eLIFE elifesciences.org

RESEARCH ARTICLE

Figures and figure supplements

Natural haplotypes of *FLM* non-coding sequences fine-tune flowering time in ambient spring temperatures in Arabidopsis

Ulrich Lutz *et al*

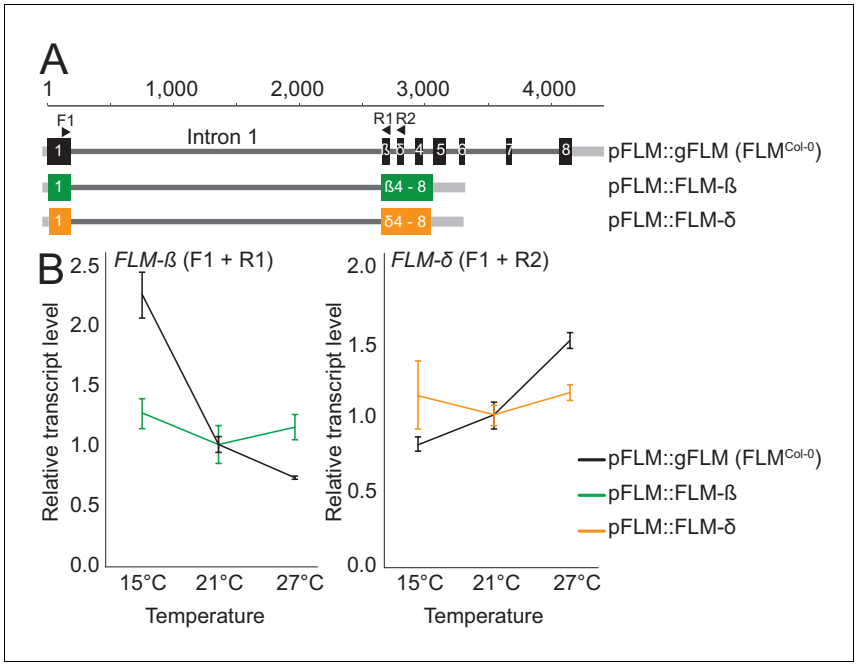

Figure 1. *FLM* intronic sequences determine basal and temperature-dependent *FLM* expression. (**A**) Schematic representation of the pFLM::gFLM (FLM^{Col-0}), pFLM::FLM-ß, and pFLM::FLM-δ constructs. Scale is set to 1 according to the A of the ATG start codon. Black boxes represent exons, grey lines 5'- or 3'-untranslated regions, dark-grey lines introns. Arrows indicate primer binding sites; the forward primer F1 was used to amplify *FLM-ß* and *FLM-δ* with reverse primers R1 and R2, respectively. (**B**) Mean and SD (three biological replicates) of qRT-PCR analyses of *FLM-ß* and *FLM-δ* expression in ten day-old plants harbouring the transgenes shown in (**A**). For normalization, the respective values measured at 21°C were set to 1.

DOI: 10.7554/eLife.22114.002

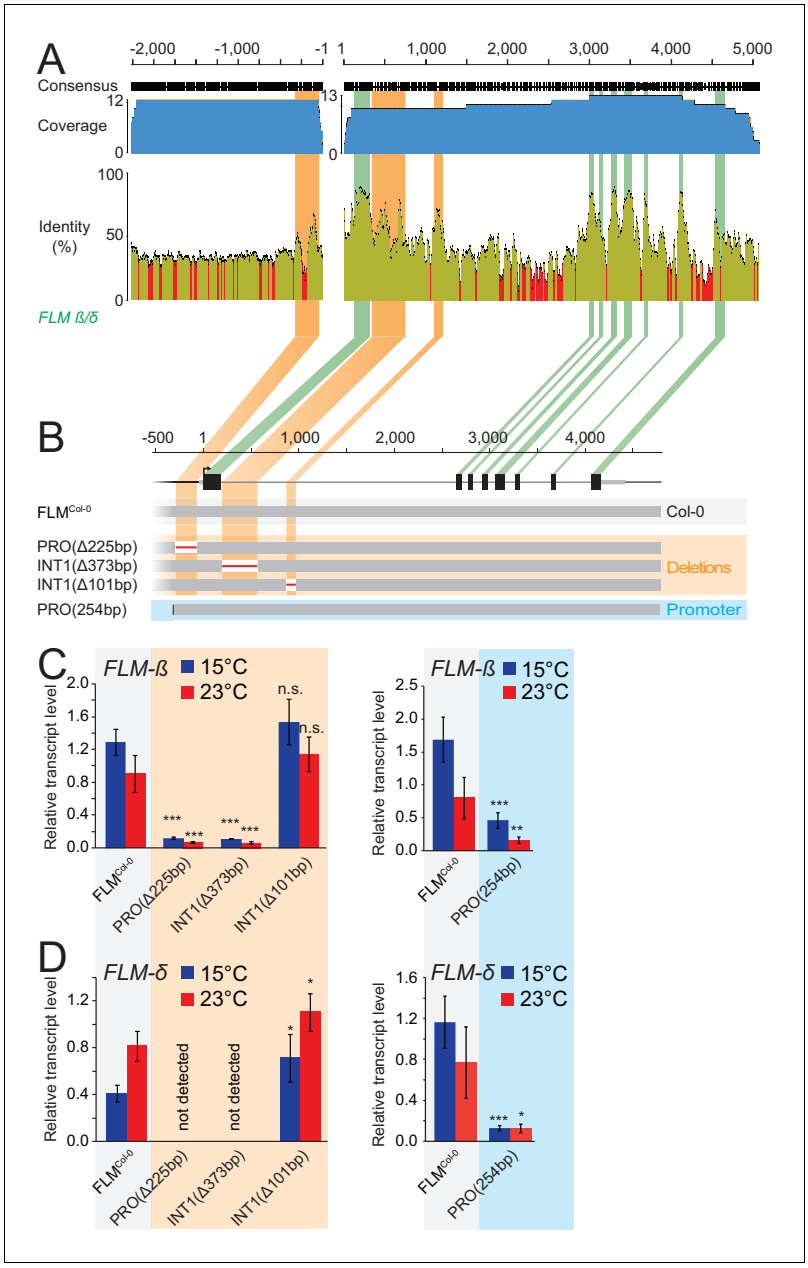

Figure 2. Phylogenetic footprinting identifies promoter and intron 1 regions required for *FLM* expression. (**A**) Phylogenetic footprinting of promoter and genomic regions of *FLM* and putative *FLM* orthologs from six *Brassicaceae* species. Coverage is shown in blue, identities are shown in ocre (≥30%) and red (<30%). Exons are displayed in green, regions with high non-coding sequence conservation are displayed in orange. (**B**) Schematic illustration of the *FLM* genomic region and FLM^{Col-0} transgenic variants used for expression analysis. (**C**) and (**D**) Mean and SD (four replicate pools with five to ten independent T2 transgenic lines) of qRT-PCR analysis of *FLM-β* (**C**) and *FLM-δ* (**D**) expression at 15°C and 23°C in ten day-old seedlings of 21–40 bulked T2 transgenic lines. Student's t-tests: *, p≤0.05; **, p≤0.01; ***, p≤0.001; n.s., not significant.

DOI: 10.7554/eLife.22114.003

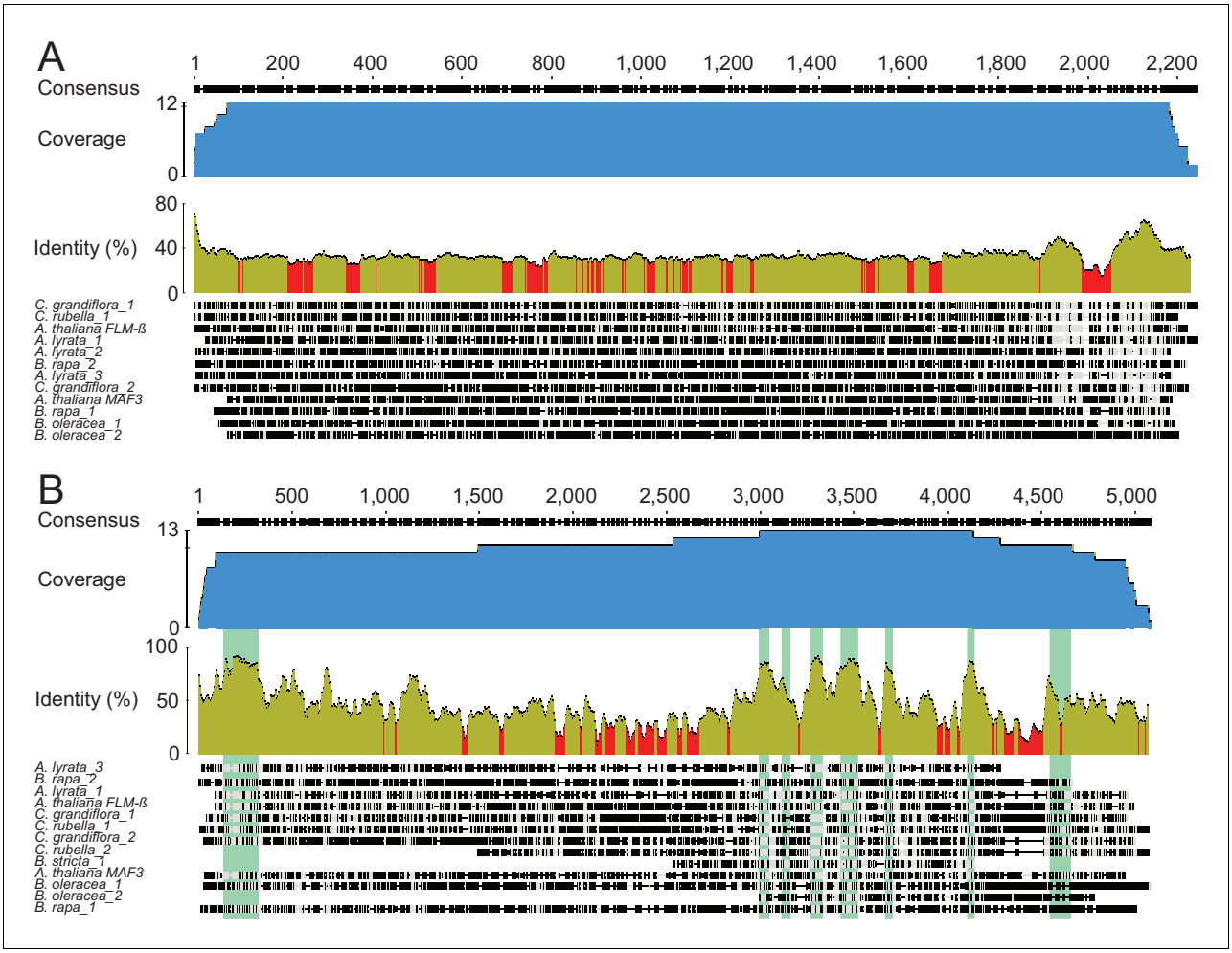

Figure 2—figure supplement 1. Phylogenetic footprinting of promoter and genomic regions of *FLM* and putative *FLM* orthologs from six *Brassicaceae* species. (**A**) and (**B**) Sequence alignments of promoter (**A**) and gene sequences (**B**) of *FLM* and *MAF3* as well as putative *FLM* orthologs from six *Brassicaceae* species. Coverage is shown in blue, identities are based on a sliding window word size = 30 and displayed in ocre (≥30%) and red (<30%). *FLM* exonic regions are shown in green (**B**). Numbering is indicated according to the position in the alignment.

DOI: 10.7554/eLife.22114.004

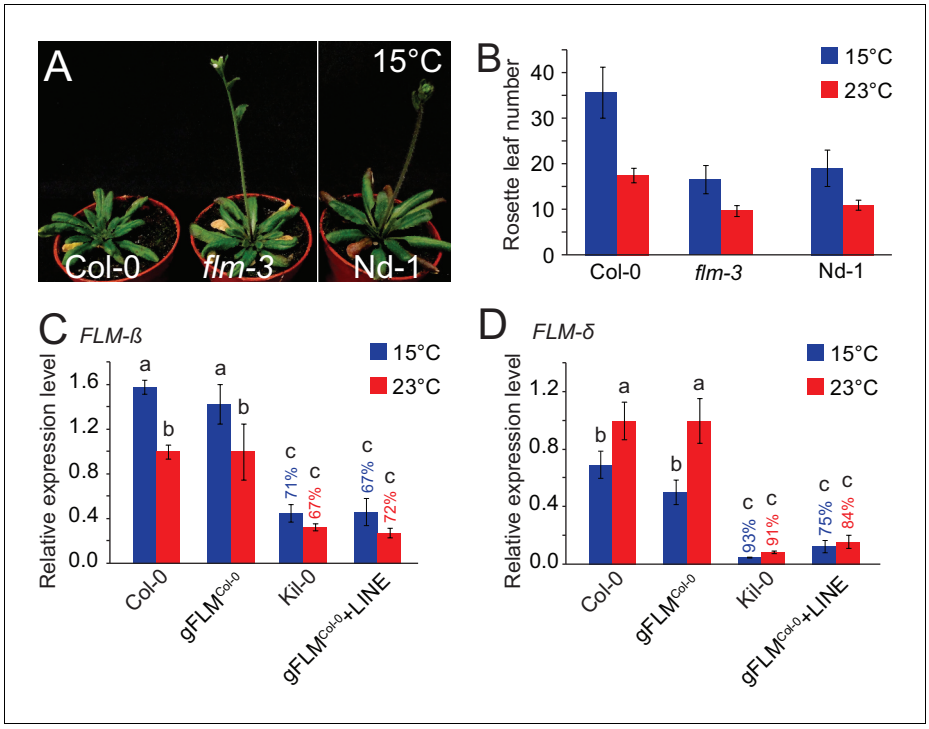

Figure 2—figure supplement 2. The gFLM^{Col-0} and gFLM^{Col-0}+LINE transgenes faithfully recapitulate *FLM* expression in the Col-0 and Kil-0 accessions. (**A**) Representative photographs of 42 day-old plants of Col-0, *flm-3*, and Nd-1 grown at 15˚C in long day photoperiod. The plant images were spliced together but originate from the same photograph as indicated by a white vertical line. (**B**) Quantitative flowering time analysis of Col-0, *flm-3*, and Nd-1 grown at 15˚C and 23˚C in long day photoperiod. (**C**) and (**D**) qRT-PCR analyses of *FLM-β* (**C**) and *FLM-δ* (**D**) expression at 15˚C and 23˚C in ten day-old *flm-3* mutants complemented with a gFLM^{Col-0} or gFLM^{Col-0}+LINE allele (***Lutz et al., 2015***). Pools of 35–40 independent T2 transgenic lines were used in comparison to Col-0 and Kil-0 homozygous lines. Shown is the mean and SD of three (Col-0, Kil-0) and four replicate pools comprising each 8 to 10 transgenic lines. Percentages refer to the reduction in *FLM* expression between the gFLM^{Col-0}+LINE transgenic line in comparison to the gFLM^{Col-0} line or between Kil-0 and Col-0, respectively. Similar letters indicate no significant difference of total leaf number (Tukey HSD, $p<0.05$).

DOI: 10.7554/eLife.22114.005

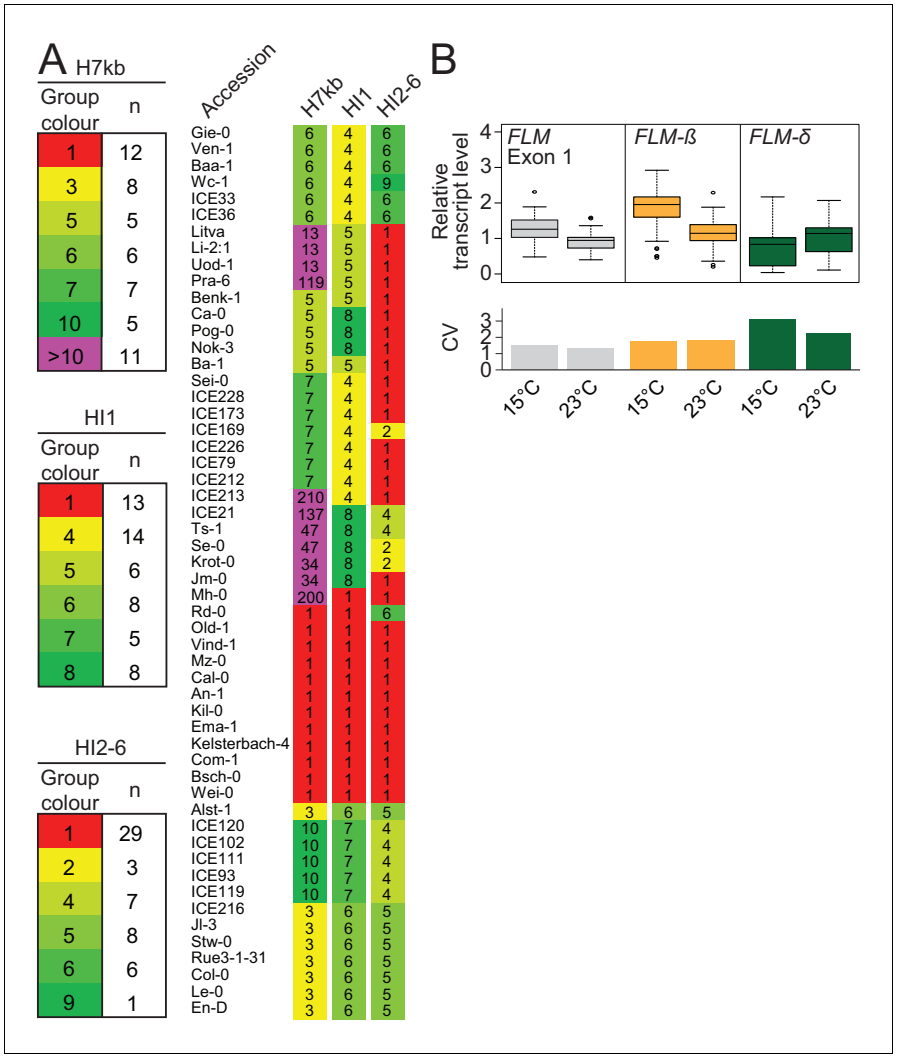

Figure 3. *FLM* haplotype analysis of 776 accessions identifies major haplotypes determined by non-coding variation. (**A**) Haplotype group affiliation of 54 accessions of the *FLM* haplotype set based on 45 SNPs for either the 7 kb *FLM* locus (H7kb), intron 1 (HI1) or introns 2–6 (HI2-6). Group numbering and colouring are according to group size. (**B**) Summarized expression values of total *FLM* (exon 1), *FLM-ß*, and *FLM-δ* expression of the *FLM* haplotype set. Outliers were determined based on 1.5 x IQR (interquartile range). The coefficient of variation (CV) is shown in the lower graph.

DOI: 10.7554/eLife.22114.006

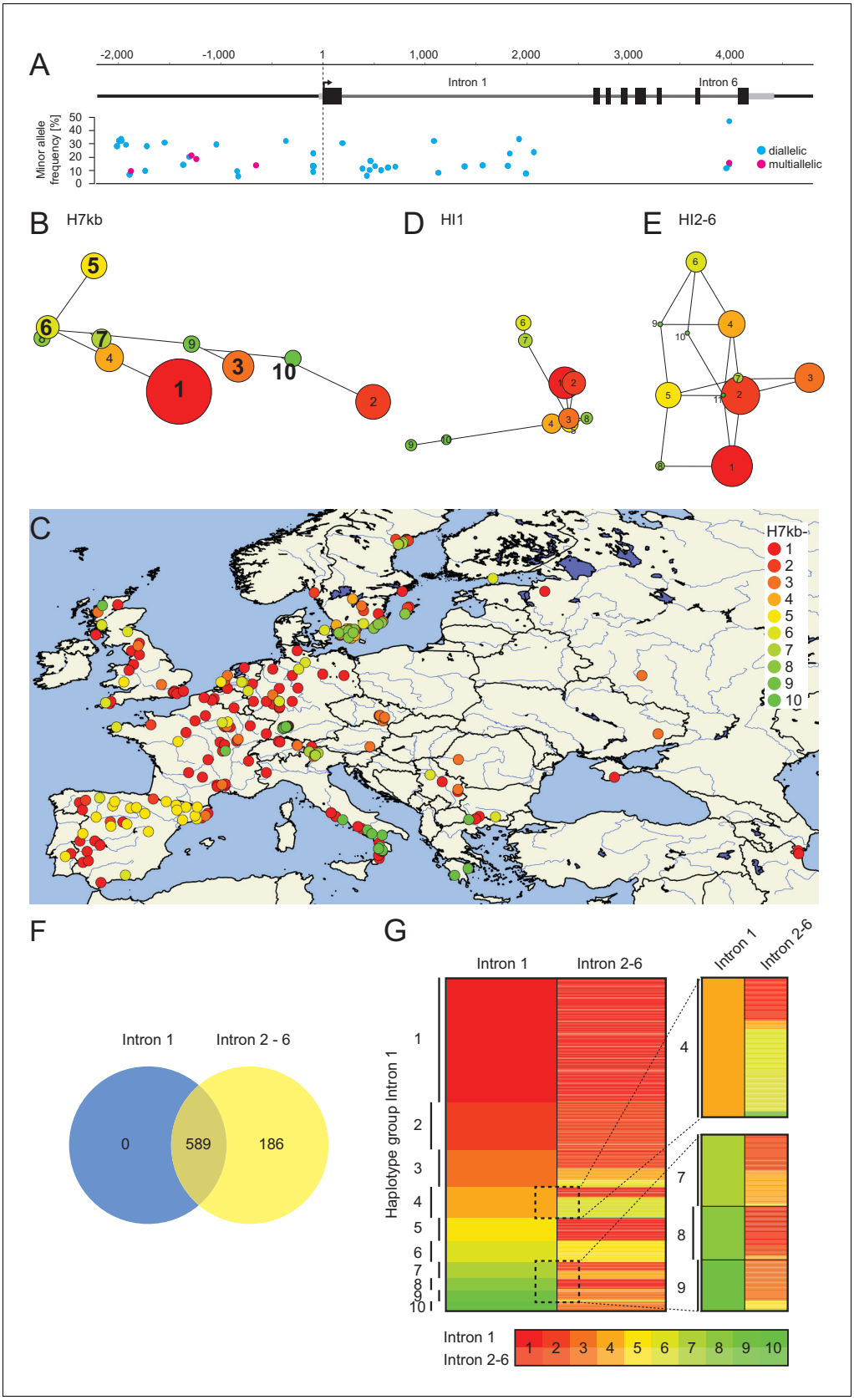

Figure 3—figure supplement 1. Haplotype analysis based on 45 SNPs from 776 accessions. (**A**) Representation of the position, the frequency and the type of 45 SNPs from 776 accessions used for haplotype clustering. (**B**), (**D**), and (**E**) Haplotype network analysis of the ten (**B**) most frequent groups of a

Figure 3—figure supplement 1 continued on next page

Figure 3—figure supplement 1 continued

7 kb *FLM* genomic region based on 45 SNPs or the ten (**D**) or eleven (**E**) largest haplotype groups when only SNPs of intron 1 (HI1; **D**) or introns 2–6 (HI2-6; **E**) are considered. Circle sizes and the color code illustrate number of accessions in each group with group 1 representing 174 (**B**), 220 (**D**) and 356 (**E**) accessions, respectively. In (**B**), bolt numbers indicate groups from which representative accessions were chosen. (**C**) Geographic distribution of the ten most frequent groups of a 7 kb *FLM* genomic region. (**F**) Venn diagram indicating the overlap of accessions included in the ten largest HI1 and HI2-6 haplotype groups. (**G**) Clustering of the haplotype groups of the 589 accessions included in the ten largest HI1 and HI2-6 haplotype groups. Accessions were sorted according to their HI1 and HI2-6 haplotype group numbers. Numbers on the left indicate the HI1 haplotype group number as also shown in (**D**). An enlargement of a subset of accessions highlighted by the dotted boxes is shown on the right. The color coding for haplotype group numbers, which illustrates the group number and group size as shown in (**D**) and (**E**) is specified below the graph.

DOI: 10.7554/eLife.22114.007

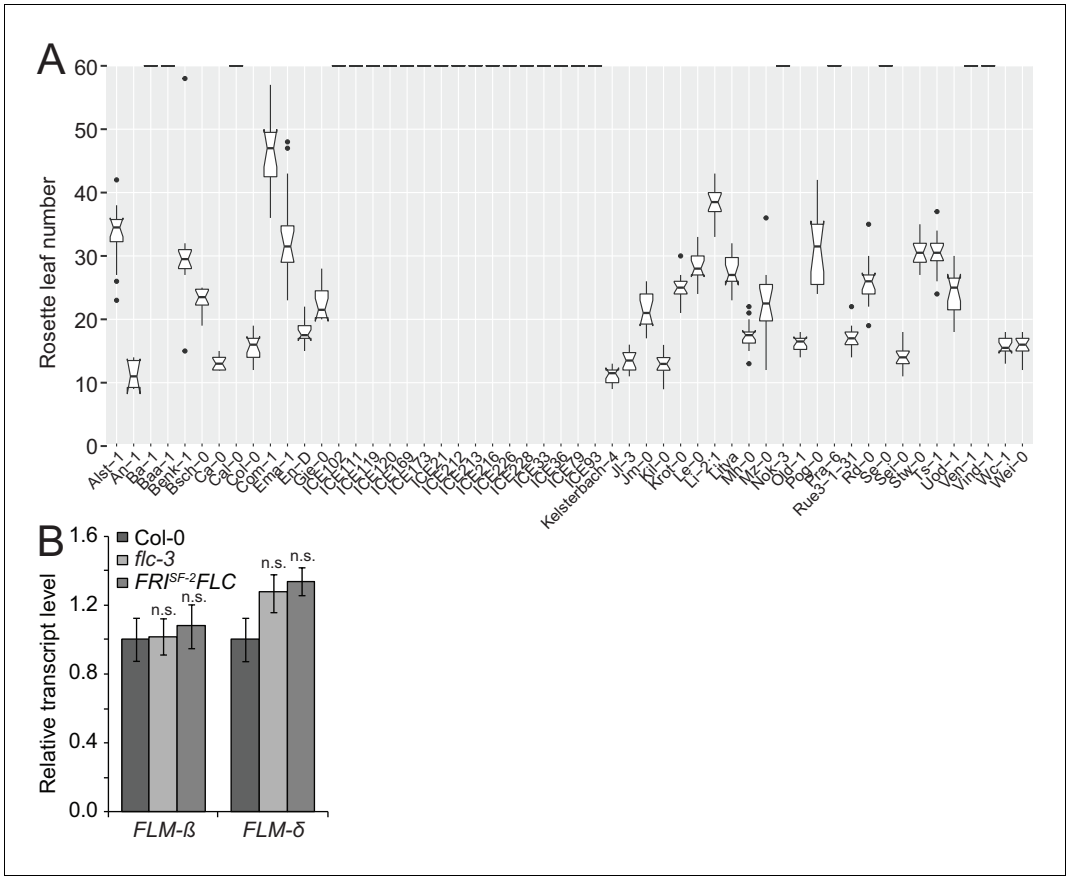

Figure 3—figure supplement 2. Flowering time of the accessions of the *FLM* haplotype set suggest a vernalization requirement for many accessions. (**A**) Flowering time analysis of the accessions of the *FLM* haplotype set. Extremely late flowering accessions or accessions that had not flowered are represented at the 60 rosette leaves margin. Outliers were determined based on 1.5 x IQR (interquartile range). (**B**) Mean and SD (three biological replicates) in qRT-PCR analyses of *FLM-β* and *FLM-δ* expression in twelve day-old Col-0, *flc-3* and Col-0 with a *FRI*$^{Sf-2}$*FLC* vernalization module (***Michaels and Amasino, 1999***).

DOI: 10.7554/eLife.22114.008

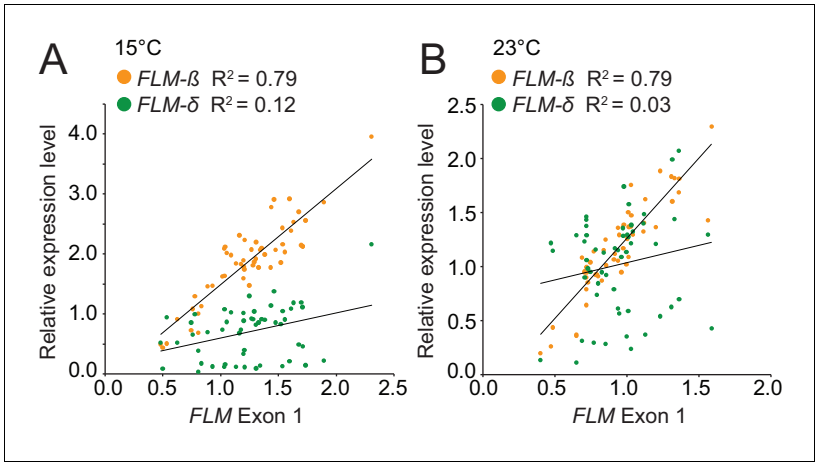

Figure 3—figure supplement 3. Correlation analysis identifies *FLM-β* as the major *FLM* transcript at 15˚C and 23˚C in the *FLM* haplotype set accessions. (**A**) and (**B**) Correlation (simple linear regression) between *FLM* expression as determined by qRT-PCR over *FLM* exon 1 with measurements of *FLM-β* and *FLM-δ* splice variant abundance at 15˚C (**A**) and 23˚C (**B**) in the 54 accessions of the *FLM* haplotype set. *FLM-ß*, 15˚C, $p<0.0001$; *FLM-δ*, 15˚C, $p=0.009$; *FLM-ß*, 23˚C, $p<0.0001$; *FLM-δ*, 23˚C, $p>0.05$.

DOI: 10.7554/eLife.22114.009

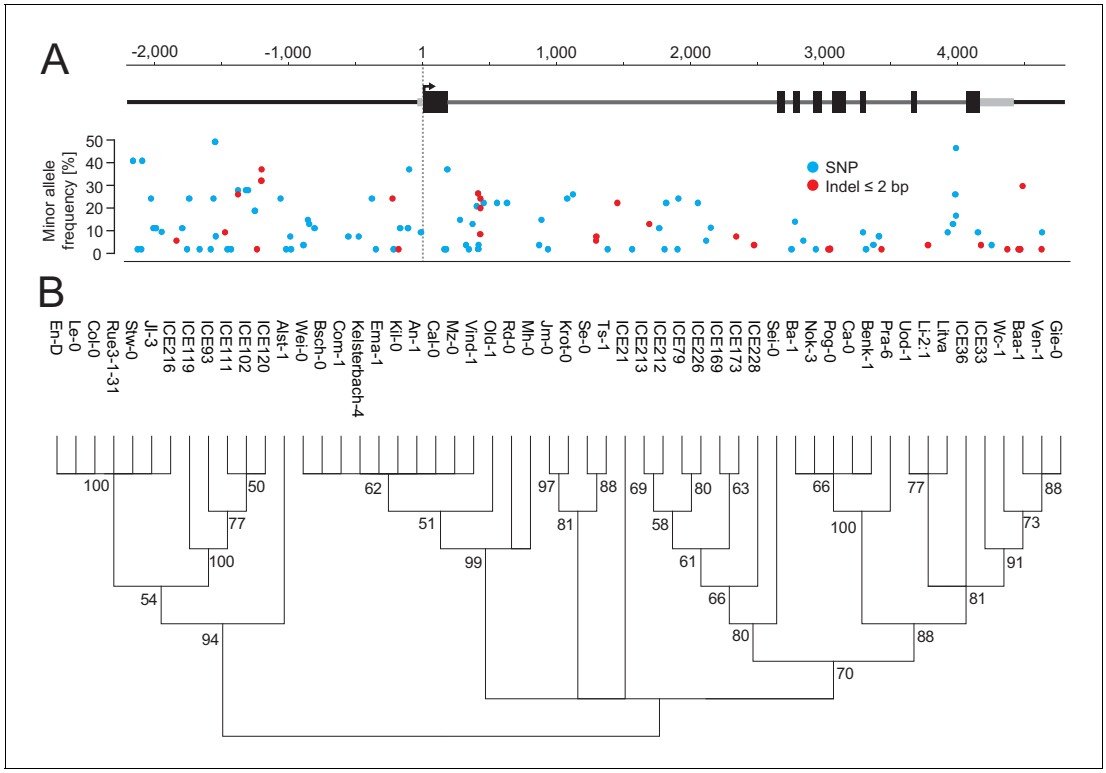

Figure 3—figure supplement 4. 119 polymorphic sites among the accessions of the *FLM* haplotypes set. (**A**) Position, frequency and type of each the 119 polymorphic sites among 54 accessions of the *FLM* haplotype set used for simple single locus association tests. (**B**) Neighbour-joining tree showing the genetic relationship among 54 accessions of the *FLM* haplotype set based on the 119 polymorphic sites shown in (**A**). Bootstrap values are indicated at each branch.

DOI: 10.7554/eLife.22114.010

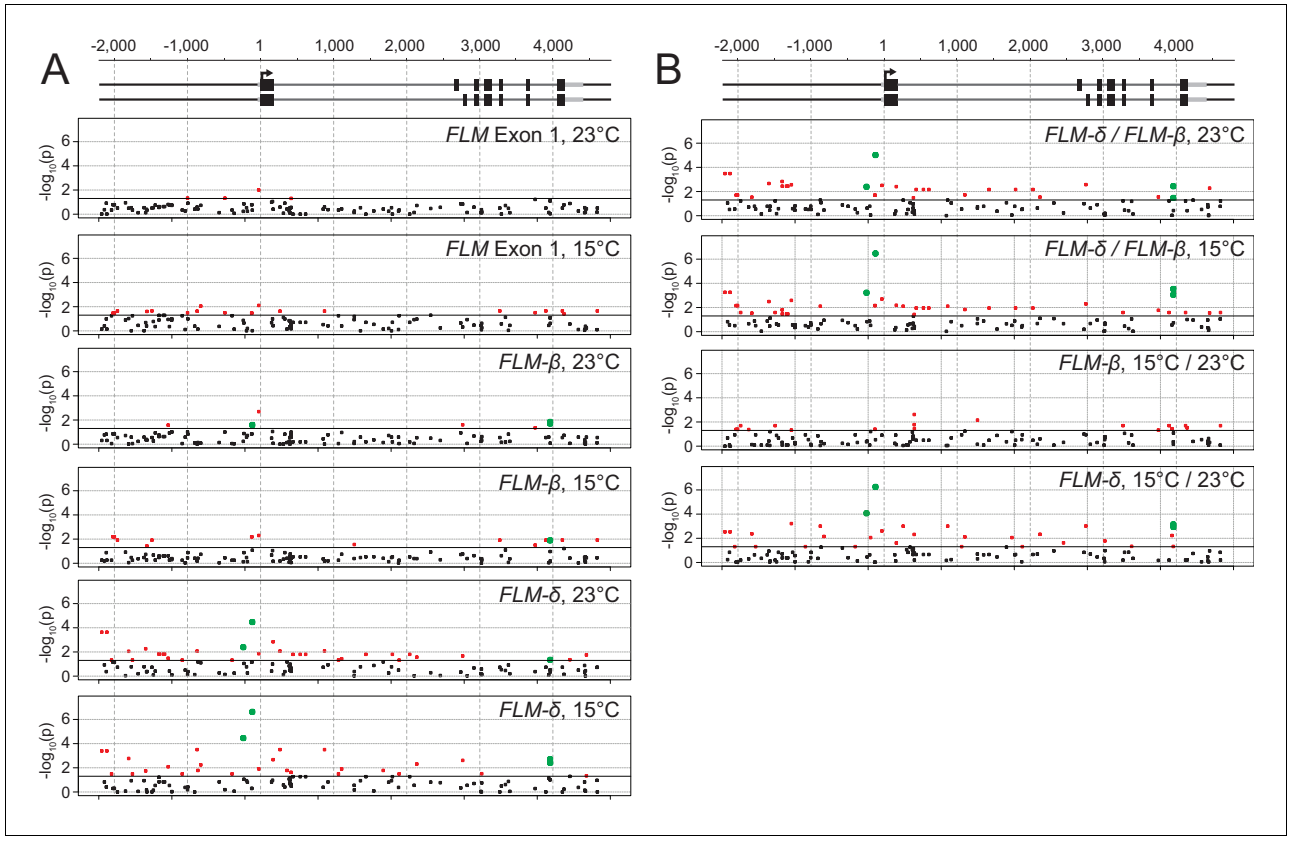

Figure 3—figure supplement 5. Results from the simple single locus association test between the 119 polymorphic sites with *FLM* expression values. (**A**) and (**B**) Representation of the –log(10)-transformed p-values of the simple single locus association tests of each of the 119 sites that were included in the analysis. The black horizontal line corresponds to a significance threshold of p=0.05. The black dots represent polymorphic sites that associate with p>0.05, the red dots those with p<0.05. The green dots represent sites with p<0.05 that were chosen for detailed experimental analysis as shown ***Figure 4A***. The *FLM* genomic gene model is as introduced in ***Figure 1***. Each graph represents the analysis of the indicated expression trait. (**A**) shows all comparisons with relative expression values and (**B**) shows all comparisons of relative changes of transcript levels between *FLM-β* and *FLM-δ* in response to changes in temperature (15°C to 23°C) or ratios between the transcript levels (*FLM-δ*/*FLM-β*) at 15°C and 23°C as a readout for the relative abundance of the two splice forms with hypothesized antagonistic functions. Quantitative values are summarized in ***Supplementary file 5***.
DOI: 10.7554/eLife.22114.011

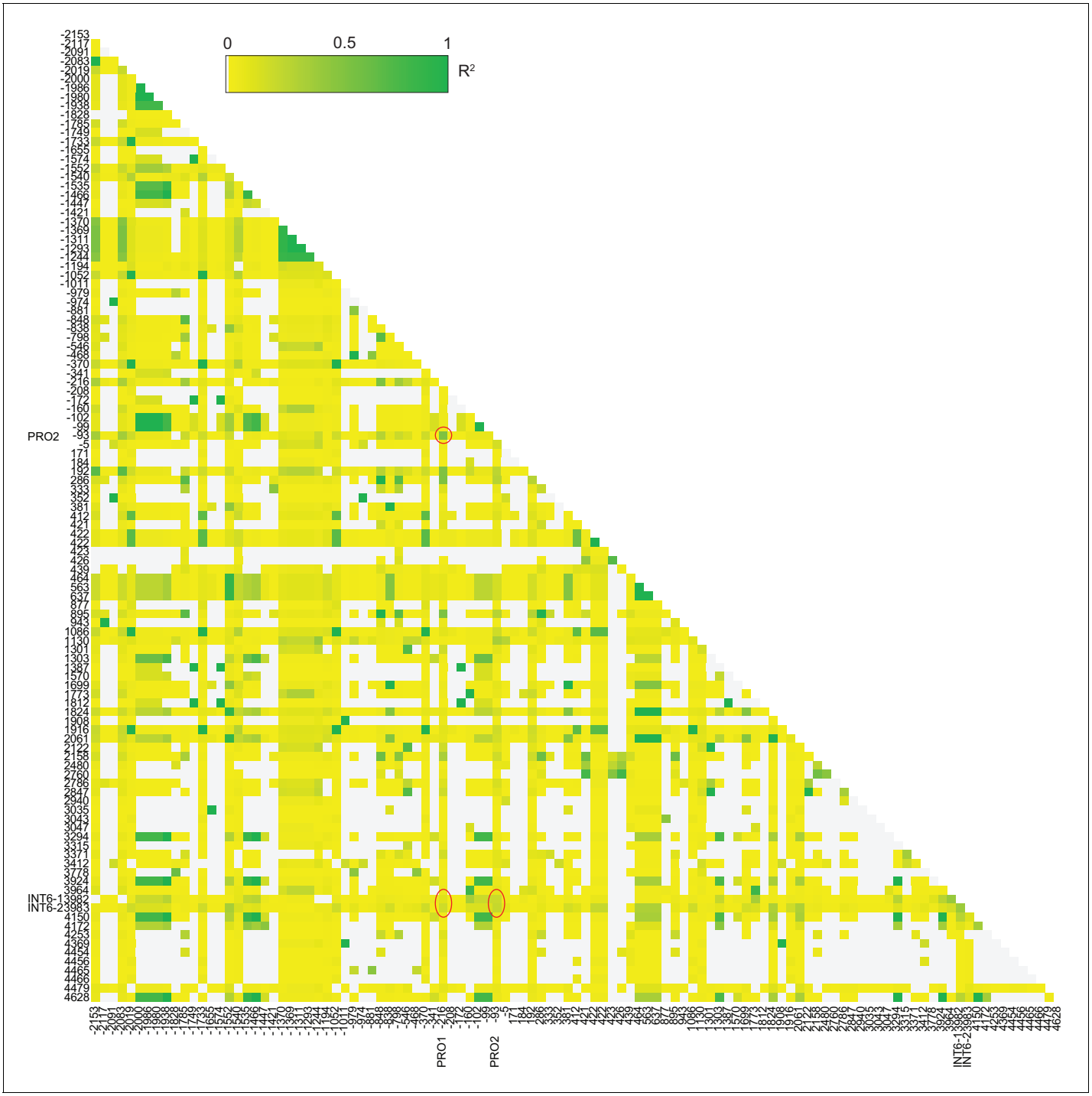

Figure 3—figure supplement 6. Linkage analysis for the simple single locus association test. Pairwise linkage analysis (R^2) of the sites used for the simple single locus association test among the 54 accessions of the *FLM* haplotype set. Only the 109 biallelic sites of the 119 sites are shown. Red circles show the sites PRO1, PRO2, and two biallelic SNPs of the nucleotide triplet INT6 (at bp +3975 and +3976). Position is set to 1 according to the A of the ATG start codon.

DOI: 10.7554/eLife.22114.012

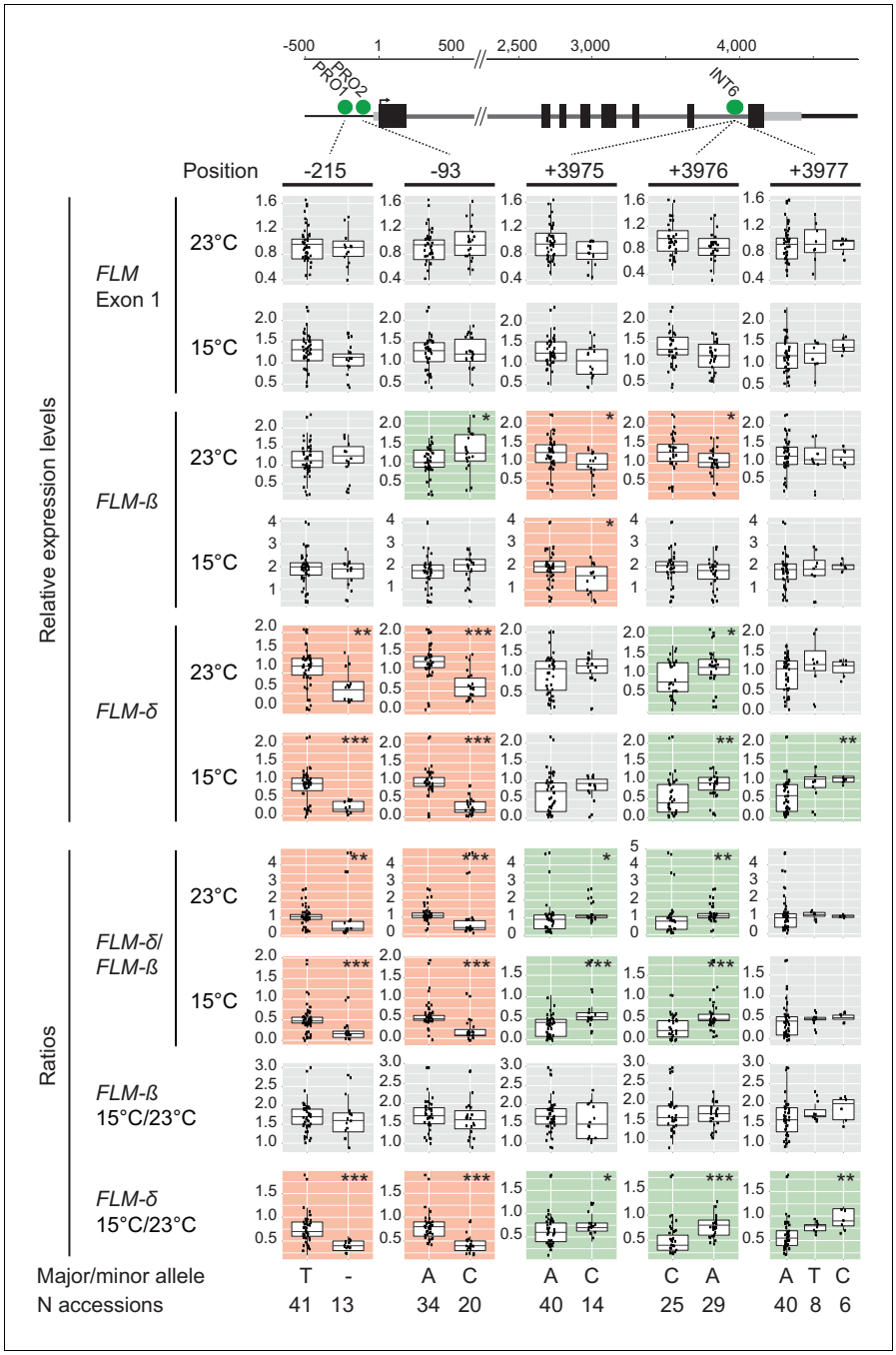

Figure 3—figure supplement 7. Effects of the PRO1, PRO2 and INT6 polymorphisms on *FLM* gene expression. Effects of the PRO1, PRO2 and INT6 alleles on relative expression levels and ratios of these expression values among the 54 accessions of the FLM haplotype set. The background color of each graph indicates that the minor allele associates with upregulation (green) and downregulation (red), respectively. p-values of the association tests are shown: *, p<0.05; **, p≤0.01; ***, p<0.001 as shown in each graph. Non-significant comparisons (p>0.05) are shown with a grey background. Single values are represented as jittered dots. Scale and numbering is set to 1 according to the A of the ATG start codon. Quantitative values are summarized in ***Supplementary file 5***.
DOI: 10.7554/eLife.22114.013

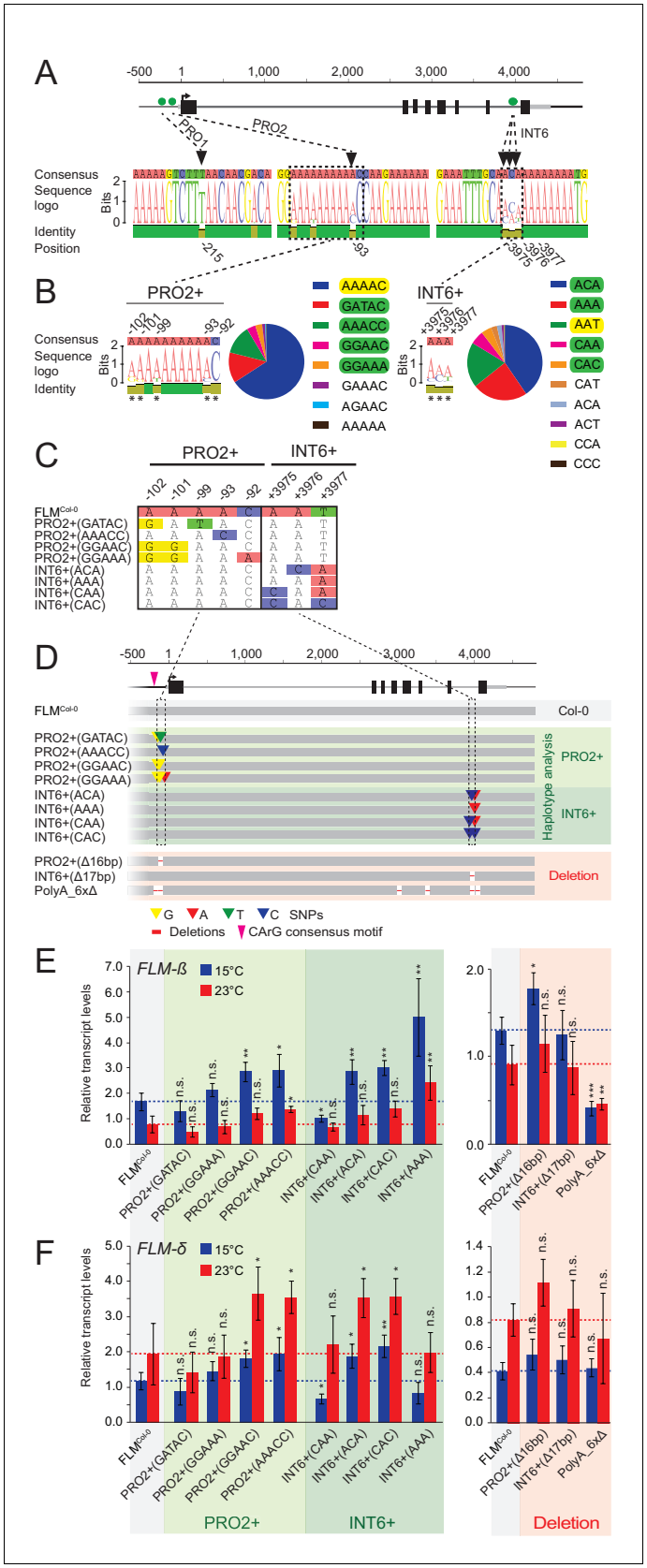

Figure 4. Polymorphisms in PRO2+ and INT6+ sites influence basal and temperature-dependent *FLM* expression. (**A**) Schematic representation of the *FLM* genomic locus as shown in ***Figure 1***. Green dots indicate the positions of

Figure 4 continued on next page

Figure 4 continued

the PRO1, PRO2, and INT6 sites that were chosen for further investigation. Sequence logos display the allele frequencies at these sites among the 54 accessions of the *FLM* haplotype set. (**B**) Sequence logos and allele frequency distribution of PRO2+ and INT6+ polymorphisms among 840 Arabidopsis accessions. All polymorphic residues are marked with asterisks and the respective allele frequencies are indicated by the sequence logo. The Col-0 reference haplotype is marked in yellow, haplotypes chosen for further investigation are marked in green. (**C**) and (**D**) Alignment of the PRO2+ and INT6+ polymorphisms (**C**) and schematic representation of the FLM^{Col-0} reference construct (pFLM::gFLM) as well as the FLM^{Col-0} variants selected for transgenic analysis in the *FLM* deletion accession Nd-1. Bases that differ between FLM^{Col-0} and the variants are coloured. Deletions in the deletion constructs of the PRO2+ and INT6+ sites as well as the PolyA motifs are displayed with red lines (not drawn to scale). (**E**) and (**F**) Mean and SD (four replicate pools with four to eleven independent T2 transgenic lines) of qRT-PCR analyses of *FLM-β* (**E**) and *FLM-δ* (**F**). For easier comparison, the values obtained with FLM^{Col-0} are displayed as red and blue dotted lines. Student's t-tests: *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; n.s., not significant.

DOI: 10.7554/eLife.22114.014

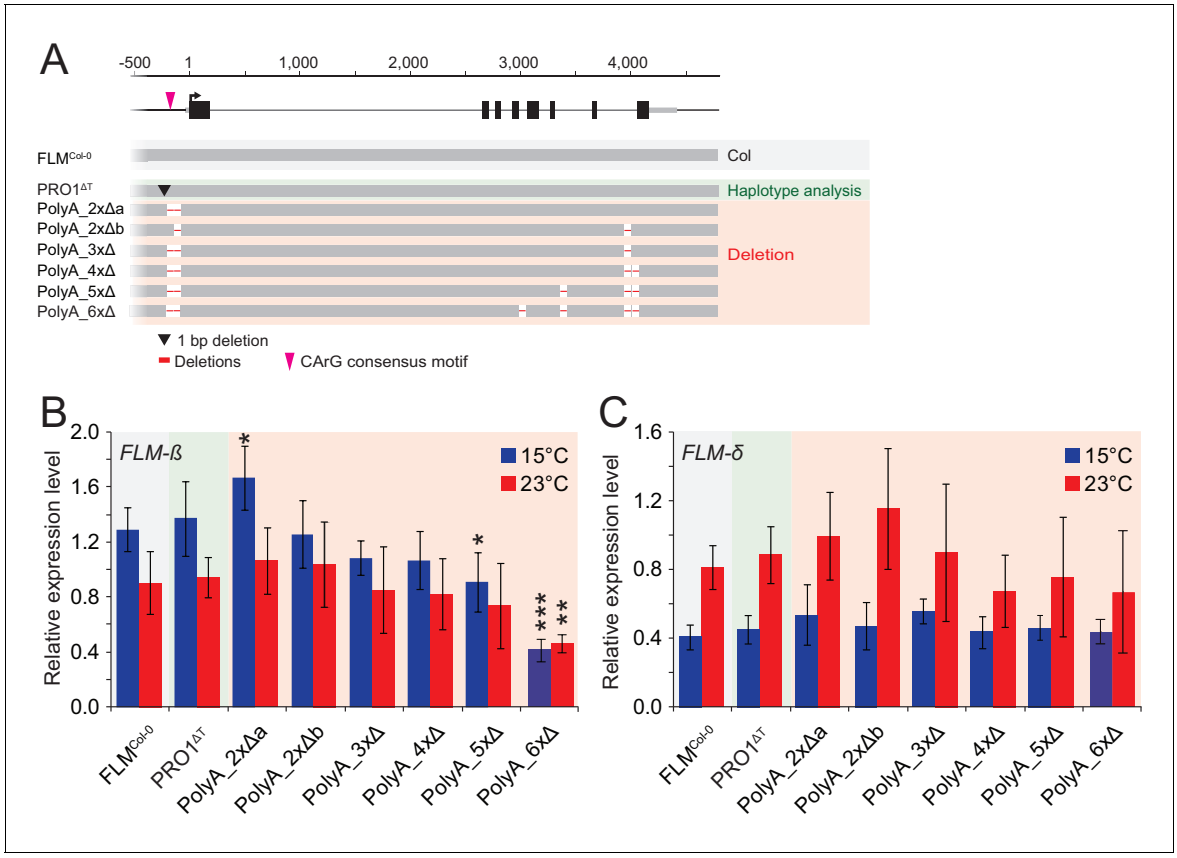

Figure 4—figure supplement 1. Increasing the number of PolyA motif deletions results in the gradual reduction of *FLM-ß* expression. (**A**) Schematic representation of the FLM^{Col-0} reference construct (pFLM::gFLM) as well as the *FLM* PolyA deletion variants selected for transgenic analysis in Nd-1. Deletions of the PolyA motifs are displayed with red lines (not drawn to scale). (**B**) and (**C**) Mean and SD (four replicate pools comprising between five and ten independent T2 transgenic lines) from qRT-PCR analyses of *FLM-β* (**B**) and *FLM-δ* (**C**) expression in ten day-old seedlings. Student's t-test: *p≤0.05; **p≤0.01; ***p≤0.001; otherwise not significant.

DOI: 10.7554/eLife.22114.015

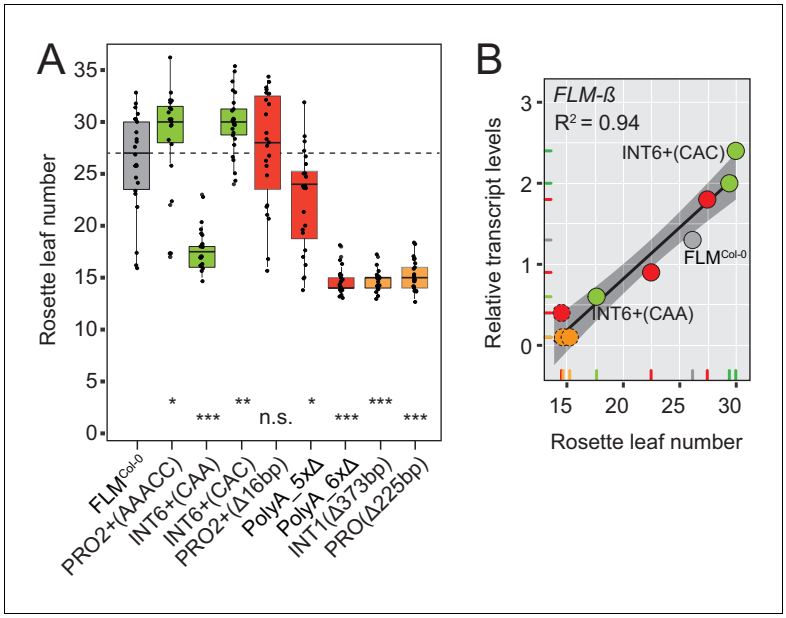

Figure 5. *FLM-ß* expression shows high correlation with flowering time of transgenic plants. (**A**) Box plot of quantitative flowering time analysis (rosette leaf number) of independent T2 transgenic lines at 15˚C and long day photoperiod. Ten plants of three replicate pools were analysed for each construct. The data were corrected for the anticipated 25% of non-transgenic Nd-1 segregants (see ***Figure 5—figure supplement 1D*** for the uncorrected analysis). Single values are shown as jittered dots, the colour represents the type of variant as introduced in ***Figure 4D***, the median of the FLM^{Col-0} reference is indicated as dotted line. Wilcoxon rank test: *p≤0.05; **p≤0.01; ***p≤0.001; n.s., not significant. (**B**) Correlation (simple linear regression) between *FLM-ß* and qRT-PCR expression data as presented in ***Figure 5—figure supplement 1A,B*** and flowering time analysis as shown in (**A**). Datapoints of INT6+ variants with contrasting *FLM* expression and FLM^{Col-0} are designated. The color code corresponds to (**A**). The variants PolyA_6xΔ, $INT1^{\Delta 373bp}$, and $PRO^{\Delta 225bp}$, which do not respond in a linear manner are shown as dotted circles. The shaded areas indicate the 95% confidence intervals; p<0.0001. Note, that the flowering time data shown in (**A**) were corrected by removing values of non-transgenic T2 segregants. An analysis using the uncorrected data is shown in ***Figure 5—figure supplement 1E,F***.

DOI: 10.7554/eLife.22114.016

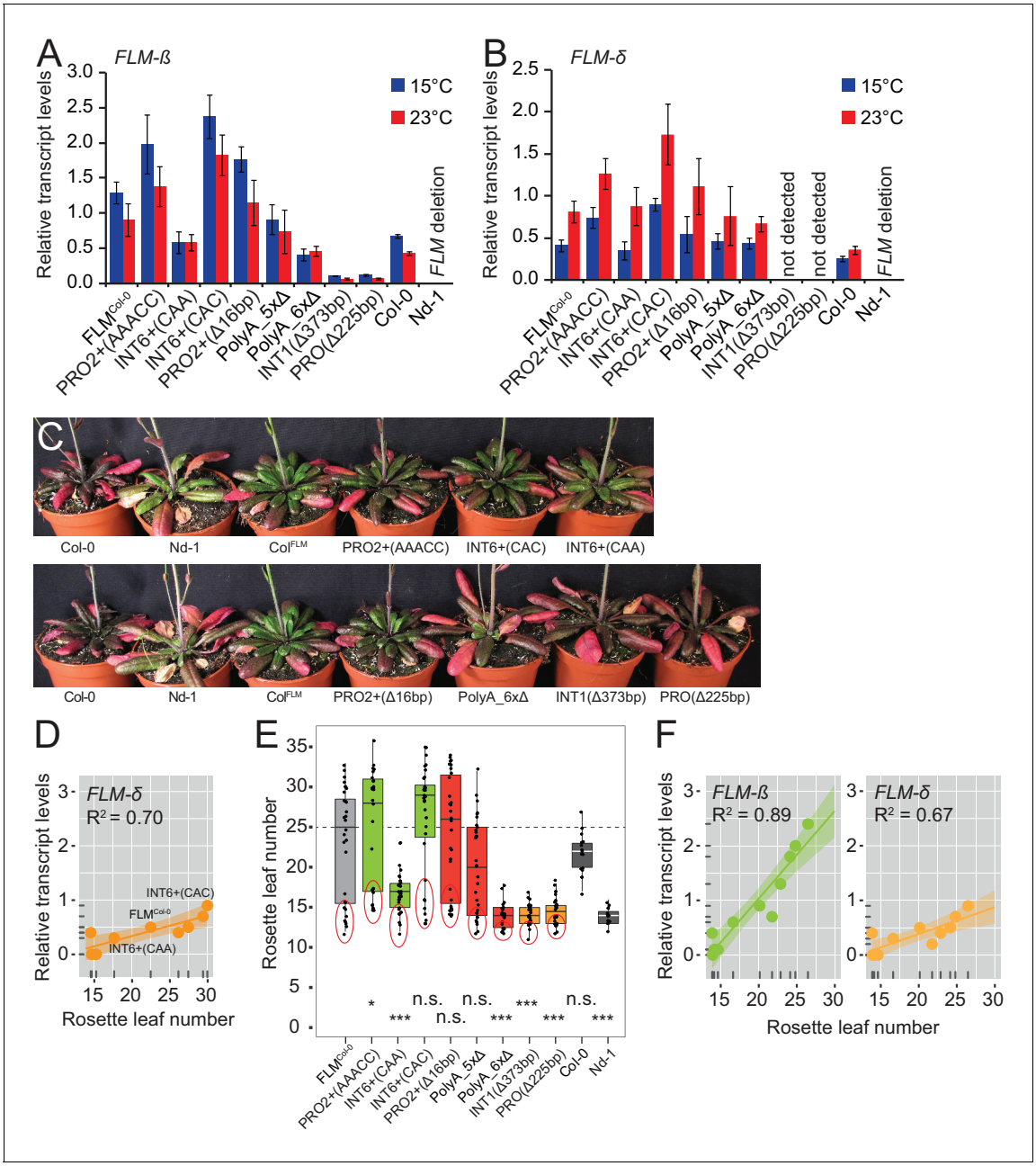

Figure 5—figure supplement 1. *FLM-ß* expression correlates with flowering time in transgenic lines carrying different *FLM* alleles. (**A**) and (**B**) Results from qRT-PCR analyses of *FLM-β* (A) and *FLM-δ* (B) expression in ten day-old seedlings of pools (n = 18–40) of independent T2 transgenic lines. Error bars indicate SD of four replicate pools comprising between four and ten transgenic lines or three Col-0 replicates. (**C**) Photographs of representative 57 day-old plants expressing $FLM^{Col\text{-}0}$ and PRO2+ and INT6+ variants grown at 15°C in long-day photoperiod. Col-0 and Nd-1, which served as genetic background for the transgenic analysis, are shown in each photograph for comparison. Plants of both pictures are from the same experiment but were photographed separately. (**D**) Correlation (simple linear regression) between *FLM-δ* expression data as presented in (B) and flowering time as shown in ***Figure 5A***. Data points of INT6+ variants with contrasting *FLM* expression and $FLM^{Col\text{-}0}$ are designated. p=0.005. The shaded areas indicate the 95% confidence intervals. (**E**) Box plot of quantitative flowering time analysis (rosette leaf number) of independent T2 transgenic lines at 15°C and long day photoperiod. Ten plants of three replicate pools were analysed for each construct (uncorrected data). Single values are shown as jittered dots, the colour represents the type of variant as introduced in ***Figure 4D***, the median of the $FLM^{Col\text{-}0}$ reference is indicated as dotted line. The 25% of the plants that show a very early flowering that most likely represent the wild type Nd-1 segregants are indicated with a red circle and these were extracted for correction as presented in ***Figure 5A***. Wilcoxon rank test: *p≤0.05; **p≤0.01; ***p≤0.001; n.s., not significant. (**F**) Correlation (simple linear regression) between *FLM-ß* and *FLM-δ*. qRT-PCR expression data as presented in (**A**) and (**B**) and flowering time analysis as shown in (**D**). *FLM-ß*, p<0.0001; *FLM-δ*, p=0.002. The shaded areas indicate the 95% confidence intervals.

Figure 5—figure supplement 1 continued on next page

Figure 5—figure supplement 1 continued

DOI: 10.7554/eLife.22114.017

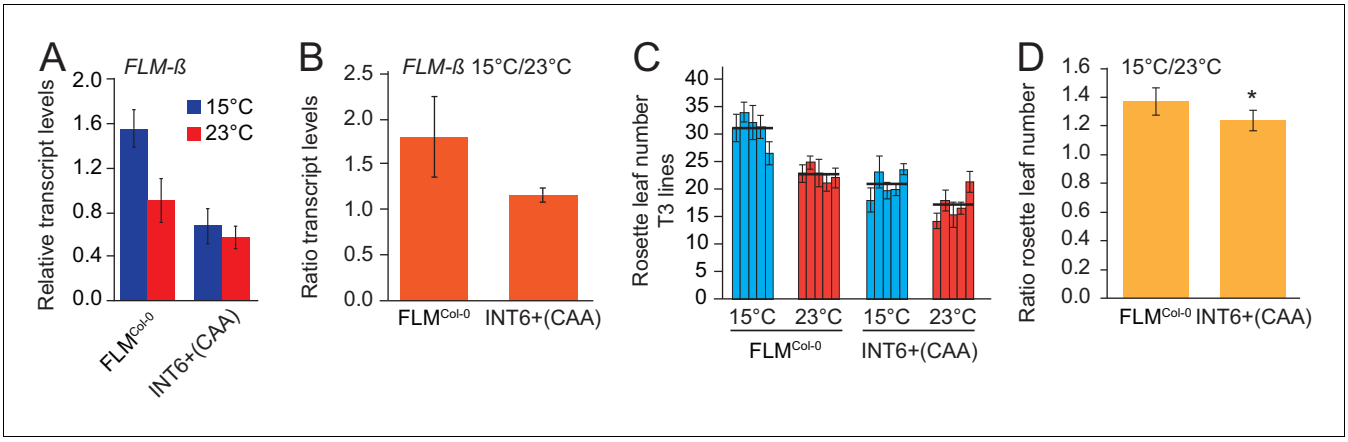

Figure 6. The INT6+^{CAA} polymorphism reduces temperature-sensitivity of *FLM* expression and flowering. (**A**) qRT-PCR analysis and (**B**) expression ratios of *FLM-β* of FLM^{Col-0} (INT6+^{AAT}) and INT6+^{CAA} grown in 15°C and 23°C. Statistical tests of temperature-sensitivity are described in the text. (**C**) Means and SD of quantitative flowering time analysis (rosette leaf number). n = 5 (15°C) and 8 (23°C) replicates from five independent homozygous T3 transgenic lines grown at 15°C and 23°C in long day photoperiod. (**D**) Mean and SD of ratios of rosette leaf numbers from the analysis shown in (**C**). Student's t-test: *p<0.05; **p≤0.01.

DOI: 10.7554/eLife.22114.018

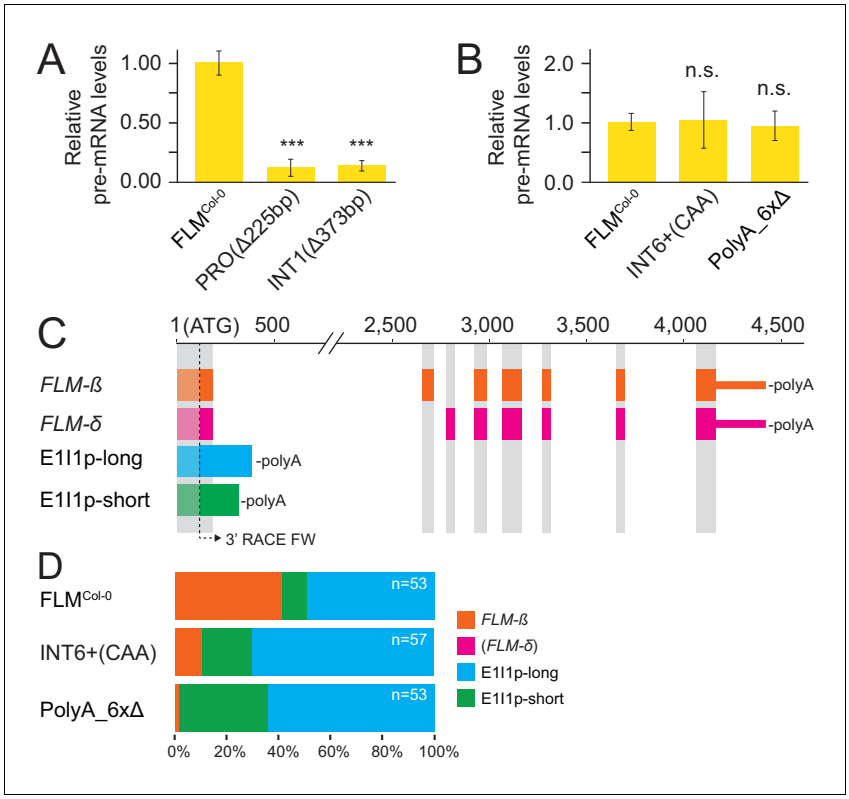

Figure 7. *FLM* polymorphisms affect *FLM* transcription or splicing at the expense of *FLM-ß*. (**A**) and (**B**) Mean and SD (n = 3) of *FLM* pre-mRNA levels. Student's t-test: ***p≤0.001; n.s. = not significant. (**C**) Schematic representation of *FLM* cDNAs detected more than once (n = 163). *FLM-δ* transcripts were not detected and are only shown for completeness. Grey areas correspond to *FLM* exons of the *FLM-ß* and *FLM-δ* gene models as specified in ***Figure 1A***. The arrow indicates the position of the 3' RACE primer used in combination with an oligo (dT) reverse primer to detect polyadenylated transcripts. (**D**) Frequency distribution of the transcripts displayed in (**B**) with total number of sequences depicted in the graph.

DOI: 10.7554/eLife.22114.019

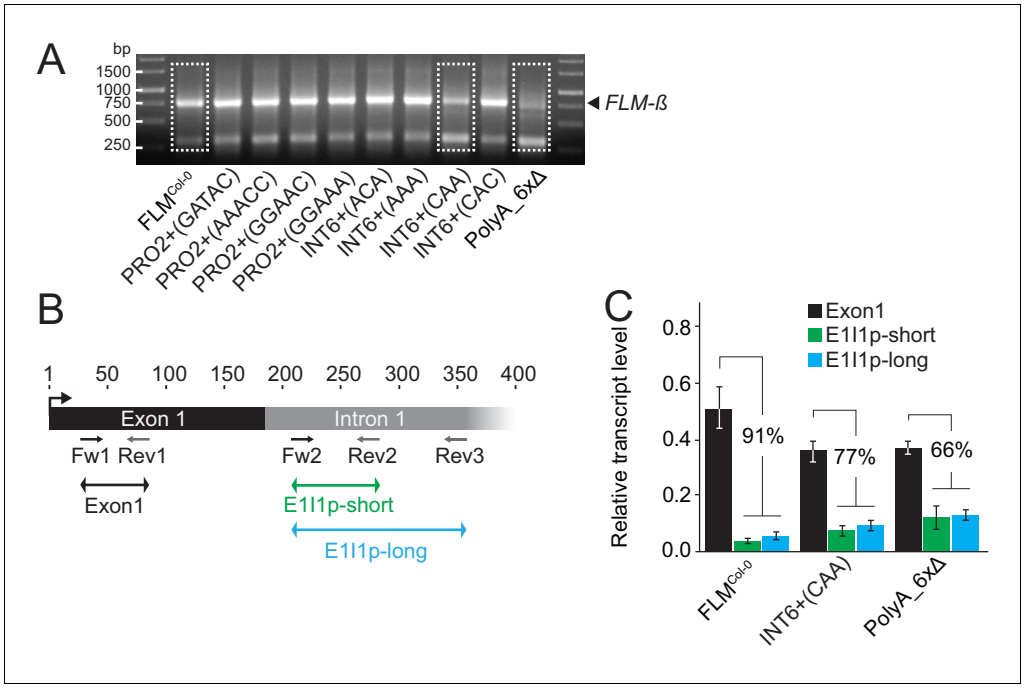

Figure 7—figure supplement 1. Transgenic lines expressing different *FLM* alleles display differential *FLM* expression and splicing patterns. (**A**) Image of an agarose gel with the results of a 3'-RACE PCR. The primer used for 3' amplification is indicated in ***Figure 7C***. The gel portions from the FLM^{Col-0}, INT6+^{CAA} and PolyA_6x**Δ** amplification products framed with a dotted line were purified, cloned and subjected to DNA sequencing. (**B**) Schematic representation of the primers used for the qRT-PCR quantification of total *FLM* levels (exon 1) and the E1l1p intron 1 sequence-containing transcripts as indicated by arrows below the gene model. The black box corresponds to *FLM* exon 1, the grey box to the proximal portion of intron 1. Numbering is respective to the A of the ATG start codon. (**C**) Mean and SD of qRT-PCR analyses of intron 1 sequence-containing transcripts from four replicate pools of independent T2 lines. The relative decrease in E1l1p expression values compared to exon 1 transcript is indicated for each line in the graph. Note that the comparison of the expression values between the fragments are approximations since the primer efficiencies may not be identical. Primer sequences are listed in ***Supplementary file 7***.

DOI: 10.7554/eLife.22114.020

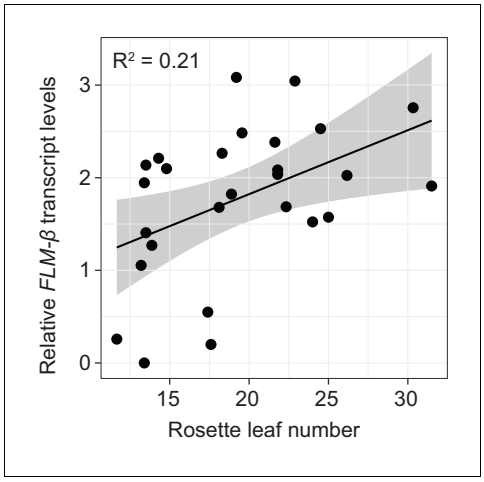

Figure 8. *FLM-ß* expression shows high correlation with flowering time in Arabidopsis accessions. (**A**) Correlation (simple linear regression) of *FLM-ß* transcript levels detected in transgenic plants and natural accessions with identical PRO2+ and INT6+ haplotypes. When the contribution of individual haplotypes was assessed, PRO2+^{AAAAC} INT6+^{CAA} showed an exceptional response in accessions and was excluded. ***Figure 8—figure supplement 1E*** shows the complete analysis. Horizontal and vertical error bars depict SD of four replicate transgenic line pools or SD of the accessions. Data of the transgenic lines was taken from ***Figure 4E***. The grey area indicates the 95% confidence interval. (**B**) Correlation of *FLM-ß* transcript levels with flowering time (n = 8–10) as measured in rosette leaf number of summer-annual accessions. p=0.011.

DOI: 10.7554/eLife.22114.021

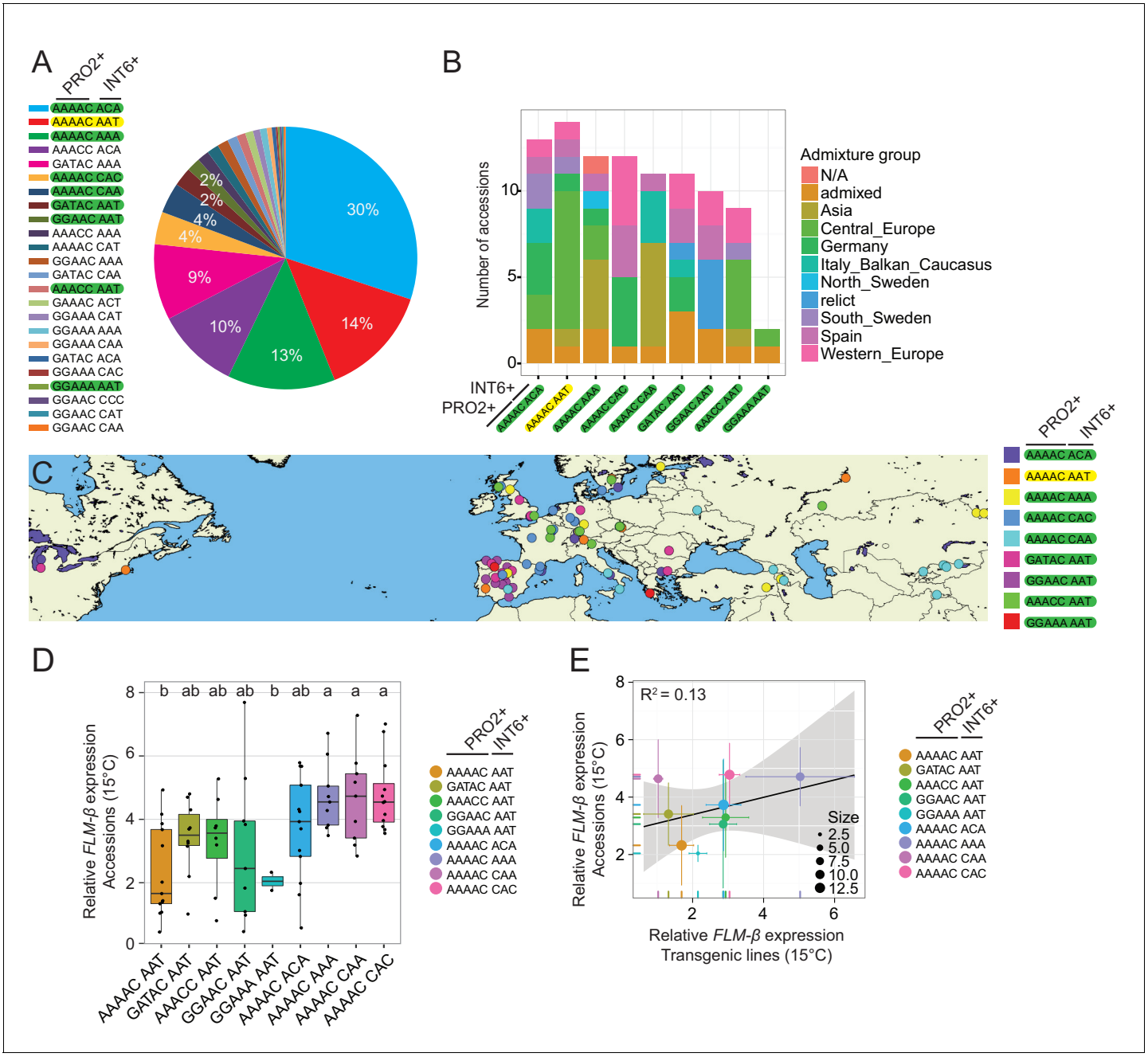

Figure 8—figure supplement 1. Geographic and genetic distribution of the PRO2+/INT6+ haplotypes among 840 accessions. (**A**) Pie chart with the relative distribution of the PRO2+/INT6+ haplotypes among 840 accessions. The Col-0 haplotype is marked in yellow, haplotypes analyzed as part of this study are marked in green. Haplotypes only represented by a single accession are not included. (**B**) Number of accessions selected from each of the nine PRO2+/INT6+ groups (average 10, in total 94 accessions). The color code indicates the genetic group membership (k = 9) of every accession as retrieved from the 1001 Genomes ADMIXTURE tool (http://tools.1001genomes.org/) (***The 1001 Genomes Consortium, 2016***). (**C**) Geographic distribution of the accessions shown in (**B**). (**D**) *FLM-β* transcript levels of accessions shown in (**B**) and (**C**) grouped by the PRO2+/INT6+ haplotype. Similar letters indicate no significant difference of total leaf number (Tukey HSD, $p<0.05$). (**E**) Correlation (simple linear regression) between *FLM-β* transcript levels as detected in the transgenic analysis and in accessions with identical PRO2+/INT6+ haplotypes from plants grown at 15°C. The PRO2^{+AAAAC} INT6+^{CAA} haplotype, which was removed in ***Figure 8A***, is shown and highlighted with a red arrow. Circle sizes represent the number of accessions analysed of each group as specified in the legend. Horizontal and vertical error bars depict SD of four replicate transgenic line pools or SD of the accessions of each group. Data for the transgenic line analysis were taken from ***Figure 4E***. The grey area indicates the 95% confidence interval.
DOI: 10.7554/eLife.22114.022

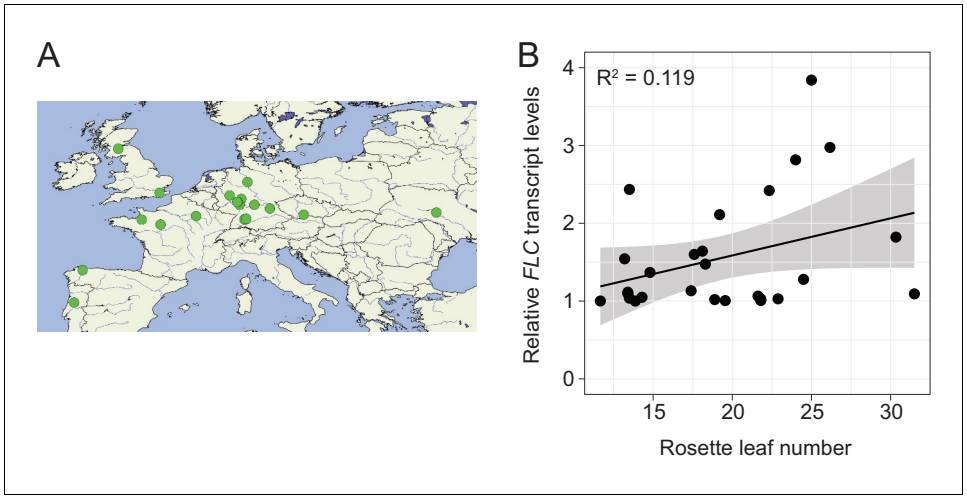

Figure 8—figure supplement 2. Geographic distribution of summer-annual accessions. (**A**) Geographic distribution of the 27 summer-annual accessions as described in ***Figure 8B***. (**B**) Correlation of *FLC* transcript levels with flowering time data (n = 8–10) as measured in rosette leaf number of summer-annual accessions. p>0.05.
DOI: 10.7554/eLife.22114.023

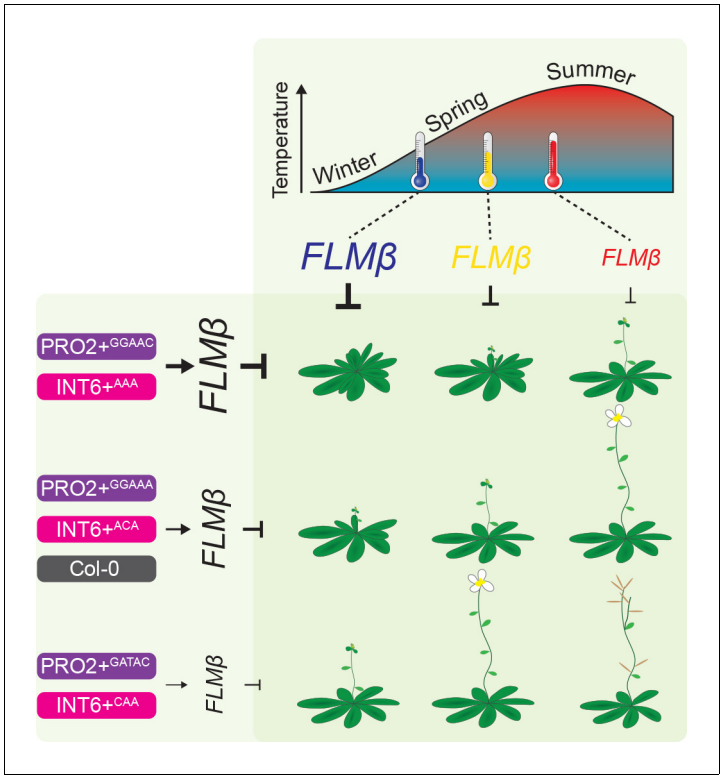

Figure 9. Model of the proposed role of PRO2+ and INT6+ haplotypes and temperature on *FLM-ß* abundance and flowering. The abundance of the flowering repressor *FLM-ß* decreases in response to higher temperature and flowering is consequently accelerated (***Figures 1B*** and ***4E***) (***Lee et al., 2013***; ***Posé et al., 2013***; ***Lutz et al., 2015***). Note, that previous studies showed an especially prominent effect of *FLM* in a range from 9°C to 21°C (long-day photoperiod) (***Lee et al., 2013***; ***Posé et al., 2013***; ***Lutz et al., 2015***). At the genetic level, *FLM-ß* abundance is triggered by the PRO2+ (purple) and/or the INT6+ (pink) haplotype and flowering time correlated to *FLM-ß* abundance (***Figure 4E***). Among the PRO2+/INT6+ combinations tested, the Col-0 (grey) reference allele (PRO2 $+^{AAAAC}$/INT6$+^{AAT}$) showed intermediate *FLM-ß* levels (***Figure 4E***). We suggest, that changes in flowering time due to changing ambient temperature can be precisely buffered by modifying the PRO2+ and INT6+ regions, as illustrated by similar plant symbols.

DOI: 10.7554/eLife.22114.024