Natural Variation in *Arabidopsis lyrata* Vernalization Requirement Conferred by a *FRIGIDA* Indel Polymorphism

Helmi Kuittinen, Anne Niittyvuopio, Paula Rinne, and Outi Savolainen

Department of Biology, University of Oulu, Oulu, Finland

Species share homologous genes to a large extent, but it is not yet known to what degree the same loci have been targets for natural selection in different species. Natural variation in flowering time is determined to a large degree by 2 genes, *FLOWERING LOCUS C* and *FRIGIDA*, in *Arabidopsis thaliana*. Here, we examine whether *FRIGIDA* has a role in differences in flowering time between and within natural populations of *Arabidopsis lyrata*, a close outcrossing perennial relative of *A. thaliana*. We found 2 *FRIGIDA* sequence variants producing potentially functional proteins but with a length difference of 14 amino acids. These variants conferred a 15-day difference in flowering time in an association experiment in 2 Scandinavian populations. The difference in flowering time between alleles was confirmed with transformation to *A. thaliana*. Because the north European late-flowering populations harbor both late- and early sequence variants at intermediate frequencies and the late-flowering variant is most frequent in the southern early flowering European population, other genetic factors must be responsible for the flowering time differences between the populations. The length polymorphism occurs at high frequencies also in several North American populations. The occurrence of functional variants at intermediate frequencies in several populations suggests that the variation may be maintained by balancing selection. This is in contrast to *A. thaliana*, where independent loss-of-function mutations at the *FRIGIDA* gene are responsible for differences between populations and local adaptation.

Introduction

Molecular population genetics now allows examining the nature of genetic variation even in complex adaptations (review by Hoekstra and Coyne 2007). It is possible to resolve whether the phenotypic variation is due to structural or regulatory changes (Chen et al. 2007). Further, we can compare the genetic architecture and the molecular genetic details of similar adaptations in different species, currently especially in model organisms. This also improves our understanding of the evolutionary processes governing the natural variation (Mitchell-Olds et al. 2007). Adaptation to similar environmental pressures in different species or populations can occur by mutations in either the same genes (parallel evolution) or the different genes (convergent evolution). Both parallel and convergent evolution have been documented in animals and humans (e.g., Hoekstra and Nachman 2003; Colosimo et al. 2005; Tishkoff et al. 2007), but in plants only few studies exist (Kivimäki et al. 2007). If a specific phenotype could be reached in many different ways, directional selection on alternative genes could give rise to the same adaptations in different species, leading to convergent evolution (Nachman 2005). Genes and genetic pathways are redundant to a high degree in plants (Lin et al. 1999), and this type of evolution could be expected to be abundant. However, genes that have few developmental constraints may be more likely to be targets for positive selection than genes having high constraints. In this case, the same gene would be likely to be used in different species for the same adaptations and parallel evolution would be more likely.

The timing growth and reproduction are critical to the fitness of plants and animals. These traits mostly display complex quantitative variation (Mitchell-Olds et al. 2007). Initiation of reproduction is an important transition in a plant’s life, often related to life history trade-offs (Silvertown and Doust 1993). Early flowering is often associated with drought escape or shortness of growing season. Fitness consequences of timing of flowering have been demonstrated in both annual (Franke et al. 2006; Sherrard and Maherali 2006) and perennial plants (Hall and Willis 2006; Giménez-Benavides et al. 2007; Sandring et al. 2007).

In many plants, a long period of cold temperature (vernalization) is required for flowering. The cold requirement in biennial or perennial plants postpones the first flowering to the second growing season. Vernalization requirement versus no requirement segregates in natural populations of many plant species in taxonomically diverse groups (Michaels and Amasino 2000). For example, winter-annual *Arabidopsis thaliana* accessions flower only after exposure to winter, but summer-annual accessions flower already in their first season (Napp-Zinn 1969). Both types flower in constant growth chamber conditions without vernalization but the winter accessions much later than the summer accessions (Karlsson et al. 1993; Nordborg and Bergelson 1999). Similar summer and winter types are found, for example, in different accessions of wheat and barley, wild beet, and *Brassica*.

Signals from the environment are mediated by a number of genetic pathways converging at integrator genes to initiate flowering (Simpson and Dean 2002). Different genes seem to govern natural variation in vernalization requirement, even in closely related plant taxa. The major players in *A. thaliana* are flowering repressor *FLOWERING LOCUS C* (*FLC*) and its positive regulator *FRIGIDA* (*FRI*) (Johanson et al. 2000; Gazzani et al. 2003; Lempe et al. 2005; Shindo et al. 2005). In *Brassica*, *FLC* is an important determinant of natural variation of vernalization requirement (e.g., Osborn et al. 1997). However, in beet (*Beta vulgaris*) the bolting gene responsible for natural variation in flowering habit is not *FLC*, although *FLC* retains a conserved function in the vernalization pathway (Reeves et al. 2007). In a more distant species, wheat, natural variation for vernalization requirement is regulated by the *Arabidopsis* *AP1* homolog and a *CO*-like gene (Yan et al. 2007).
2004), the homologues of which are not known to have a vernalization-related function in *A. thaliana*.

Thus, FRI is the most important regulator of natural flowering time variation in *A. thaliana*, although it may have modest role in other species. A large share of variation in flowering time among European natural accessions of *A. thaliana* is due to independent loss-of-function mutations at FRI, and this variation is believed to lead to adaptation to local conditions (Le Corre et al. 2002). Here, we study the role of FRI in flowering time variation in a close relative, *Arabidopsis lyrata*, and compare it with that in *A. thaliana*.

*Arabidopsis lyrata* is becoming a promising model species for evolutionary and ecological studies, facilitated with the whole-genome sequencing being underway. Since their divergence 5 MYA (Koch et al. 2000), *A. thaliana* and *A. lyrata* have acquired different life history strategies. Whereas *A. thaliana* is a weedy selfer with a worldwide distribution, *A. lyrata* is a perennial outcrosser that occurs in small disjunct populations in the northern hemisphere in ecologically variable habitats (Mitchell-Olds 2001). Perenniarity and outcrossing mode of reproduction have given rise to stable, geographically differentiated populations with a great opportunity to study the genetic basis of local adaptation in the framework of classical population genetics (Riihimäki et al. 2005). Populations of *A. lyrata* differ in their flowering times, the southern populations flowering earlier than the northern populations in growth chamber conditions (Riihimäki and Savolainen 2004). Variation in the time of initiation of flowering has been shown to have fitness consequences in natural populations of *A. lyrata*, selection differing among years and sites (Sandring et al. 2007).

In this study, we show evidence that an insertion/deletion (indel) polymorphism in the FRI gene influences flowering time in *A. lyrata* using sequence analysis, association studies, and transformation experiments. This evidence together with worldwide patterns of variation suggests that the functional polymorphism may be maintained by balancing selection.

**Materials and Methods**

**Population Flowering Experiment**

Flowering time variation was characterized in a growth chamber experiment using 3 *A. lyrata* populations, 1 from Central Europe (Plech, Germany) and 2 from Northern Europe (Mjällom, Sweden and Spitzenstulen, Norway). The plants were exposed to 2 treatments, vernalization and control (no vernalization) treatments. Plants were grown from seeds originating from lab-generated crosses. These crosses were made between randomly chosen parents grown from seeds collected randomly from each wild population. The sample size was 80 plants per population, originating from 10 to 20 families per population. Seeds were germinated at 20 °C in 8-h photoperiod in petri dishes and then planted in 7 × 7 cm pots with 1:1 peat and gravel. The plants were grown in a growth chamber at 22 °C in 20-h photoperiod.

The light was produced by Power Star HQI bulbs. Half of the plants were transferred to the cold room (4 °C, 8-h photoperiod) when the rosettes were 4 weeks old, and they were returned back to the growth chamber after 9 weeks of vernalization treatment. Half of the plants remained as nonvernalized controls. The nonvernalized plants were germinated later so that both treatments grew together after return from the vernalization to the growth chamber. Plants from the different treatments and populations were randomized within blocks. Flowering was scored as the number of days from planting until the visible flower bud was seen, excluding the time spent in cold room in the vernalized treatment. Because normality was not obtained with transformations, we used the original data. The block effect was not significant (Kruskal–Wallis test, *P* = 0.904), and we analyzed the data as a completely randomized design. The significance of the difference in flowering between the populations in each treatment was tested with nonparametric Kruskal–Wallis tests.

**Sequencing**

Sequence variation was studied in 9 populations from Europe and North America from a south–north gradient (fig. 1), obtaining altogether 295 alleles. Primers to amplify the FRI gene region were designed using information from the *A. thaliana* genomic sequence for initial polymerase chain reaction (PCR). The region was amplified in 3 overlapping fragments using primer pairs 5′-AACAGGACGAGTGGAAATTAGGG-3′ and 5′-ACAAATCCATACAAACTCTGC-3′, 5′-CTTCGGCAATCTCATGTCTC-3′ and 5′-GTAACTACTCACAGAGGCTGT-3′, and 5′-TCA-TGATTCTCTATCTACC-3′ and 5′-ACGAGAAC-AAGGAAGGGAACAGAG-3′. The whole coding sequence and introns (ca. 2,600 bp), and stretches from the promoter (22 bp) and 3′ flanking (213 bp) region were obtained for 133.5 individuals. For 14 individuals, the almost complete coding region (lacking 66 and 54 bp from the 5′ and 3′ ends) was obtained in 1 PCR, using either 5′-GGGGAGCCATTCTAACC-3′ or 5′-CGATTTCACTGCTGGACAG-3′ as the forward primer and either 5′-GGGGAGGACCTGTATGGG-3′ or 5′-GGGGTGCTATGATGACT-3′ as the reverse primer. The temperature profile was 3:00 min at 94 °C; 35 cycles of 1:00 min at 94 °C, 1:00 min at 50 °C, 2:00 min at 72 °C; 6:00 min at 72 °C in a Roebocycler PCR machine. The reactions contained 0.05 U/μl Promega Taq polymerase, 200 μM deoxynucleoside triphosphate, 1.5 mM MgCl₂, 0.25 μM each primer, 50 mM Tris–HCl (pH 9.0), and 0.1% Triton X-100 in a 25-μl volume. Both strands were directly sequenced using the PCR product as a template. In the case of an indel heterozygote, primers were designed after the indel region. Sequencing reactions were run in the ABI PRISM 377 or 3730 sequencing machines, and contigs were aligned using the Sequencher program. All chromatograms were visually inspected, and special attention was paid so that genotypes at polymorphic sites were correctly scored. Haplotypes of multiply heterozygous individuals were determined with the PHASE program (Stephens et al. 2001) (267 haplotypes) and verified by sequencing progenies of the individuals. Twenty-eight haplotypes were determined with the algorithm of Clark (1990) and with the PHASE program, and the 1 singleton mutation was assigned to the haplotype with allel specific amplification and sequencing. The sequences are
deposited in the EMBL database under accession numbers EU177129–EU177423.

Sequence Analysis

The level of nucleotide diversity (\(\pi\)) was estimated in the total data, within populations and within groups of short and long indel alleles. The divergence between allele groups was estimated with net nucleotide distance (\(D_n\)). Levels of linkage disequilibrium (LD) were estimated using Kelly’s (1997) ZnS measure. To test the fit of the data with the neutral Wright–Fisher population model, Tajima’s (1989) \(D\) and \(H\) statistics of Fay and Wu (2000) were calculated and their significance was tested with coalescent simulations. The congruence of the ratio of synonymous to nonsynonymous polymorphisms with the ratio of divergence at synonymous and nonsynonymous sites was tested with the test of McDonald and Kreitman (1991). The congruence of the level of synonymous polymorphism with the divergence along the gene was tested with the tests of Hudson, Kreitman, and Aguadé (HKA; Hudson et al. 1987) using a sliding window analysis (DNA slider program, McDonald 1998). Arabidopsis thaliana FRI sequence (GenBank accession number NC_003075) was used as the outgroup sequence. The level of polymorphism was higher in the first part of the gene (exon 1) than in the second part of the gene (exon 2 and 3), and the levels of polymorphism and divergence were compared separately in these 2 parts with those in the ADH region (Savolainen et al. 2000 and O. Savolainen, unpublished data). The ADH sample contained 5 Plech, 4 Mjällom, 4 Mayodan, 4 Karhumäki, and 2 Iceland alleles, and the FRI sample had individuals from the same populations in corresponding ratios. The overall and pairwise genetic differentiation between populations was measured as \(F_{ST}\) (Hudson et al. 1992). A population tree was constructed with the Neighbor-Joining method using Nei’s (1987) genetic distance and treating each site as separate character. The confidence was assessed bootstrapping the data over nucleotide sites. The phylogenetic analysis was made with the PHYLIP program (Felsenstein 2005). All sequence analyses were made with the DnaSP 4.00 program (Rozas et al. 2003), unless stated otherwise.

Expression Analysis

A 14-amino acid indel polymorphism was detected among the FRI sequences. We checked with a reverse transcription (RT) assay whether both the long and the short alleles were expressed. Total genomic RNA was isolated from fresh leaves of 2 adult plants grown in 20:4 photoperiod, one of which was homozygous for the long allele and the other homozygous for the short allele. cDNA was synthesized with the RT enzyme (Invitrogen, Carlsbad, CA) using a polyT primer. The cDNA was used as a template in a PCR reaction with forward primer 5’-CTACGTATTCTCTCATTCTACC-3’ and reverse primer 5’-GGGTCTAATGATGACTACT-3’. The amplified region contained the 42-bp indel polymorphism. PCR products were run on agarose gel to check their length. The fragments were subsequently sequenced to verify that the sequence corresponded to the expected FRI genotype.

Transformation

To test whether A. lyrata FRI long and short alleles delay flowering in A. thaliana and if there is difference in flowering time between the alleles, the early flowering wild-type A. thaliana (Col-4) plants were transformed with A. lyrata 35S::FRI constructs. The Col-4 ecotype has functional FLC and nonfunctional fri alleles. 35S::FRI(SP32) and 35S::FRI(SP33) were constructed by cloning 2.3-kb fragments containing FRI-coding region from Spiterstulen plants that were homozygous for the short allele (SP32) and for the long allele (SP33) into pBI121 binary vector between XbaI and SmaI sites adjacent to the CaMV 35S promoter. In addition to the indel in the third exon, the SP32 allele differed from SP33 by having 139Ala, 307Met, 328Ser, 343Lys, and 347Ser.
and 437Lys, whereas the SP33 allele had Thr, Ile, Cys, and Arg in corresponding positions (fig. 2). Col-4 plants were transformed with these constructs by floral dipping (Clough and Bent 1998) using Agrobacterium strain C58C1 pGV3850. Transgenic T1 plants were selected on 1/2 Murashige–Skoog plates (kanamycin 50 µg/ml) and 26 independent transformants (T1) were obtained from 35S::FRI(SP32) and 15 from 35S::FRI(SP33). Segregation analysis was done to select for lines with a single copy of the foreign gene and to identify homozygous lines. Two independent homozygous transformants for each allele (SP32/32 and SP32/45 for the short allele and SP33/9 and SP33/17 for the long allele) were selected in the T3 generation. Wild Col-4 plants and Col-FRI plants (Col ecotype with A. thaliana FRI from late-flowering St-2 ecotype; Lee and Amasino 1995) were used as controls. The phenotype was scored in 2 environmental conditions. The imbibed seeds were exposed to 4 °C during 30 days (vernalization treatment) or 3 days (no vernalization). After 30 or 3 days cold treatment, the seeds were planted on peat–vermiculite mixture (1:1) and placed in a growth chamber in 18-h photoperiod. Flowering time was measured as the number of days from germination to bolting, in a sample ranging between 19 and 24 for each line × treatment combination. The difference in flowering time between transformed lines was tested with nonparametric Kruskal–Wallis tests.

Associations of Indel Genotypes with Flowering Time

The FRI short/long indel polymorphism was scored in Spiterstulen and Mjällom individuals that had been growing in the population flowering experiment. A fragment transversing the indel was amplified using the forward primer 5′-TCATCGTATTCTCCTATCTACC-3′ and the reverse primer 5′-ACCGAGAAACAAGGAACAGAG-3′ in standard PCR conditions. The PCR products were run on agarose gel and the 2 homozygotes and the heterozygote were genotyped. The association of flowering time with the genotype was tested with nonparametric Kruskal–Wallis tests separately in the 2 populations and in the 2 conditions (vernalized and control).

Worldwide FRI-Indel Genotyping

To obtain a comprehensive picture of the extent of the 42-bp indel polymorphism among natural populations, the polymorphism was genotyped in altogether 20 populations (sample sizes are given in fig. 1). The long/short polymorphism was scored as described above for the association study. The fit of the genotype frequencies with those expected in panmictic populations (Hardy–Weinberg law) was tested with chi-square tests in each population.

Results

Flowering Experiments

When not vernalized, the mean time till bolting in the southern Plech population was 45 days and in the northern populations Mjällom and Spiterstulen 83 and 74 days. The difference between the populations was significant (P < 0.001). After vernalization, the differences in bolting times were smaller but populations were still significantly different (P < 0.001). Mjällom flowered on average at 37, Spiterstulen at 40, and Plech at 32 days (fig. 3). Thus, the Scandinavian populations responded more strongly to vernalization, leading to a considerably reduced difference between populations after vernalization. This situation is similar to A. thaliana where late-flowering, winter-annual ecotypes respond to vernalization more strongly than early flowering, summer-annual ecotypes, resulting in shorter flowering times in all ecotypes (Karlsson et al. 1993).

Sequence Variation

We obtained sequence data from 9 populations (fig. 1 and table 1). Most sequences had an open reading frame coding for a full-length protein (fig. 2), that is, they did not have mutations that would lead to a major defect in the protein (premature STOP codon, out-of-frame indel, or change in the START codon). There were only 3 such defective mutations among the 295 sequences. However, there was abundant nonsynonymous and synonymous polymorphism within the populations. The first part of the gene (exon 1) harbored more variation than the second part (exons 2 and 3) of the gene. This was especially pronounced for the synonymous sites, but the level of nonsynonymous polymorphism also varied in the same direction (table 1 and fig. 4).

Within populations, there was generally a lower level of nonsynonymous than synonymous nucleotide diversity. The median $\pi_S/\pi_S$ ratio within populations was 0.39 and the mean 0.59 (influenced strongly by the high value in Mjällom). With the pooled data, the ratio was 0.25. The ratio of nonsynonymous to synonymous divergence ($K_S/K_S$) between A. lyrata and A. thaliana was 0.30. These results suggest that the gene is under functional constraint.

In spite of a high level of nonsynonymous polymorphisms, there was only 1 site where populations had been fixed for alternative amino acids (fig. 2). This was a nonradical change from glutamic acid to aspartic acid. Many silent polymorphisms were shared among populations too. The pairwise divergence between populations ($F_{ST}$) at silent sites ranged from 0.11 to 0.63 (see supplementary table S1, Supplementary Material online), and for all populations $F_{ST}$ was 0.41. The topology of the Neighbor-Joining tree of populations based on Nei’s genetic identity was similar to that based on microsatellites, although branch lengths

![Image](fig_2.png) Amino acid polymorphisms in Arabidopsis lyrata populations. The number of each haplotype is shown in the last column. There was 1 amino acid change that was fixed between at least 2 populations (blue). Major alterations of the protein (premature STOP codon, out-of-frame indel, and change in the START codon) are shown with red and the 14-amino acid deletion allele with yellow color. Polymorphic amino acids highlighted in gray correspond to predicted coiled coil domains (Johanson et al. 2000).
seemed to vary (Muller et al. 2007; see supplementary fig. S2, Supplementary Material online).

The most striking mutation was a 42-bp indel, corresponding to a 14-amino acid length difference at the protein level (fig. 2). Both alleles were transcribed, based on the RT assay. This polymorphism was found in several populations in Europe and North America. Additional genotyping revealed that the short allele was not present in Alaskan and Karhuma¨ki populations but occurred at low frequencies in southern populations in both continents and had highest frequency in Scandinavian and North Eastern American populations (fig. 1). Genotypes were in Hardy–Weinberg equilibrium in all populations.

Several tests were used to find signals of balancing or positive directional selection in the sequence data. In addition to using a pooled data set, we partitioned the data in individual populations. We also contrasted the long and short allele groups and the first and second half of the gene.

The ratio of number of synonymous polymorphisms to number of fixed differences between A. lyrata and A. thaliana was significantly higher in the first (exon 1) than in the second part (exons 2 and 3) of the gene using a sliding window analysis in the total data set (HKa test, table 1). When the ADH gene was used as a reference against which the first and second halves were tested, there was again an excess of synonymous polymorphism in the first half of the FRI gene (P = 0.013) but not in the second half (P = 0.3715). Out of the 5 individual populations that had enough polymorphism for the sliding window analysis, Plech and Bohemia had significant results and Ithaca and Karhuma¨ki marginally significant results. Thus, this tendency seems to exist not only in the pooled data but also in many populations.

The frequency distributions of nucleotide polymorphisms (Tajima’s D) at all sites were as expected in neutral situation in all populations, except for the Russian Karhuma¨ki and the Swedish Mjällom populations, that had a significant excess of intermediate frequency polymorphisms (positive Tajima’s D) (table 1). The Tajima’s D values for nonsynonymous sites were very similar with those using all sites.

The ratio of polymorphisms to divergence between A. lyrata and A. thaliana was similar at synonymous and nonsynonymous sites resulting in nonsignificant McDonald–Kreitman tests in all populations and in the pooled data. The test was also nonsignificant when applied separately to the first exon and to the second and third exons.

There were no consistent differences between the long and short allele groups within populations, in terms of nucleotide diversity, LD, Tajima’s D, or other neutrality tests (table 1). Similar patterns in sequence variation, combined with the information that the long and short alleles occur in both North America and Europe, suggest that the indel polymorphism is relatively old. The net nucleotide difference (Dn) between the long and short allele groups did not seem to vary much along the length of the gene in different populations, perhaps except for an increase at the end in some populations (see supplementary fig. S3, Supplementary Material online).

Transformation

We transformed the early flowering A. thaliana Col-fri line with both short and long indel alleles and studied their influence on flowering. Flowering times between all 4 transformed lines differed significantly (P < 0.001) from each other. Both short and long alleles delayed flowering complementing the loss-of-function FRI allele in A. thaliana Col-fri line (fig. 5). However, the long allele lines flowered on average at 45.5 days and the short allele lines on average at 36.3 days. Note that A. thaliana functional FRI sequence, conferring late flowering, is similar to the long allele in terms of the indel. Vernalization reduced flowering time in all lines.

Association Analysis

Time to flowering was significantly associated with the indel genotypes in nonvernalized plants in Spiterstulen (P < 0.001) and it was close to significant in Mjällom (P = 0.062). Long allele homozygotes flowered 15 days later than heterozygotes in both populations. When the plants were vernalized, all plants flowered early and the difference between genotypes was reduced (fig. 5). The difference in flowering times between indel genotypes was borderline significant in Spiterstulen (P = 0.052) and not significant in Mjällom (P = 0.788).

Discussion

We have shown here that different A. lyrata populations flower at different times in a common garden experiment in growth chamber. However, after 9 weeks vernalization at the rosette stage they flower equally early. Thus, populations have differences in vernalization...
requirement, resembling the situation between A. thaliana winter and summer annuals. This result forms a complement to an earlier study (Riihimäki and Savolainen 2004) that found differences among A. lyrata populations in response to seed vernalization.

The transformations presented here showed that coding FRI sequence from A. lyrata can complement the native A. thaliana FRI, suggesting that the FRI gene has a similar function (it confers a vernalization requirement) in both species. Note that the functional copies are in nonhomologous positions in A. lyrata and A. thaliana (Kuittinen et al. 2004). Knowing that the homologous genes have a similar function, it is interesting to compare the roles of FRI in natural flowering time variation in the 2 species.

In A. thaliana, variation in the coding region of FRI is responsible for much of the differences between populations in vernalization requirement (Le Corre et al. 2002; Gazzani et al. 2003; Shindo et al. 2005). Several independent loss-of-function mutations, caused by out-of-frame deletions and premature STOP codons in the first exon, give rise to local adaptation of natural populations in A. thaliana (Le Corre et al. 2002; Caicedo et al. 2004; Stinchcombe et al. 2004). Based on sequence variation, we found little evidence that coding single nucleotide polymorphisms in

<p>| Table 1 |
| Analysis of FRI Sequences |</p>
<table>
<thead>
<tr>
<th>Sites</th>
<th>S</th>
<th>$\pi$(tot)</th>
<th>$\pi$(syn)</th>
<th>$\pi$(nonsyn)</th>
<th>$\pi$/(\pi_c)</th>
<th>ZnS</th>
<th>Tajima’s D</th>
<th>HKA (Sliding Window)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reykjavik (n = 28)</td>
<td>2,517</td>
<td>24</td>
<td>2.91</td>
<td>4.74</td>
<td>1.62</td>
<td>0.34</td>
<td>0.261 NS</td>
<td>0.679 NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>11</td>
<td>3.28</td>
<td>8.64</td>
<td>1.63</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>822</td>
<td>4</td>
<td>1.36</td>
<td>0.37</td>
<td>1.66</td>
<td>4.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short allele (n = 8)</td>
<td>2,529</td>
<td>6</td>
<td>0.59</td>
<td>0.59</td>
<td>0.18</td>
<td>0.31</td>
<td>—*</td>
<td>$-1.639$ NS</td>
</tr>
<tr>
<td>Long allele (n = 20)</td>
<td>2,559</td>
<td>24</td>
<td>3.37</td>
<td>4.34</td>
<td>2.3</td>
<td>0.53</td>
<td>0.340 NS</td>
<td>1.056 NS</td>
</tr>
<tr>
<td>Lom (n = 26)</td>
<td>2,380</td>
<td>22</td>
<td>3.08</td>
<td>4.08</td>
<td>2.12</td>
<td>0.52</td>
<td>0.249 NS</td>
<td>0.978 NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>10</td>
<td>3.83</td>
<td>7.43</td>
<td>2.72</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>834</td>
<td>4</td>
<td>1.25</td>
<td>0.39</td>
<td>1.51</td>
<td>3.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short allele (n = 5)</td>
<td>2,531</td>
<td>6</td>
<td>0.95</td>
<td>1.89</td>
<td>1.16</td>
<td>0.61</td>
<td>—*</td>
<td>$-1.455$ NS</td>
</tr>
<tr>
<td>Long allele (n = 21)</td>
<td>2,422</td>
<td>18</td>
<td>2.46</td>
<td>4.2</td>
<td>1.56</td>
<td>0.37</td>
<td>0.428 NS</td>
<td>0.706 NS</td>
</tr>
<tr>
<td>Spiterstulen (n = 38)</td>
<td>2,529</td>
<td>22</td>
<td>2.59</td>
<td>2.3</td>
<td>1.6</td>
<td>0.69</td>
<td>0.347 NS</td>
<td>0.852 NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>8</td>
<td>2.68</td>
<td>3.91</td>
<td>2.3</td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>834</td>
<td>4</td>
<td>0.78</td>
<td>0.54</td>
<td>0.85</td>
<td>1.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short allele (n = 13)</td>
<td>2,531</td>
<td>8</td>
<td>1.04</td>
<td>2.19</td>
<td>1.04</td>
<td>0.48</td>
<td>0.549 NS</td>
<td>0.095 NS</td>
</tr>
<tr>
<td>Long allele (n = 23)</td>
<td>2,572</td>
<td>16</td>
<td>0.79</td>
<td>0.6</td>
<td>0.37</td>
<td>0.61</td>
<td>0.752 NS</td>
<td>$-1.916$*</td>
</tr>
<tr>
<td>Mjällom (n = 30)</td>
<td>2,212</td>
<td>11</td>
<td>2.11</td>
<td>1.01</td>
<td>1.74</td>
<td>1.73</td>
<td>0.384 NS</td>
<td>2.201*</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>5</td>
<td>2.49</td>
<td>1.88</td>
<td>2.67</td>
<td>1.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>792</td>
<td>1</td>
<td>0.51</td>
<td>0</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short allele (n = 22)</td>
<td>2,231</td>
<td>7</td>
<td>1.2</td>
<td>0.23</td>
<td>1.25</td>
<td>5.54</td>
<td>0.569 NS</td>
<td>1.244 NS</td>
</tr>
<tr>
<td>Long allele (n = 8)</td>
<td>2,277</td>
<td>2</td>
<td>0.22</td>
<td>0.59</td>
<td>0.18</td>
<td>0.31</td>
<td>—*</td>
<td>$-1.310$ NS</td>
</tr>
<tr>
<td>Karhumäki (n = 32)</td>
<td>2,571</td>
<td>18</td>
<td>3.02</td>
<td>9.96</td>
<td>1.16</td>
<td>0.12</td>
<td>0.473 NS</td>
<td>2.500*</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>11</td>
<td>4.98</td>
<td>16.06</td>
<td>1.56</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>876</td>
<td>3</td>
<td>1.47</td>
<td>3.76</td>
<td>0.76</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bohemia (n = 32)</td>
<td>2,465</td>
<td>41</td>
<td>5.59</td>
<td>11.95</td>
<td>3.3</td>
<td>0.28</td>
<td>0.241 NS</td>
<td>1.286 NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>21</td>
<td>7.7</td>
<td>20.26</td>
<td>3.84</td>
<td>0.18</td>
<td>1.355 NS</td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>876</td>
<td>8</td>
<td>2.97</td>
<td>3.45</td>
<td>2.82</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plech (n = 50)</td>
<td>2,159</td>
<td>56</td>
<td>6.5</td>
<td>16.71</td>
<td>4.56</td>
<td>0.27</td>
<td>0.105 NS</td>
<td>0.422 NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>891</td>
<td>34</td>
<td>11.15</td>
<td>27.17</td>
<td>6.3</td>
<td>0.23</td>
<td>1.031 NS</td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>780</td>
<td>10</td>
<td>3.14</td>
<td>4.94</td>
<td>2.59</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ithaca (n = 32)</td>
<td>2,529</td>
<td>21</td>
<td>2.52</td>
<td>6.23</td>
<td>1.17</td>
<td>0.19</td>
<td>0.159 NS</td>
<td>0.762 NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>942</td>
<td>12</td>
<td>4.08</td>
<td>10.88</td>
<td>1.98</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>834</td>
<td>2</td>
<td>0.48</td>
<td>1.16</td>
<td>0.27</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short allele (n = 23)</td>
<td>2,529</td>
<td>17</td>
<td>2.38</td>
<td>5.64</td>
<td>1.03</td>
<td>0.18</td>
<td>0.217 NS</td>
<td>1.108 NS</td>
</tr>
<tr>
<td>Long allele (n = 9)</td>
<td>2,571</td>
<td>13</td>
<td>2.23</td>
<td>5.61</td>
<td>1.18</td>
<td>0.21</td>
<td>0.495 NS</td>
<td>0.937 NS</td>
</tr>
<tr>
<td>Mayodon (n = 27)</td>
<td>2,270</td>
<td>13</td>
<td>1.31</td>
<td>2.77</td>
<td>1.2</td>
<td>0.43</td>
<td>0.426 NS</td>
<td>$-0.399$ NS</td>
</tr>
<tr>
<td>Exon 1</td>
<td>922</td>
<td>10</td>
<td>2.56</td>
<td>5.35</td>
<td>1.71</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 2 + 3</td>
<td>843</td>
<td>2</td>
<td>0.49</td>
<td>0</td>
<td>0.64</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All populations (n = 295)</td>
<td>1,914</td>
<td>88</td>
<td>5.74</td>
<td>12.83</td>
<td>3.23</td>
<td>0.25</td>
<td>$-0.643$ NS</td>
<td>*</td>
</tr>
</tbody>
</table>

Note.—Number of segregating sites (S), nucleotide diversity ($\pi$) in all (tot), synonymous (syn), and nonsynonymous (nonsyn) sites/bp, ratio of nonsynonymous to synonymous nucleotide diversity ($\pi$/\(\pi_c\)), tests statistics and its significance for LD (ZnS) and departures from the neutral model (Tajima’s D and HKA tests). n = number of sequences.

* Could not be tested due to insufficient polymorphisms.

* $n = 132$; details are given in Materials and Methods.

* $P < 0.05$.

** $P < 0.01$. 

Role of FRI in Arabidopsis lyrata 325
FRI would account for a large part of the flowering time differences between populations. Considering the large number of both synonymous and nonsynonymous polymorphisms, there was only 1 fixed, nonradical amino acid change between populations. The sequence divergence between populations (FST values) was not high compared with neutral microsatellites of the same populations (Müller et al. 2007; supplementary fig. S4, Supplementary Material online), as would be expected in case of local adaptation (e.g., Beaumont and Balding 2004).

Among the numerous nonsynonymous polymorphisms, the most interesting was the indel of 14 amino acids in the last exon that was segregating in several populations. The indel polymorphism was associated with flowering time in the 2 studied Scandinavian populations, significantly in Spiterstulen and close to significantly in Müllom populations. Because of LD, the effective site may be a linked site outside the indel. For example, the Arg/Lys (site 437 in fig. 2) polymorphism in the same gene was completely linked with the indel at site 554 in all studied populations. The LD within populations does not decay significantly within the approximately 3,000 bp that were sequenced (O. Savolainen, unpublished data). It is possible that there is still another linked locus that causes the flowering time differences. The association was found in 2 rather distant populations, from Sweden and Norway, that have considerable differentiation (FST = 0.25) for microsatellites (Muller et al. 2007), and we can expect that such distant populations tend to have different nonrandom associations between variants. Further, as the 14-amino acid length difference is a large change in a protein and the FRI gene is the top candidate for natural variation in vernalization requirement, we find it likely that this indel itself is responsible for the flowering time differences between the scored genotypes in the association study.

When transformed to A. thaliana, the 2 alleles resulted in flowering times that had a difference in the same direction as in the association study. This also supports that variation in the FRI region itself causes flowering time variation. The 2 transformed alleles had 4 nonsynonymous differences in addition to the indel. One of them is within a predicted coiled coil domain (Johanson et al. 2000). In principle, any of the polymorphisms could have contributed to the phenotypic differences but among them the large indel remains the most promising causative polymorphism. The transformation results should be interpreted with caution because it is possible that variation in A. lyrata constructs is enhanced due to overexpression.

Considering that the indel or a linked polymorphism has such large effect on flowering, it is unlikely to be neutral to plant’s fitness. However, the frequencies of the indel were surprisingly high in many populations, in 2 continents. As plants experience natural vernalization during the winter, the flowering time variation that can be seen in a growth chamber without vernalization might not be expressed in natural conditions after the first growing season. The variation in flowering time caused by this gene could thus be hidden from natural selection in later years. However, during the first growing season the plants have an opportunity to flower if they do not have a strong vernalization requirement (Riihimäki and Savolainen 2004). Thus, it is likely that this trait is expressed at least in the first season but perhaps much less in later seasons in cold climates. It is also possible that FRI genotypes have pleiotropic effects on other traits that are under selection during all life stages. In A. thaliana, FRI is known to affect requirement for both vernalization and water use efficiency (McKay et al. 2003).
The intermediate allele frequencies in several populations argue against directional selection in favor of either indel type. It is not likely that either of the alleles is deleterious because frequencies of deleterious alleles are expected to be very low due to purifying selection. Alternatively, 1 of the alleles (e.g., the short allele) could be advantageous on its way to fixation. However, advantageous mutations are fixed very rapidly in the evolutionary timescale, and they are not expected to be segregating for long time. We could not see an increased level of LD or a reduced level of diversity in either of the alleles as would be expected for a positively selected allele (table 1). Fay and Wu’s test did not indicate positive selection.

This suggests that the most probable mode of selection acting on the long and short alleles is balancing selection, which maintains both alleles at intermediate frequencies within populations. A peak of diversity caused by divergence of the alleles that are maintained for a long time due to balancing selection would be expected around the selected site (Kreitman and Hudson 1991). This kind of pattern could not be seen (fig. 4, supplementary fig. S3, Supplementary Material online). However, the diversity peak is expected only if the alleles are so old that there has been enough time for mutations to accumulate, and if recombination is low enough so that the window of high diversity remains large (Charlesworth 2006) and the chances to detect a peak may be low for outcrossing populations (Wright and Gaut 2005). Tajima’s D is expected to be positive in case of balancing selection. There was a significantly positive D not only in Mjöllom but also in Karhumäki where the indel was not segregating. Positive Tajima’s D can result from demographic factors too, such as bottleneck associated with glaciations (Wright et al. 2006; Muller et al. 2007). The final interpretation is only possible against the genome-wide pattern, not yet available. There is no obvious explanation for the excess of silent variation in the first exon because we could not find any other features that would vary over the gene, for example, the Tajima’s D or MacDonald–Kreitman test.

Natural variation in flowering behavior is thus governed by FRI in both A. thaliana and A. lyrata. However, there are differences in patterns that may be at least partly due to life history differences. Whereas several independent loss-of-function mutations cause local adaptation in A. thaliana, the fraction of loss-of-function mutations is negligible in A. lyrata. Instead, there are more subtle changes, such as the 14-amino acid indel polymorphism that does not destroy the function. Some functionality of FRI may be indispensable in A. lyrata but optional in A. thaliana. This difference could also be due to selfing versus outcrossing mode of reproduction. The loss-of-function FRI alleles in A. thaliana are recessive (Johanson et al. 2000). In a selfing species, the probability of fixation of recessive or dominant mutations is almost equal because recessive mutations occur soon as homozygous and are exposed to selection. The short FRI variant seemed to be dominant in our experiment (fig. 3). In the outcrossing A. lyrata, the probability of the initial increase in frequency is much higher for a new dominant than recessive advantageous mutation.

Contrary to the situation in A. thaliana, the current data do not support a major role for FRI in differences between populations in flowering behavior of A. lyrata. Frequency differences of the FRIGIDA indel (or a linked) polymorphism can account for some phenotypic divergence between populations. Because the north European late-flowering populations harbor both late- and early sequence variants at intermediate frequencies and the late-flowering variant is most frequent in the southern early flowering European population, other important genetic factors must be responsible for the flowering time differences between the populations. Loss-of-function mutations of downstream genes, such as FLC, would be interesting candidates. Instead, the data suggest that FRI polymorphism would account for some within-population variation in flowering time in nonvernalized conditions. This polymorphism may be maintained by balancing selection in many populations. The lack of the polymorphism in the Alaskan and Russian populations could be due to the loss of the indel mutation in these lineages or because the mutation has occurred only after they have diverged from other lineages. Balanced polymorphism can be maintained by selection in spatially heterogeneous environments (under some fitness configurations), often mediated by pleiotropic antagonistic effects (reviewed in Hedrick 2006). Changes in water use efficiency could be such a pleiotropic effect (McKay et al. 2003). The perenniality of A. lyrata may also offer wider possibilities for fitness trade-offs for different flowering alternatives. Flowering at small size, as perhaps conferred by short FRI, would result in early reproduction, but it on the other hand reduces chances of survival in later years. Studies that associate fitness-related traits at different life stages with different FRI genotypes in the field conditions would give direct evidence on the nature of evolutionary forces maintaining polymorphisms.

Supplementary Material

Supplementary table S1 and figures S2–S4 are available at Molecular Biology Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Caroline Dean for providing us with the FRI sequence, to R. M. Amasino for sending FRI-Col seeds, and to Juha Lemmetyinen for pBI121 binary vector and Agrobacterium tumefaciens, to C. H. Langley, Barbara Mable, and Rodney Mauricio for A. lyrata seeds, to Soile Suvi Sarlin and Tuomas Toivainen for laboratory assistance, and to Montserrat Aguadé and Pekka Pamilo for comments on an earlier version of the manuscript. This work was supported by grants from the Graduate School of Population Genetics to P.R., from Academy of Finland to O.S. (77311, 211489), and from the Finnish Cultural Foundation to A.N.

Literature Cited


Lauren McIntyre, Associate Editor

Accepted November 17, 2007