction for multiple tests. Moreover, the variation in π_S among pathways is also significant (KW test, $\chi^2 = 8.4154$, $P = 0.038$), and this may contribute to the insignificance of a test of differences in π_A/π_S (KW test, $\chi^2 = 4.5844$, $P = 0.205$).

Rates of molecular evolution based on between-species divergence show significant differences among pathways as well. In comparisons between Col-0 and *A. lyrata*, the ratios of nonsynonymous to synonymous substitution, K_a/K_s (KW Test, $\chi^2 = 18.8665$, df = 3, $P = 0.0003$) and K_a (KW test, $\chi^2 = 10.0371$, df = 3, $P = 0.018$) are significantly different among pathways, whereas K_s is not (KW test, $\chi^2 = 5.1113$, df = 3, $P = 0.16$, see table 1). Posthoc pairwise MWU tests between all four pathways indicated that significant differences in K_a/K_s are attributable to lower rates of protein evolution in the photoperiod ($P = 0.00039$) and gibberellic acid ($P = 6.4 \times 10^{-5}$) pathways compared with the vernalization genes. The comparison between vernalization and integrator pathways ($P = 0.0096$) is marginally insignificant after Bonferroni correction. Therefore, proteins in the vernalization pathway (average $K_a/K_s = 0.28$) appear to be evolving faster than photoperiod (average $K_a/K_s = 0.15$) and gibberellic acid (average $K_a/K_s = 0.12$) genes, and possibly faster than integrator genes (average $K_a/K_s = 0.20$) as well (see table 1). A possible explanation for this pattern is that genes in different pathways experience different levels of pleiotropic constraint. Many genes in the gibberellic acid and photoperiod pathways (such as the photoreceptor loci) have pleiotropic effects during development (Schepens et al. 2004; Schwechheimer 2008), whereas the vernalization genes appear to function primarily to control flowering time response to overwintering seasonal cues (but see Chiang et al. 2009).

The Evolution of the *A. thaliana* Flowering Time Gene Network

Comparative sequence data in genomewide studies have led to systematic analyses of evolutionary patterns among genes and their relation to various features of organismal genomes (Wolfe and Li 2003; Cork and Purugganan 2004). Over the last few years, analysis of genome-scale data resulted in several novel observations, providing insights into the roles of genomic architecture and gene function on molecular evolution and suggesting that the processes that govern gene diversification may be fairly complex (Williams and Hurst 2000; Pál et al. 2001; Urrutia and Hurst 2003). Several of these studies have focused on examining the molecular evolution of gene networks that direct particular metabolic, physiological or developmental traits (Jordan et al. 2003; Qin et al. 2003; Makino et al. 2006; Flowers et al. 2007; Kim et al. 2007; Greenberg et al. 2008; Presser et al. 2008). The increasing numbers of well-characterized molecular genetic networks provide opportunities to dissect patterns of gene evolution within a network context and determine how natural selection operates within functional genetic systems.

Despite its clear ecological significance (Engelmann and Purugganan 2006), there is limited population genetic evidence of widespread adaptive evolution across the flowering time genetic network. Rather, evidence of purifying selection is found throughout the network in both coding and noncoding DNA suggesting that a significant amount of the standing variation in this system of genes is deleterious (Weinreich and Rand 2000; Bustamante et al. 2002; Wright and Andolfatto 2008). In principle, multi-niche (Levene 1953) or other models of balancing selection could contribute to the excess of polymorphism observed in MK tests (but see Gillespie 1999). However, although different adaptive models including spatially varying or fluctuating selection have been suggested as mechanisms that may account for excesses of replacement polymorphism in MK tests at single loci (e.g., Verrelli and Eanes 2000, 2001), accounting for the large excesses observed across many background fragments and flowering time genes seems likely to be at odds with a primary role for adaptive processes in generating this pattern.

The absence of evidence for adaptive evolution in the flowering time gene network contrasts with patterns of polymorphism and divergence for a set of disease resistance loci, of which half of the *R* genes in Arabidopsis appear to be under diversifying selection (Bakker et al. 2006). This may indicate either that individual flowering time genes 1) do not have a history of recurrent selection, or 2) selection may not act in a specieswide mechanism, but may be more limited to local populations or geographically distinct regions of the species range. This latter hypothesis has been offered as an explanation for the low frequency of many

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

polymorphisms that contribute to natural flowering time variation in *A. thaliana* (Alonso-Blanco et al. 2005, e.g., *FLM*, Werner et al. 2005; *FRL1*, Schläppi 2006; *CRY2*, El-Assal et al. 2001; *HUA2*, Doyle et al. 2005) and could contribute to patterns of variation observed at the *FRI* locus (Le Corre 2005; Toomajian et al. 2006). However, evidence of a preponderance of deleterious polymorphisms in our population genetic analysis suggests that the low frequency of functional alleles may be explained by purifying selection. If true, this would indicate that a fraction of the standing genetic variation for flowering time may be maladaptive and maintained by mutation-selection balance, as observed in other quantitative traits (Keightley and Hill 1990; Barton and Keightley 2002; Zhang and Hill 2005). Slightly deleterious polymorphisms may therefore dominate the mutational landscape of this ecologically important genetic network and deleterious mutations may play a large role in shaping the genetic architecture of this life-history transition.

Supplementary Material

Supplementary figure S1 and supplementary tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Stephen I. Wright who kindly provided alignments of background fragments. We would also like to thank three anonymous reviewers for helpful comments that improved the manuscript and the Purugganan laboratory for reading an initial draft. This work was supported by an NSF FIBR grant to M.D.P. (co-PI).

Literature Cited

Alonso-Blanco C, Mendez-Vigo B, Koornneef M. 2005. From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. Int J Dev Biol. 49: 717–732.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215: 403–410.

Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T, Amasino RM, Sharrock RA. 1997. A deletion in the *PHYD* gene of the *Arabidopsis Wassilewskija* ecotype defines a role for phytochrome D in red/far-red light sensing. Plant Cell. 9:1317–1326.

Bakker EG, Toomajian C, Kreitman M, Bergelson J. 2006. A genome-wide survey of *R* gene polymorphisms in *Arabidopsis*. Plant Cell. 18:1803–1818.

Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Chory J, Weigel D. 2006. The *PHYTOCHROME C* photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. Nat Genet. 38:711–715.

Barton NH, Keightley P. 2002. Understanding quantitative genetic variation. Nat Rev Genet. 3:11–21.

Bäurle I, Dean C. 2006. The timing of developmental transitions in plants. Cell. 125:655–664.

Bustamante CD, Fledel-Alon A, Williamson S, et al. (11 co-authors). 2005. Natural selection on protein-coding genes in the human genome. Nature. 437:1153–1157.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. 2002. The cost of inbreeding in *Arabidopsis*. Nature. 416:531–534.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci USA. 101:15670–15675.

Charlesworth J, Eyre-Walker A. 2008. The McDonald–Kreitman test and slightly deleterious mutations. Mol Biol Evol. 25: 1007–1015.

Chiang GCK, Deepak B, Kramer EM, Amasino RM, Donohue K. 2009. Major flowering time gene, *FLOWERING LOCUS C*, regulates seed germination in *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 106:11661–11666.

Clarke JH, Mithen R, Brown JKM, Dean C. 1995. QTL analysis of flowering time in *Arabidopsis thaliana*. Mol Gen Genet. 248:278–286.

Coberly C, Rausher MD. 2008. Pleiotropic effects of an allele producing white flowers in *Ipomoea purpurea*. Evolution. 62: 1076–1085.

Comeron JM. 1995. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol. 41:1152–1159.

Cork J, Purugganan M. 2004. The evolution of molecular genetic pathways and networks. Bioessays. 26:479–484.

DeRose-Wilson LJ, Gaut BS. 2007. Transcription-related mutations and GC content drive variation in nucleotide substitution rates across the genomes of *Arabidopsis thaliana* and *Arabidopsis lyrata*. BMC Evol Biol. 7:66.

Doyle MR, Bizzell CM, Keller MR, Michaels SD, Song JD, Noh YS, Amasino R. 2005. *HUA2* is required for the expression of floral repressors in *Arabidopsis thaliana*. Plant J. 41:376–385.

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

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD. Forthcoming. Candidate gene association mapping of *Arabidopsis* flowering time. Genetics.

Ehrenreich IM, Purugganan MD. 2006. The molecular genetic basis of plant adaptation. Am J Bot. 93:953–962.

El-Assal SE-D, Alonso-Blanco C, Peeters AJM, Raz V, Koornneef M. 2001. A QTL for flowering time in *Arabidopis* reveals a novel allele of *CRY2*. Nat Genet. 29:435–440.

Engelmann K, Purugganan MD. 2006. The molecular evolutionary ecology of plant development: flowering time in *Arabidopsis thaliana*. Adv Bot Res. 44:505–524.

Fay JC, Wyckoff GJ, Wu C-I. 2001. Positive and negative selection on the human genome. Genetics. 158:1227–1234.

Fay JC, Wyckoff GJ, Wu C-I. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature. 415:1024–1026.

Filiault DL, Wessinger CA, Dinneny JR, Lutes J, Borevitz JO, Weigel D, Chory J, Maloof JN. 2008. Amino acid polymorphisms in *Arabidopsis phytochrome B* cause differential response to light. Proc Natl Acad Sci USA. 105:3157–3162.

Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. Mol Biol Evol. 24: 1347–1354.

Foxe JP, Dar VUN, Zheng H, Nordborg M, Gaut BS, Wright SI. 2008. Selection on amino acid substitutions in Arabidopsis. Mol Biol Evol. 25:1375–1383.

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

Gillespie JH. 1999. The role of population size in molecular evolution. Theor Pop Biol. 55:145–156.
Greenberg AJ, Stockwell SR, Clark AG. 2008. Evolutionary constraint and adaptation in the metabolic network of *Drosophila*. Mol Biol Evol. 25:2537–2546.
Huang XQ, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.
Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. 2000. Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science. 290:344–347.
Jordan IK, Wolf YI, Koonin EV. 2003. No simple dependence between protein evolution rate and the number of protein–protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol. 3:Art. No. 1
Keightley P, Hill WG. 1990. Variation maintained in quantitative traits with mutation selection balance—pleiotropic side effects on fitness traits. Proc R Soc Lond B. 242:95– 100.
Kim PM, Korbel JO, Gerstein MB. 2007. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. Proc Natl Acad Sci USA. 104:20274–20279.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
Komeda Y. 2004. Genetic regulation of time to flower in *Arabidopsis thaliana*. Ann Rev Plant Biol. 55:521.
Koornneef M, Alonso-Blanco C, Vreugdenhil D. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. Ann Rev Plant Biol. 55:141–172.
Kuittinen H, Sillanpaa MJ, Savolainen O. 1997. Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. Theor Appl Genet. 95:573–583.
Le Corre V. 2005. Variation at two flowering time genes within and among populations of Arabidopsis thaliana: comparison with markers and traits. Mol Ecol. 14:4181–4192.
Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Mol Biol Evol. 19:1261–1271.
Levene H. 1953. Genetic equilibrium when more than one ecological niche is available. Am Nat. 87:331–333.
Makino T, Suzuki Y, Gojobori T. 2006. Differential evolutionary rates of duplicated genes in protein interaction network. Gene. 385:57–63.
Mathews S, McBreen K. 2008. Phylogenetic relationships of B-related phytochromes in the Brassicaceae: redundancy and the persistence of phytochrome D. Mol Phyl Evol. 49: 411–423.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in Drosophila. Nature. 351:652–654.
Mitchell-Olds T, Schmitt J. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. Nature. 441:947–952.
Mouradov A, Cremer F, Coupland G. 2002. Control of flowering time: interacting pathways as a basis for diversity. Plant Cell. 14:S111–S30.
Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
Nordborg M, Hu TT, Ishino Y, et al. (23 co-authors). 2005. The pattern of polymorphism in Arabidopsis thaliana. PLOS Biol. 3:1289–1298.
Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD. 2004. Linkage disquilibrium mapping of Arabidopsis CRY2 flowering time alleles. Genetics. 167:1361–1369.
Olsen KM, Womack A, Garrett AR, Suddith JI, Purugganan MD. 2002. Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics. 160: 1641–1650.
Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. Genetics. 158:927–931.
Presser A, Elowitz MB, Kellis M, Kishony R. 2008. The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication. Proc Natl Acad Sci USA. 105:950–954.
Qin H, Lu HHS, Wu WB, Li WH. 2003. Evolution of the yeast protein interaction network. Proc Natl Acad Sci USA. 100: 12820–12824.
Rausher MD, Lu Y, Meyer K. 2008. Variation in constraint versus positive selection as a explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol. 67:137–144.
Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. Genetics. 132:1161–1176.
Scarcelli N, Cheverud JM, Schaal BA, Kover PX. 2007. Antagonistic pleiotropic effects reduce the potential adaptive value of the *FRIGIDA* locus. Proc Natl Acad Sci USA. 104:16986–16991.
Schepens I, Duek P, Fankhauser C. 2004. Phytochrome-mediated light signalling in *Arabidopsis*. Curr Op Plant Biol. 7:564–569.
Schläppi MR. 2006. *FRIGIDA LIKE 2* is a functional allele in Landsberg erecta and compensates for a nonsense allele of *FRIGIDA LIKE 1*. Genetics, genomics, and molecular evolution. Plant Physiol. 142:1728–1738.
Schlenke TA, Begun DJ. 2003. Natural selection drives *Drosophila* immune system evolution. Genetics. 164: 1471–1480.
Schwechheimer C. 2008. Understanding gibberellic acid signaling—are we there yet? Curr Op Plant Biol. 11:9– 15.
Shapiro JA, Huang W, Zhang CH, et al. (11 co-authors). 2007. Adaptive genic evolution in the *Drosophila* genomes. Proc Natl Acad Sci USA. 104:2271–2276.
Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. Plant Physiol. 138:1163–1173.
Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. Proc Natl Acad Sci USA. 101:4712–4717.
Subramaniam B, Rausher MD. 2000. Balancing selection on a floral polymorphism. Evolution. 54:691–695.
Swarbreck D, Wilks C, Lamesch P, et al. (16 co-authors). 2008. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. 36: D1009–D1014.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.
Thornton K. 2003. libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics. 19:2325–2327.
Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang C, Zheng H, Zhao K, Calabrese P, Dean C, Nordborg M. 2006. A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. PLOS Biol. E137:0732–0738.
Ungerer MC, Halldorsdottir AA, Modliszewski JL, Mackay TFC, Purugganan MD. 2002. Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. Genetics. 160: 1133–1151.
Ungerer MC, Halldorsdottir SS, Purugganan MD, Mackay TFC. 2003. Genotype-environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. Genetics. 165:353–365.
Urrutia AO, Hurst LD. 2003. The signature of selection mediated by expression on human genes. Genome Res. 13:2260–2264.

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021

Verrelli BC, Eanes WF. 2000. Extensive amino acid polymorphism at the Pgm locus is consistent with adaptive protein evolution in *Drosophila melanogaster*. Genetics. 156: 1737–1752.

Verrelli BC, Eanes WF. 2001. Clinal variation for amino acid polmorphisms at the Pgm locus in *Drosophila melanogaster*. Genetics. 157:1649–1663.

Wang Q, Sajja U, Rosloski S, Humphrey T, Kim MC, Bomblies K, Weigel D, Grbic V. 2007. *HUA2* caused natural variation in shoot morphology of *A. thaliana*. Curr Biol. 17: 1513–1519.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Pop Biol. 7:256–276.

Weinig C, Ungerer MC, Dorn LA, Kane NC, Toyonaga Y, Halldorsdottir SS, Mackay TFC, Purugganan MD, Schmitt J. 2002. Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. Genetics. 162: 1875–1884.

Weinreich DM, Rand DM. 2000. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics. 156:385–399.

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D. 2005. Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci USA. 102:2460–2465.

Williams EJ, Hurst LD. 2000. The proteins of linked genes evolve at similar rates. Nature. 407:900–903.

Wolfe KH, Li WH. 2003. Molecular evolution meets the genomics revolution. Nat Genet. 33:255–265.

Wright SI, Andolfatto P. 2008. Selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. Ann Rev Ecol Evol Syst. 39:193–213.

Zhang XS, Hill WG. 2005. Genetic variability under mutation selection balance. Trends Ecol Evol. 20:468–470.

Zhao KY, Aranzana MJ, Kim S, et al. (11 co-authors). 2007. An Arabidopsis example of association mapping in structured samples. PLOS Genet. 3:e4.

Patricia Beldade, Associate Editor

Accepted July 13, 2009

Downloaded from https://academic.oup.com/mbe/article/26/11/2475/1012321 by guest on 30 July 2021