Allelic Variation in the Perennial Ryegrass FLOWERING LOCUS T Gene Is Associated with Changes in Flowering Time across a Range of Populations1[W]

Leif Skøt, Ruth Sanderson, Ann Thomas, Kirsten Skøt, Danny Thorogood, Galina Latypova, Torben Asp, and Ian Armstead*

Institute of Biological, Environmental, and Rural Sciences, Aberystwyth University, Aberystwyth, Ceredigion SY23 3EB, United Kingdom (L.S., R.S., A.T., K.S., D.T., G.L., I.A.); and Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Research Centre Flakkebjerg, Aarhus University, 4200 Slagelse, Denmark (T.A.)

The Arabidopsis (Arabidopsis thaliana) FLOWERING LOCUS T (FT) gene and its orthologs in other plant species (e.g. rice [Oryza sativa] OsFTL2/Hd3a) have an established role in the photoperiodic induction of flowering response. The genomic and phenotypic variations associated with the perennial ryegrass (Lolium perenne) ortholog of FT, designated LpFT3, was assessed in a diverse collection of nine European germplasm populations, which together constituted an association panel of 864 plants. Sequencing and genotyping of a series of amplicons derived from the nine populations, containing the complete exon and intron sequences as well as 5’ and 3’ noncoding sequences of LpFT3, identified a total of seven haplotypes. Genotyping assays designed to detect the genomic variation showed that three haplotypes were present in approximately equal proportions and represented 84% of the total, with a fourth representing a further 11%. Of the three major haplotypes, two were predicted to code for identical protein products and the third contained two amino acid substitutions. Association analysis using either a mixed model with a relationship matrix to correct for population structure and relatedness or structured association with further correction using genomic control indicated significant associations between LpFT3 and variation in flowering time. These associations were corroborated in a validation population segregating for the same major alleles. The most “diagnostic” region of genomic variation was situated 5’ of the coding sequence. Analysis of this region identified that the interhaplotype variation was closely associated with sequence motifs that were apparently conserved in the 5’ region of orthologs of LpFT3 from other plant species. These may represent cis-regulatory elements involved in influencing the expression of this gene.

1 This work was supported by a Strategic Programme Grant and Follow-on Fund award from the UK Biotechnology and Biological Sciences Research Council.

* Corresponding author; e-mail ipa@aber.ac.uk.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Ian Armstead (ipa@aber.ac.uk).

1W The online version of this article contains Web-only data. www.plantphysiol.org/cgi/doi/10.1104/pp.110.169870
ortholog of Arabidopsis (Arabidopsis thaliana) FLOWERING LOCUS T (AtFT; Kojima et al., 2002) and of perennial ryegrass FT3 (LpFT3).

The AtFT gene codes for a phosphoethanolamine-binding protein (PEBP), a family of proteins that are known to be of key importance in the floral induction process in plants (Yamaguchi et al., 2005; Ahn et al., 2006; Yan et al., 2006; Jaeger and Wigge, 2007; Tamaki et al., 2007; Igasaki et al., 2008; Komiya et al., 2008; Turck et al., 2008). In Arabidopsis, AtFT is believed to function as an integrator of different flowering induction pathways (Komeda, 2004; Wu et al., 2008; Michaels, 2009), and the effects of modulating FT expression on the flowering phenotype have been extensively studied (Komeda, 2004; Moon et al., 2005; Wigge et al., 2005; Hisamatsu and King, 2008; King et al., 2008; Turck et al., 2008). There is also strong evidence that the FT protein itself can act as a transmissible “florigen,” the signal molecule that moves from the leaf to the shoot apex in floral meristem induction (Corbesier et al., 2007; Tamaki et al., 2007). The direct experimental manipulation of FT orthologs has also demonstrated the potentially pivotal role of FT in monocot systems (Hayama et al., 2003; Hayama and Coupland, 2004; Yan et al., 2006) and has been associated with quantitative variation in flowering time in a number of monocot genetic and physiological studies (King et al., 2006; Bonnin et al., 2008; Hemming et al., 2008; Li et al., 2008; Kikuchi et al., 2009; Sasani et al., 2009; Stracke et al., 2009), including ryegrass (Arnstein et al., 2004, 2008).

Linkage disequilibrium (LD)-based association genetics provide means for statistically validating the coincidence of allelic and phenotypic variations. These approaches have been widely used in animal systems and in human genetics, but more recently, a number of suitable populations have been developed for plant species (Skøt et al., 2005, 2007; Nordborg and Weigel, 2008; Stich et al., 2008; van Berloo et al., 2008; McMullen et al., 2009; Smith et al., 2009; Stracke et al., 2009; Waugh et al., 2009; Yao et al., 2009; Zhang et al., 2009). Skøt et al. (2007) described the development of a perennial ryegrass population consisting of nine populations of diverse European origin suitable for candidate gene evaluation of flowering phenotype through association analysis (the LD family). This study confirmed the role of LpHd1 (an ortholog of rice OsHd1 and Arabidopsis CONSTANS) in affecting time of flowering. In this work, we describe the range and molecular nature of allelic variation for LpFT3 in the same population and demonstrate that there are significant associations between this variation and flowering time. Furthermore, these are confirmed in a validation population segregating for these polymorphisms.

RESULTS

Allelic Variation in LpFT3

A total of 3,668 bp containing the target gene LpFT3 was sequenced from the perennial ryegrass bacterial artificial chromosome (BAC) clone (accession no. FN993916), which was used to design internal primers. Initial sequencing of a nested 1,500-bp segment, containing the complete LpFT3 coding sequence, from 288 clones isolated from a representative sample of the association mapping panel (LD family) revealed the presence of three major haplotypes representing 36% (A), 30% (B), and 25% (C) of the total. Subsequent genotyping and sequencing identified four further haplotypes present at lower frequencies (D–G; Supplemental Fig. S1). From these sequences, the four genotyping primer pairs FT-LD1 to -4 were designed, which spanned insertion/deletion (indel)-type variation that distinguished the haplotypes (Supplemental Fig. S2).

The protein predicted from haplotypes B, C, D, and F consisted of 177 amino acids and was predicted to have three exons, identical to GenBank accession ABC33722 (LpFT3, perennial ryegrass). The protein derived from haplotypes A and G could be distinguished from those derived from B, C, D, and F by a Val/Ile substitution at position 23 and an Ala/Pro substitution at position 156. The haplotype E protein could be distinguished from the B-type protein just by the Val substitution at position 23 (Supplemental Fig. S3).

Genotyping of the Nine Perennial Ryegrass Populations

Assaying of the nine populations constituting the LD family with markers FT-LD1 to FT-LD4 allowed for the discrimination of six alleles with FT-LD1, three with FT-LD2, and four with FT-LD3/4 (Supplemental Table S1). A further haplotype, G, was identified with FT-LD4 (identical to haplotype A, except for a single base deletion at position 1,635). However, except for the FT-LD4 E allele confirmation, none of the FT-LD4 data were included, since FT-LD4 could not discriminate all of the other haplotypes. Overall, approximately 84% of the population consisted of combinations of the A, B, and C haplotypes, with a further 11% contributed by the E haplotype. However, there was considerable variation between the nine populations in the frequency of different genotype classes, varying from two in Ba11304 to 15 in Ba10158 and 16 in Ba10113 (Supplemental Tables S2 and S3).

Genotype/Phenotype Associations

Flowering time of the LD family showed a bimodal distribution in both years, and there were generally earlier flowering times in 2005 compared with 2004 (Fig. 1). To what extent this was due to yearly environmental variation or to the difference in cultivation conditions is difficult to ascertain. A more detailed statistical analysis of the associations between populations, years, and flowering time has been presented (Skøt et al., 2007). The variance component estimate was 73% for the effect of subpopulation within the LD family. Thus, when considering the three major alleles, a, b, and c, the general observation that genotypes with a and/or b alleles tended to be later flowering than...
genotypes with c alleles (Supplemental Tables S4 and S5; Supplemental Figs. S4 and S5) is confounded by the considerable population structure within the LD family (Supplemental Table S6; Supplemental Fig. S6).

To establish the most suitable analytical model for identifying significant population-wide genotype/phenotype associations in the presence of such population structure and relatedness, comparisons were made between a simple ANOVA, a general linear model incorporating structured associations (SA), SA plus genomic control (GC), and a mixed linear model incorporating a relationship matrix as a random effect. The expected $P$ values plotted against observed genome-wide “neutral” amplified fragment length polymorphism (AFLP) markers are shown in Supplemental Figure S7. While the simple and SA models gave inflated $P$ values, both the SA+GC model and the mixed linear model indicated good correction for population structure and relatedness. The relationship matrix corrected for both structure between populations and individual relatedness, while GC corrected for relatedness not accounted for by the SA model alone. The $P$ values for the three FT-LD markers increased substantially from the simple model to SA+GC but were still remarkably low for the mixed model (Table I).

Part of the explanation may be found in the composition of the relationship matrix. Experimenting with a matrix diagonal lower than 2 gave no restricted maximum likelihood (REML) values, perhaps indicating a nonpositive semidefinite matrix, which is a requirement (Kang et al., 2008). It was possible to reduce the diagonal to $1 + F_{ST} (=1.1184)$ using a kinship matrix and still get a REML output, but it had little effect on the $P$ values. Changing the matrix diagonals altered the balance between $V_g$ and $V_c$ values. However, using the sum of $V_g$ and $V_c$ in the calculation of the percentage of the phenotypic variance accounted for by the marker made these values relatively insensitive to matrix changes. As for the $P$ values, the genetic effect was estimated based on a relationship matrix with 2 as the diagonal, which was slightly more conservative than other estimates.

Whichever model was used, the FT-LD1 $a$ and $c$ alleles both had highly significant effects. At the FT-LD2 locus, all three alleles had significant effects, as did the $a$ and $b$ alleles at FT-LD3. Four alleles (FT-LD1 $a$ and $c$, FT_LD2 $a$, and FT_LD3 $a$) accounted for most of the phenotypic variation (Table I). The lack of effect of the FT-LD1 $b$ allele can partly be explained by its recessive nature.

These LD family-wide significant results were corroborated using within-population ANOVA, which was not influenced by population structure to the same degree (Supplemental Fig. S8). This indicated a range of significant LpFT3 genotype/phenotype associations, similar to those identified using the mixed model within-relationship matrices (Supplemental Table S7), and served as a further indication of a role for LpFT3 variation in influencing the flowering time phenotype.

The second experiment was carried out to confirm the effect of genotypic variation associated with the FT-LD1 marker on flowering time without the confounding effect of population structure present in the LD family. We analyzed a population of 563 genotypes derived from a polycross (see “Materials and Methods”) segregating for the $a$, $b$, $c$, and $e$ alleles, termed the validation population. This population contained all the expected 10 genotype classes, although progeny containing the $b$ or $c$ allele were relatively underrepresented (Supplemental Table S8A). A comparison of the 10 resultant genotype classes in relation to flowering time (ANOVA) indicated highly significant effects of allelic substitution (Supplemental Table S8B). Variance component analysis showed that genotypes accounted for 13.8% of the total variance. It also indicated the apparent dominance of the $c$ allele, as the four genotypic classes containing the $c$ allele were the earliest to flower (Fig. 2). The overall effects of allelic substitution on flowering time were similar to those observed in the LD family (compare Fig. 2 with Supplemental Fig. S9).

Two sets of genotypes, $bb$, $be$, and $ee$ on the one hand and $cc$, $ca$, and $aa$ on the other, are segregants from the two different pairs of grandparents. Population structure can definitely be discounted within these, as they are progeny from full sib matings. The $bb$ genotypes had significantly later flowering time than $ee$ and $be$ genotypes (Fig. 2). Similarly, $aa$ geno-

Figure 1. Flowering times for the LD family in 2004 and 2005.
types had significantly later flowering time than cc and ccα genotypes (Fig. 2). These findings confirm the results of the association analysis in the LD family.

**DISCUSSION**

The outbreeding, heterozygous natures of perennial ryegrass and other forage grasses of importance in temperate agriculture and the broad, but still relatively well characterized, gene pool lend themselves to association genetics approaches for linking genotype with phenotype. While working exclusively with natural variation limits the degree to which the fine control of phenotype can be analyzed in terms of laboratory-based molecular biology, there is the potential for a high degree of biological and agricultural relevance to the findings. Previous work in rice has established the role of *FL* as a key gene in the induction to flowering, with a similar, although not completely analogous, role to that observed for *AtFT* in Arabidopsis (Hayama and Coupland, 2004). Comparative genetics and genomics have also indicated that *FT* is a candidate gene underlying both flowering and vernalization QTLs in a number of *Lolium* and *Hordeum* species studies as well as showing an expression profile compatible with its role as a flowering promoter (King et al., 2006; Armstead et al., 2008; Hemming et al., 2008; Stracke et al., 2009). Fine genetic mapping in Arabidopsis has also indicated that variation at *FT* (specifically, its promoter region) underlies flowering QTLs in natural populations (Schwartz et al., 2009). In this study, a direct link between genotype and flowering phenotype has been established for *LpFT3* across different perennial ryegrass populations using association genetics and a catalogue of the underlying allelic variation assembled. This link between *LpFT3* genotypic variation and annual flowering time in perennial ryegrass is not, in itself, unexpected. What is of greater interest, however, is that variation was identified both in the predicted protein sequence and in noncoding

---

**Table 1. Association analysis of flowering times in 2004 and 2005 with the major alleles at the three loci in the LpFT3 gene**

Probabilities for four models are shown: simple, ANOVA without any correction for population structure; SA, structured association; SA+GC, structured association and genomic control; EMMA MLM, mixed linear model as executed with the program EMMA with relationship matrix included as a random effect. Percentage marker effect estimates the percentage of phenotypic variance accounted for by the marker and was obtained as described in “Materials and Methods.” NS, Not significant.

<table>
<thead>
<tr>
<th>Locus/Allele</th>
<th>Year</th>
<th>Simple</th>
<th>SA</th>
<th>SA+GC</th>
<th>EMMA MLM</th>
<th>Percentage Marker Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT-LD1a</td>
<td>2004</td>
<td>&lt;10$^{-17}$</td>
<td>4 × 10$^{-12}$</td>
<td>3 × 10$^{-10}$</td>
<td>4 × 10$^{-29}$</td>
<td>12.4</td>
</tr>
<tr>
<td>FT-LD1b</td>
<td>2004</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>8 × 10$^{-3}$</td>
<td>0.7</td>
</tr>
<tr>
<td>FT-LD1c</td>
<td>2004</td>
<td>&lt;10$^{-17}$</td>
<td>6 × 10$^{-12}$</td>
<td>5 × 10$^{-10}$</td>
<td>3 × 10$^{-30}$</td>
<td>13.4</td>
</tr>
<tr>
<td>FT-LD1e</td>
<td>2004</td>
<td>5 × 10$^{-8}$</td>
<td>NS</td>
<td>NS</td>
<td>1 × 10$^{-7}$</td>
<td>1.0</td>
</tr>
<tr>
<td>FT-LD2a</td>
<td>2004</td>
<td>&lt;10$^{-17}$</td>
<td>1 × 10$^{-8}$</td>
<td>2 × 10$^{-7}$</td>
<td>2 × 10$^{-21}$</td>
<td>9.0</td>
</tr>
<tr>
<td>FT-LD2b</td>
<td>2004</td>
<td>6 × 10$^{-15}$</td>
<td>1 × 10$^{-2}$</td>
<td>2 × 10$^{-2}$</td>
<td>7 × 10$^{-5}$</td>
<td>1.6</td>
</tr>
<tr>
<td>FT-LD2c</td>
<td>2004</td>
<td>3 × 10$^{-11}$</td>
<td>7 × 10$^{-3}$</td>
<td>1 × 10$^{-2}$</td>
<td>3 × 10$^{-5}$</td>
<td>1.8</td>
</tr>
<tr>
<td>FT-LD3a</td>
<td>2004</td>
<td>&lt;10$^{-17}$</td>
<td>4 × 10$^{-9}$</td>
<td>1 × 10$^{-7}$</td>
<td>5 × 10$^{-20}$</td>
<td>8.5</td>
</tr>
<tr>
<td>FT-LD3b</td>
<td>2004</td>
<td>3 × 10$^{-15}$</td>
<td>2 × 10$^{-3}$</td>
<td>6 × 10$^{-3}$</td>
<td>1 × 10$^{-5}$</td>
<td>1.9</td>
</tr>
<tr>
<td>FT-LD3c</td>
<td>2004</td>
<td>3 × 10$^{-8}$</td>
<td>4 × 10$^{-2}$</td>
<td>NS</td>
<td>1 × 10$^{-3}$</td>
<td>1.0</td>
</tr>
<tr>
<td>FT-LD1a</td>
<td>2005</td>
<td>&lt;10$^{-17}$</td>
<td>&lt;10$^{-17}$</td>
<td>2 × 10$^{-10}$</td>
<td>8 × 10$^{-44}$</td>
<td>20.2</td>
</tr>
<tr>
<td>FT-LD1b</td>
<td>2005</td>
<td>3 × 10$^{-5}$</td>
<td>2 × 10$^{-3}$</td>
<td>NS</td>
<td>1 × 10$^{-7}$</td>
<td>3.0</td>
</tr>
<tr>
<td>FT-LD1c</td>
<td>2005</td>
<td>&lt;10$^{-17}$</td>
<td>&lt;10$^{-17}$</td>
<td>&lt;10$^{-17}$</td>
<td>2 × 10$^{-78}$</td>
<td>34.0</td>
</tr>
<tr>
<td>FT-LD1e</td>
<td>2005</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>0.2</td>
</tr>
<tr>
<td>FT-LD2a</td>
<td>2005</td>
<td>&lt;10$^{-17}$</td>
<td>&lt;10$^{-17}$</td>
<td>4 × 10$^{-8}$</td>
<td>1 × 10$^{-34}$</td>
<td>16.1</td>
</tr>
<tr>
<td>FT-LD2b</td>
<td>2005</td>
<td>&lt;10$^{-17}$</td>
<td>4 × 10$^{-11}$</td>
<td>4 × 10$^{-5}$</td>
<td>1 × 10$^{-10}$</td>
<td>5.0</td>
</tr>
<tr>
<td>FT-LD2c</td>
<td>2005</td>
<td>6 × 10$^{-3}$</td>
<td>2 × 10$^{-3}$</td>
<td>5 × 10$^{-2}$</td>
<td>4 × 10$^{-2}$</td>
<td>0.4</td>
</tr>
<tr>
<td>FT-LD3a</td>
<td>2005</td>
<td>&lt;10$^{-17}$</td>
<td>&lt;10$^{-17}$</td>
<td>2 × 10$^{-8}$</td>
<td>5 × 10$^{-8}$</td>
<td>15.7</td>
</tr>
<tr>
<td>FT-LD3b</td>
<td>2005</td>
<td>&lt;10$^{-17}$</td>
<td>2 × 10$^{-11}$</td>
<td>3 × 10$^{-5}$</td>
<td>6 × 10$^{-11}$</td>
<td>4.9</td>
</tr>
<tr>
<td>FT-LD3e</td>
<td>2005</td>
<td>NS</td>
<td>1 × 10$^{-3}$</td>
<td>4 × 10$^{-2}$</td>
<td>5 × 10$^{-2}$</td>
<td>0.4</td>
</tr>
</tbody>
</table>

---

Figure 2. Flowering times of the 10 genotype groups in the validation population. The error bars represent the LSD (P ≤ 0.05) between genotype group mean values.
regions, most diagnostically within the region 5′ of the coding sequence and so possibly in candidate cis-regulatory regions.

**Predicted Protein Variation within the Haplotypes**

Comparative analysis of FT family and related proteins indicates a high degree of sequence conservation (Faure et al., 2007; Danilevskaya et al., 2008), and the FT proteins predicted for the perennial ryegrass A and B haplotypes are consistent with these. Secondary structure predictions (data not shown) do not indicate a marked effect of the amino acid substitutions that distinguish the A and B haplotype proteins, and neither of these substitutions is associated directly with regions of FT with proposed or defined functions (Hanzawa et al., 2005; Ahn et al., 2006; Faure et al., 2007; Danilevskaya et al., 2008). The amino acid residue at the equivalent to position 23 in LpFT3 is relatively unconserved in FT family proteins (Faure et al., 2007); by contrast, an alignment of 248 FT, FT-like, and other closely related plant PEBP proteins (TERMINAL FLOWER1 [TFL1], TFL1-like, MOTHER OF FT [MFT], MFT-like, and maize [Zea mays] CENTRORADIALIS [ZCN]) indicates that the Pro at the equivalent of position 156 (perennial ryegrass B haplotype protein) is a highly conserved residue and the Ala substitution is unique to perennial ryegrass haplotype A (Supplemental Fig. S10). In this study, allele h was identified as a haplotype A/B recombinant, with the recombination occurring at or near the beginning of the third exon. Therefore, this haplotype would be predicted to code for a protein derived from A-type exons 1 and 2 and B-type exon 3. While only a limited set of comparisons is available for the h allele, the results from this study suggest that the A haplotype-associated phenotype was conferred by variation 5′ of the A/B recombination in allele h and not directly by the Ala/Pro substitution. However, a contrary observation is that in exon 3 of haplotype E, six of the seven single-nucleotide polymorphisms (SNPs) that could distinguish the A from the B/C haplotypes were A-type SNPs with one B/C-type SNP. The latter conferred the retention of (reversion to?) Pro in E haplotype proteins (Supplemental Fig. S1). Either this latter base change occurred only in the A haplotype, subsequent to the divergence of the A and E haplotypes, or there has been reversion to the ‘wild-type’ protein in the E haplotypes. If the latter is true, there may be some functional significance of the Ala/Pro substitution, at least in some circumstances.

**FT Genotype and Flowering Phenotype**

While strong population structure existed within the LD family, the association analysis implemented was able to show that different LpFT3 alleles and genotype classes were significantly associated with the flowering time variation. Additionally, while it is still relatively rare that such associations are confirmed in other populations, we were able to do this in a validation population in which population structure could be discounted. This is particularly useful in that QTLs have been evaluated for flowering time in a number of different perennial ryegrass crosses, and markers based directly on LpFT3 have been mapped in at least two of these mapping populations (Armstead et al., 2008). Thus, the association we have established between LpFT3 genotype and flowering phenotype in this study allows us to reevaluate this previous work. The WSC F2 mapping population was a cross between early- and late-flowering genotypes that segregated markedly for heading date. The major and highly significant QTL was associated with the LpFT3 region. Subsequent sequencing of LpFT3 haplotypes from this population identified FT-LD1 bb, bc, and cc segregating genotype classes (Armstead et al., 2004, 2008), which were associated with average flowering times of 61.1 d (bb), 42.7 d (bc), and 39.2 d (cc) after April 1. This is in line with the observation from the LD family that the FT-LD1 c allele is generally associated with earlier flowering times than the FT-LD1 b allele and is consistent with the results obtained for the validation family (Fig. 2). In the F2 mapping family, there is an overall dominant effect of the c allele over the b allele. This relationship is also evident in the validation population but less clear in the LD family: in Ba10284, the bc average values are closer to the bb rather than the cc average values; in Ba10278, the heterozygote is approximately midrange. In two further mapping families, the ILGI (Armstead et al., 2008) and VRN (Jensen et al., 2005; Studer et al., 2008) families, b and e alleles were segregating as FT-LD1 be and ee genotype classes. QTLs for flowering time were present at or near the LpFT3 genomic location on C7 in both families, but in neither family was the associated effect large, which is not inconsistent with the comparison of be and ee genotype classes in the validation family (Fig. 2). In the LD family, the be/ee genotype comparison was only present in the Ba10158 population, in which no LpFT3 genotype comparisons were significant, so it is not possible to draw any parallels.

**Comparative Genomic Analysis**

The sequence differences in the 5′ region were particularly interesting because of their potential to represent variation in cis-regulatory elements that can influence gene expression. Comparison of the 5′ regions of the perennial ryegrass haplotypes indicates that they are, in general, strongly conserved. When the 5′ regions from *Lolium* are compared with the 500-bp 5′ region of rice *OsFT* (LOC_Os06g06320.1, 12006. m05358|1000 bp_40730p177|2938157.2939159; http://rice.plantbiology.msu.edu/index.shtml), three particular motifs are conserved across the two species in terms of relative positions and distances from the ATG initiation codon (motifs 1–3; Fig. 3). Similar motifs are also present in relatively conserved positions in the Triticeae and *Brachypodium distachyon* FT 5′ regions,
although the homology extends beyond the motifs between these species and perennial ryegrass due to their closer evolutionary distances. Motifs 2 and 3 are also present in similar positions in maize ZCN15 (GenBank accession no. EU241906), a PEBP protein with strong homology to FT (Supplemental Fig. S3), and motif 1 is also possibly represented in ZCN15, although reduced in size and at approximately 410 bp upstream of the ATG as opposed to 179, 187, 238, and 239 bp upstream of the ATG for wheat (Triticum aestivum), perennial ryegrass, rice, and B. distachyon, respectively. Searches in PLACE and Consite confirmed that the sequence of motif 2 was consistent with that of a TATA box but also with the binding consensus sequence (plus strand) for the MADS box transcription factor AGL3 from Arabidopsis. Similar searches did not identify any obvious plant-derived cis-regulatory sequence models for motif 3, although the conserved sequence CTAGACATGG in the perennial ryegrass haplotypes, which overlaps motif 1 (GACATGG), is also consistent with the binding consensus sequence (minus strand) for AGL3 (Consite prediction). Clearly, there is a high probability of obtaining matches to short transcription factor-binding site target sequences by chance, and biological relevance is not indicated from this evidence alone. However, it is interesting that both motifs 1 and 3 are juxtaposed to regions within the perennial ryegrass 5′ sequences that show indel-type variation, which could affect binding affinities (Fig. 3; Supplemental Fig. S1), with motif 1 positioned just upstream of the only region that distinguishes the A, B, and C haplotypes.

In animals and yeast, the transcriptional control of PEBPs has been associated with CCAAT transcription binding domains, targeted by Nuclear Factor Y (NFY)/Heme Activator Protein (HAP) protein complexes, although in plants, the identification of conserved CCAAT domains in gene promoters is less clear (Siefers et al., 2009). A model for the regulation of AtFT in Arabidopsis involves the interaction of the 5′ promoter region with a trimeric complex of the proteins CONSTANS, NFY-B (also known as HAP3), and NFY-C (HAP5; Ben-Naim

Figure 3. Representation of approximately 500 bp 5′ of the ATG start codon of FT3 from perennial ryegrass (haplotypes A–F), F. pratensis (Fp), Italian ryegrass (Lolium multiflorum; Lm), darnel (Lolium temulentum; Lt), wheat (Ta), B. distachyon (Bd), rice (Os), and Z. mays (Zm). For perennial ryegrass A to F, Fp, Lm, and Lt, thick black horizontal bars represent aligned sequences, where breaks represent gaps in the sequence alignments. Vertical dotted lines illustrate the positions of the conserved motifs 1, 2, and 3. For Ta, Bd, Os, and Zm, thin black horizontal bars represent the 500 bp 5′ of the ATG start codon (sequence nonaligned), and thick black horizontal bars illustrate the relative positions of the conserved motifs 1, 2, and 3. The sequence alignments of motifs 1, 2, and 3 are given below.
et al., 2006; Kumimoto et al., 2008; Turck et al., 2008; Siefers et al., 2009). Kumimoto et al. (2008) demonstrated that a trimeric complex of Arabidopsis-derived NFY-B proteins (closely related to yeast HAP3) in combination with yeast HAP2 and HAP5 could significantly bind to a 44-bp sequence starting 890 bp upstream of the ATG initiation codon in the Arabidopsis FT gene. This 44-bp sequence was identified as it contained the closest candidate CCAAT sequence (minus strand) to the ATG (Fig. 4). Interestingly, this 44-bp sequence also contains an overlapping GGACAT motif (shaded black; minus strand) to the ATG initiation; and (3) there has been divergence of sequence variation next to the GGACAT motif might be the first is the case, then the proximity of allelic, indel-type and/or specificity between dicots and monocots. If the type introduced a “bias” toward CCAAT recognition; and (3) there has been divergence of sequence and/or specificity between dicots and monocots. If the first is the case, then the proximity of allelic variation next to the GGACAT motif might be the mechanism that confers the different phenotypes associated with the perennial ryegrass A, B, and C LpFT3 haplotypes. In a related flowering induction study in perennial ryegrass, Skot et al. (2007) noted the association between flowering time and the LpHD1 (CONSTANS) gene that was also clearest in the 5′ untranslated region of that gene. In summary, while it is far from demonstrated that motif 1, as identified in this study, is significant in the control of gene expression, the close association of this conserved motif with a diagnostically variable sequence region in the LpFT3 5′ region deserves further examination. Recent work in animal systems suggests that, due to the theoretical and apparent experimental multiple redundancy of potential transcription factor-binding sites throughout eukaryotic genomes, transcription factor-mediated control of gene expression may be modulated more by quantitative binding variation than by direct, discriminatory sequence specificity (MacArthur et al., 2009; Wunderlich and Mirny, 2009). The implication of this is that a hierarchical assessment of potential DNA-binding sites will depend on developing whole-genome, statistically validated assays for such binding sites. Such analyses are beyond the tools and resources available for perennial ryegrass at this point. However, as our understanding of the control and function of cis-regulatory elements increases and the resources for perennial ryegrass and the related model B. distachyon are developed, it will be interesting to see if it can be established whether the conserved motifs identified in

Figure 4. The 44-bp sequence used by Kumimoto et al. (2008) to study the binding affinity of a NFY-B/HAP2/HAP5 protein complex to the 5′ region of the Arabidopsis FT gene. The first row illustrates the position of the CCAAT domain (shaded black; minus strand); the second row the GGACAT motif (shaded black), and the third row the mutation introduced into the CCAAT domain (shaded gray).

this study play a role in gene regulation, possibly analogous to that established for the CCAAT-binding domains known to be involved in PEBP regulation in animal and yeast biology.

CONCLUSION

In conclusion, this work has demonstrated an association between allelic variation in the LpFT3 gene and flowering time in perennial ryegrass. Analyses of an association mapping panel, a genetic mapping family, and a validation population all indicate that haplotype C associates with early flowering and that the A and B haplotypes associate with later flowering. The additive and dominance characteristics of the different haplotypes are less certain, probably influenced by other genes/LQTLs associated with flowering time. This work also illustrates the complementary properties of association and mapping populations. In the former, there are more alleles and haplotypes, but when they are very rare, their effect is difficult to assess because of the lack of statistical power. In mapping populations, fewer alleles segregate, but their effect can be assessed more accurately, as was the case for the e- and b-containing genotypes in the validation population.

Molecular characterization of LpFT3 in this cross-section of European perennial ryegrass germplasm identified three predominant haplotypes that code for two predicted different LpFT3 proteins, distinguished by two amino acid substitutions. While the protein sequence differences may influence flowering phenotype, the clearest association between phenotype and genotype is in the 5′ noncoding region, where there is indel variation in close proximity to a sequence motif that is conserved across perennial ryegrass, Festuca pratensis, wheat, rice, B. distachyon, and, possibly, maize and Arabidopsis. Thus, this region may influence gene expression (and so flowering time) by modulating transcription binding affinities, and there may be parallels with the control of other PEBPs via CCAAT-binding domains.

MATERIALS AND METHODS

Nomenclature

FT Gene/Protein Names

There is no consistent naming scheme for FT-like gene families based upon conserved syntenic relationships in monocots. In this study, we have desig-
nated the clone under study as LpFT3, in line with the name of an identical protein sequence derived from perennial ryegrass (Lolium perenne) already submitted to public databases (AB33722). However, this is likely to be orthologous to barley (Hordeum vulgare) HvFTI and rice (Oryza sativa) OsFTI2 (Faure et al., 2007; Kikuchi et al., 2009) as well as to rice Hd3a (Kojima et al., 2002) and wheat (Triticum aestivum) and barley VRN3/H3 genes (Yan et al., 2006).

**Alleles and Haplotypes**

Amplification products of primer pair LpFTF and LpFT-LD4R containing LpFT3 are referred to as haplotypes and capitalized (e.g. haplotypes A and B). Size-evaluated amplification products of markers LpFT 1 and 4 are referred to as alleles and italicized in lowercase (e.g. alleles a and b). Supplemental Table S1 shows the relationship between haplotypes and alleles.

**Identification of the Perennial Ryegrass BACs Containing LpFT3**

The perennial ryegrass (approximately five-genome coverage) BAC library has been described previously (Farrar et al., 2007; Langdon et al., 2009). Identification of LpFT3 from the perennial ryegrass BAC library was achieved by PCR screening with the primers Hd3a.1f and Hd3a.3r (for primer sequence and thermal cycling profiles, see Supplemental Table S9). The identity of LpFT3 was confirmed by direct sequencing of the Hd3a.1f/3r-positive BACs with coding sequence and protein prediction generated using FGENESH+ (http://linux1.softberry.com/berry.phtml) and by direct comparison with other monocot FT sequences obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/GenbankSearch.html). The FpFT sequence was similarly identified in a Festuca pratensis BAC library (Donnison et al., 2005).

**Derivation of Perennial Ryegrass Populations for Association Analysis and Validation**

**Association Analysis**

The perennial ryegrass LD family (association mapping panel) and the phenotypic evaluation for flowering time have been described in detail (Skøt et al., 2007). Briefly, the LD family consisted of nine populations, of which seven were natural or seminatural populations of European origin and two were varieties. Each population was represented by 96 genotypes. The entire population (9 × 96 genotypes) was evaluated for flowering time over two consecutive years (2004 and 2005), the first year as individual genotypes grown in 6-inch-diameter pots and the second year as two spaced-plant replicates in the field. DNA was extracted from each of the 864 genotypes as described previously (Skøt et al., 2005) and stored at −20°C.

**Validation Population**

The 16 parental genotypes for this population were obtained from progeny of two crosses. The first was a genotype of the late-flowering variety Bargold with a genotype of Ba10732 (genotype no. 22) from the LD family. The genotypes at the LD-FTI locus were (aa) × (cc). The second cross was a genotype of the late-flowering variety Hugo with another genotype of population Ba10732 (genotype no. 46) from the LD family. The LD-FTI genotypes were (bb) × (aa). The two varieties acted as mother plants, with the Ba10732 genotypes as pollen donors. Progeny from these two crosses were genotyped with the LD-FT1 marker to identify ac and bc genotypes. These two genotypes were used in a 1:1 mixture as parents in a polycross (i.e. one in which random mating is allowed in an isolation greenhouse). The self-incompatibility in perennial ryegrass is almost absolute, so the progeny can be assumed to be derived from crossing between the parental genotypes. The 563 progeny were obtained from 36 randomly chosen mother plants. This mating design ensured a panmictic population in which we have assumed no population structure to be present.

**LpFT3 Allele Amplification**

The LpFT3 alleles from each genotype within the LD family were initially amplified in an approximately 1,800-bp fragment using primers Hd3a.3f and FT-LD3R with HiFiTaq DNA polymerase (Roche) with the supplied PCR buffer. Equal volumes of the amplification reactions from each of the 96 genotypes within each of the nine populations were pooled, and the products were cloned into pCR4-TOPO (Invitrogen). Approximately 50 colonies from each of the populations were grown as minipreps, and the insert was sequenced. The resulting sequences were compared after alignment to identify the allelic variants for LpFT3 present in the screen. In order to avoid misscoring sequencing errors as low-level SNP/indel variations, sequence variants present in only one LpFT3 cloning product were excluded from the evaluation. After sequencing of the LpFT3-containing perennial ryegrass BAC clone, a further primer pair, LpFTF and LpFT4R, was developed that amplified the LpFT3 gene as an approximately 1,900-bp fragment. This primer pair was also used in the amplification of FT from single genotypes of Italian ryegrass (Lolium multiflorum) and dandel (Lolium temulentum).

**LD Marker Development and Analysis**

Four primer pairs (FT-LD.1F/1R, FT-LD.2F/2R, FT-LD.3F/3R, and FT-LD.4F/4R; Supplemental Table S9) were designed that spanned size polymorphisms reflecting the allelic variation within the LpFT3 haplotypes present in the LD family (for haplotype sequences and relative primer and coding sequence positions, see Supplemental Figs. S1 and S2). The reverse primer from each pair was fluorescently labeled for the detection of size polymorphisms on an ABI3130XL Genetic Analyzer (Applied Biosystems), and each genotype was scored for LpFT3 allelic complement. Subsequently, the correspondence between predicted genotype and DNA sequence was checked in 20 individuals from across the nine populations using LpFTF and LpFT4R.

The validation population had been genotypically assessed previously using primer pair Hd3a(LD)F/R (Armstead et al., 2009), which gave the same allelic discrimination as primer pair FT-LD.1F/1R, although with the reverse primer anchored in the 5′ untranslated region rather than in the first exon. This equivalence was checked by regenotyping a subset of the validation population with FT-LD.1F/1R.

**Association Analysis and ANOVA**

Statistical analysis of the association mapping experiment was carried out using a mixed linear model. We implemented this with the efficient mixed model association (EMMA) program in R (Kang et al., 2008). The association test was performed via a t test after estimating variance components using REML. Due to the multilocus nature of the markers used, genotypes containing a particular allele were scored dominantly as a binary marker. For example, the FT-LD3a marker was scored as 1 if the genotype was either homozygous or heterozygous for the a allele. The genetic relatedness was incorporated as a random effect using the relationship coefficient obtained from the software program SPAgeDi version 1.3 (Hardy and Vekemans, 2002) with the genome-wide AFLP data set described previously for this family as input (Skøt et al., 2007). One way ANOVA was carried out using GENSTAT 11.1 (www.vsni.co.uk). Within-population significance was evaluated using (so ≤ 0.05). For population Ba11304, which contained only two genotypic groups, within-population significance was evaluated using a t test. Within-population associations were also carried out using REML with relationship matrices to account for any remaining relatedness within populations.

The proportion of the phenotypic variance accounted for by an individual marker was calculated as 1 − (Vg + Vt + (with marker))/ (Vg + Vt) (without marker)). The variances for this were obtained using the REMLE option in EMMA, either incorporating the marker in the model or not.

The simple ANOVA, SA, and GC+SA were performed in R. The GC method (Devlin and Roeder, 1999) also uses genome-wide markers but has been mostly used in human genetics. As shown here, it is remarkably good at accounting for relatedness existing after correcting for the gross effect of population structure with SA. We used the genomic AFLP markers described above to obtain test statistics. This was compared with the expected value of 1 for a 1-degree of freedom χ² test or F test, in our case with 1 numerator degree of freedom. An observed average test statistic above 1 is an indication of population structure. The test statistic of the candidate markers after GC represents the observed test statistic divided by the average test statistic of the genome-wide markers.

**In Silico DNA Sequence Analyses**

DNA and protein sequences were aligned using ClustalX 2.0.10 (Larkin et al., 2007) and Macaw version 2.0.5 (Schuler et al., 1991) with manual adjustment in GeneDoc (Nicholas and Nicholas, 1997).

Primary protein structure prediction was implemented using FGENESH (http://linux1.softberry.com/berry.phtml). Predicted secondary structures...
for the proteins derived from the perennial ryegrass A and B haplotypes were obtained using PredicProtein (Rost et al., 2004) and Jpred (Cole et al., 2008).

Similarities of potential 5’ cis-regulatory elements in LpFT3 and previously identified motifs were evaluated using PLACE (http://www.dna.afrc.go.jp/PLACE/signalscan.html) and Consite (http://asp.i.uib.no/cgi-bin/CONSITE/consite/).

Accession numbers for nucleotide sequences (Supplemental Sequences S1) are as follows: F. pratensis LpFT3 (BAC), FN993915; perennial ryegrass LpFT3 (BAC), FN993916; LpFT3 haplotype A, FN993917; LpFT3 haplotype B, FN993918; LpFT3 haplotype C, FN993919; LpFT3 haplotype D, FN993920; LpFT3 haplotype E, FN993921; LpFT3 haplotype F, FN993922; LpFT3 haplotype G, FN993923; Italian ryegrass LnFT3, FN993924; darnel LfFT3, FN993925; VRNA family LpFT3 haplotype E, FN993926; LpFT3 haplotype E/B, FN993927; LpFT3 haplotype B, FN993928.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. DNA sequence alignments for LpFT3 haplotypes A to G.

Supplemental Figure S2. FT-LD marker distribution within the LpFT3 amplicons.

Supplemental Figure S3. Comparison of FT protein sequences.

Supplemental Figure S4. FT-LDI genotype frequencies in relation to flowering time.

Supplemental Figure S5. FT-LDI allele frequencies in relation to flowering time.

Supplemental Figure S6. Distribution of flowering times across the nine populations.

Supplemental Figure S7. Comparison of the cumulative probabilities with five different models applied for population structure correction.

Supplemental Figure S8. ANOVA of within-family FT-LDI to -4 genotype/phenotype comparisons.

Supplemental Figure S9. The phenotypic effect of within-population FT-LD-1 allele substitution.

Supplemental Figure S10. Alignment of a segment of 248 LpFT3-related PEBP proteins.

Supplemental Figure S11. The composition of the nine populations that constitute the LD family in terms of FT-LDI genotype and allele frequency.

Supplemental Table S1. FT-LDI to -4 allele sizes in relation to haplotype.

Supplemental Table S2. FT-LDI to -3 genotype frequencies in relation to population.

Supplemental Table S3. FT-LDI to -3 allele frequencies in relation to population.

Supplemental Table S4. FT-LDI genotype frequencies in relation to flowering time.

Supplemental Table S5. FT-LDI to -3 allele frequencies in relation to flowering time.

Supplemental Table S6. Distribution of flowering times across the nine populations.

Supplemental Table S7. Mixed model analysis of the relationship between FT-LDI-1, -2, and -3 genotypes and flowering time within each of the nine populations.

Supplemental Table S8. Genotypic segregation and association with flowering time in the validation population.

Supplemental Table S9. Primer sequences and PCR conditions.

Supplemental Sequences S1. DNA sequences of genomic regions containing Fp/LpFT3 from the F. pratensis and perennial ryegrass BACs, amplification products of perennial ryegrass haplotypes A to H from the LD family, and amplification products from darnel, Italian ryegrass, and the perennial ryegrass VRNA mapping family haplotypes.

ACKNOWLEDGMENTS

We thank Dr. Ian Mackay (National Institute of Agricultural Botany) for his advice concerning the statistical analysis.

Received April 13, 2010; accepted November 24, 2010; published November 29, 2010.

LITERATURE CITED


Armstead IP, Skot L, Turner LB, Skot K, Donnison IS, Humphreys MO, King IP (2005) Identification of perennial ryegrass (Lolium perenne (L.)) and meadow fescue (Festuca pratensis (Huds.)) candidate orthologous sequences to the rice Hd3a(Se1) and barley HvCO1 CONSTANS-like genes through comparative mapping and microsynteny. New Phytol 167: 239–247


Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2: 618–620


Hemming MN, Peacock WJ, Dennis ES, Trevaskis B (2008) Lower-
temperature and daylength cues are integrated to regulate FLOW¬
II. Roles for FLOWERING LOCUS T (FT) and gibberellin. J Exp Bot 59:
3821–3829
ERING LOCUS T/TERMINAL FLOWER 1 family in Lombardy poplar.
Plant Cell Physiol 49: 291–300
Jaeger KE, Wigge PA (2007) FT protein acts as a long-range signal in
Arabidopsis. Curri Biol 17: 1050–1054
Jensen LB, Andersen JR, Frei U, Xing YZ, Taylor C, Holm PB, Lüberssterd
TL (2005) QTL mapping of vernalization response in perennial ryegrass
(Lolium perenne L.) reveals co-locating with an orthologue of wheat
Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Esklin
association mapping. Genetics 178: 1709–1723
Kimomi R, Igekami A, Tamaki S, Yoko S, Shimamoto K (2008) HDA3a and
RFT1 are essential for flowering in rice. Development 135: 767–774
Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM,
in Arabidopsis. Dev Cell 17: 292–299
Larkin MA, Blackshields G, Brown NP, Chenra R, McGgettigan PA,
Li D, Liu C, Shen L, Wu Y, Chen H, Robertson M, Helliwell CA, Ito T,
MacArthur S, Li XY, Li JY, Brown JB, CHC, Zeng L, Gronronpa DB,
roles of 21 Drosophila transcription factors are determined by quanti¬
tative differences in binding to an overlapping set of thousands of
gene regulatory elements. Genome Biol 10: R80
McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li HH, Sun Q,
genomic region for a quantitative trait locus, Hd3, into two loci, Hda3a and
Hdb3b, controlling heading date in rice. Theor Appl Genet 104: 772–778
Nicholas HD, Nicholas HB (1997) GeneDoc: a tool for editing and annotating
multiple sequence alignments. www.nrbsc.org/ (December 15, 2010)
456: 720–723
Res 32: W231–W236
Sasani S, Hemming MN, Oliver SN, Greenup A, Tavakkol-Afshari R,
The influence of vernalization and daylength on expression of flowering¬
time genes in the shoot apex and leaves of barley (Hordeum vulgare).
J Exp Bot 60: 2169–2178
alignment construction and analysis. Proteins: Structure, Function, and
Biotechnologies 9: 130–190
Schwartz C, Balasubramanian S, Wartmann N, Michael TP, Lempe J,
Cis-regulatory changes at FLOWERING LOCUS T mediate natural
variation in flowering responses of Arabidopsis thaliana. Genetics 183:
723–732
Sievers F, Dang K, Kumimoto RW, Bynum WE IV, Tayrose G, Holt BF
transcription factors suggest potential for extensive combinatorial com¬
Skot L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J,
genes with flowering time and water-soluble carbohydrate content in
Lolium perenne (L.). Genetics 177: 535–547
Smith KE, Dobrowski MP, Cogan NOI, Spangenberg GC, Forster JW (2009)
Utilizing linkage disequilibrium and association mapping to
implement candidate gene based markers in perennial ryegrass
breeding. In T Yamada, G Spangenberg, eds, Molecular Breeding of
Forage and Turf. Springer Science+Business Media, New York, pp
259–274
mapping in sugar beet (Beta vulgaris L.). Theor Appl Genet 117: 947–954
Stracke S, Haseneyer G, Veyrietas JB, Geiger HH, Sauer S, Graner A,
Piepho HP (2009) Association mapping reveals gene action and inter¬
actions in the determination of flowering time in barley. Theor Appl
Genet 118: 259–273
(2008) Genetic characterisation of seed yield and fertility traits in
igen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol 59:
573–594
(2008) Diversity and linkage disequilibrium analysis within a selected
set of cultivated tomato (Lycopersicon esculentum Mill) lines. Theor
Appl Genet 117: 89–101
of whole genome association scans in barley. Curr Opin Plant Biol 12:
218–222
Wigge PA, Kim MC, Jaeger KE, Busch W, Schimid M, Lohmann JU, Weigel
D (2005) Integration of spatial and temporal information during floral
Wu CY, You CJ, Li CS, Long T, Chen GX, Byrne ME, Zhang QF (2001)
RIDI, encoding a Cups/His2-type zinc finger transcription factor, acts as
a master switch from vegetative to floral development in rice. Proc Natl
Acad Sci USA 105: 12915–12920
are revealed by analysis of binding motifs. Trends Genet 25: 434–440
OF FT (TSF) acts as a floral pathway integrator redundantly with FT.
Plant Cell Physiol 46: 1175–1189
Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik
gene VRN3 is an orthologue of FT. Proc Natl Acad Sci USA 103: 19581–
19586
agronic traits on chromosome 2A of wheat. Genetica 137: 67–75
and linkage disequilibrium in barley assessed by DArT markers. Theor
Appl Genet 119: 43–52