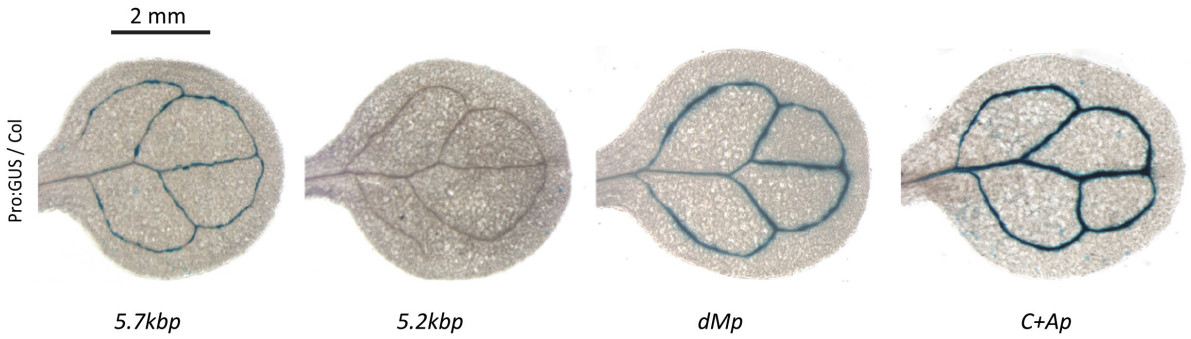
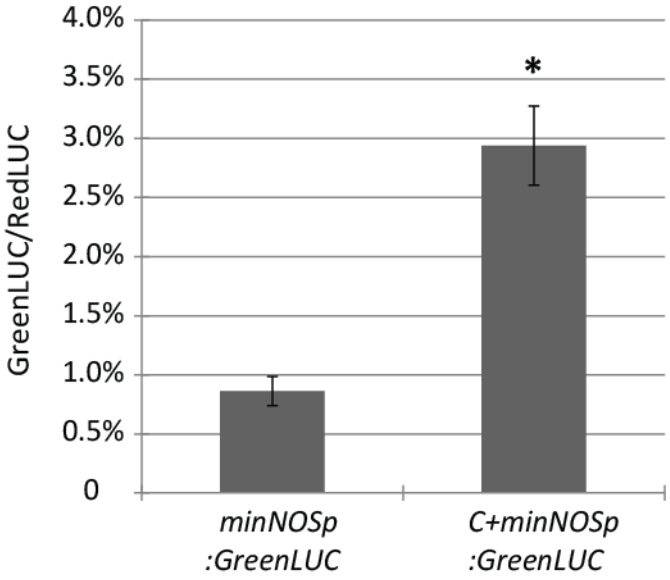
Supplementary Figures and Tables

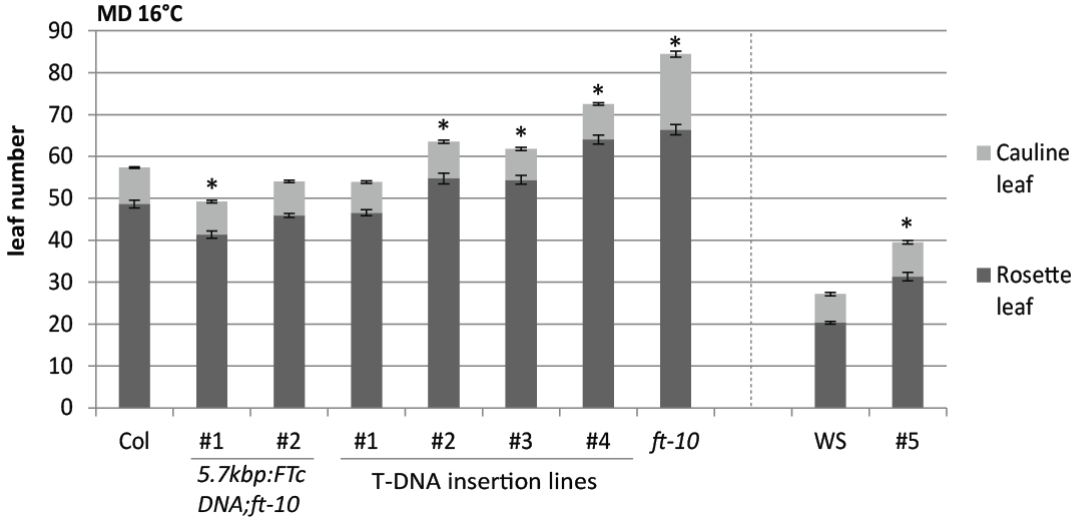


Supplementary Figure 1. Histochemical localization of GUS activity in cotyledons of transgenic Col-0 expressing the *GUS* reporter gene under the control of *FT* promoters as indicated. Leaves were collected at ZT 16 after growth for 12 LDs on GM media.



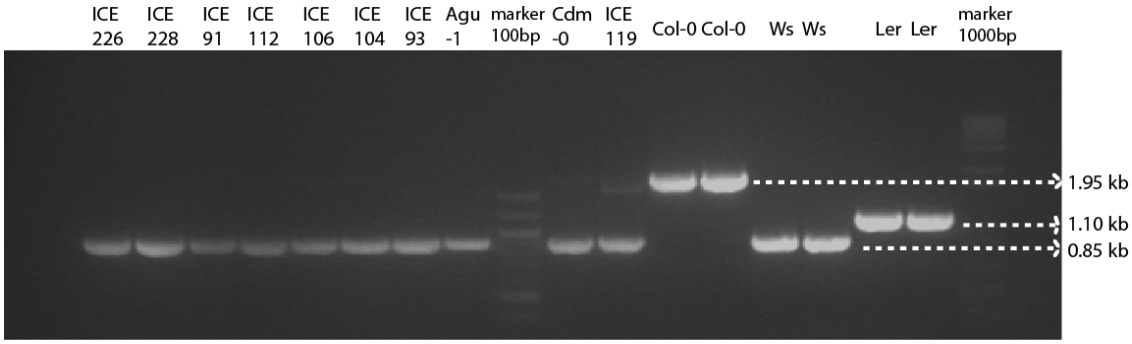
Supplementary Figure 2. *Block C* enhances a minimal promoter in transient assays

Block C fused to *NOSminp* drives reporter gene *GreenLUC* expression in transient leaf bombardment assays. *35S::RedLUC* co-bombarded with *promoter::GreenLUC* was used as the internal control. Values represent mean ± SE. The asterisk (*****) indicates statistically significant differences relative to the control *NOSmin::GreenLUC*. Statistical significance was determined using the Student's t-test (p<0.01).

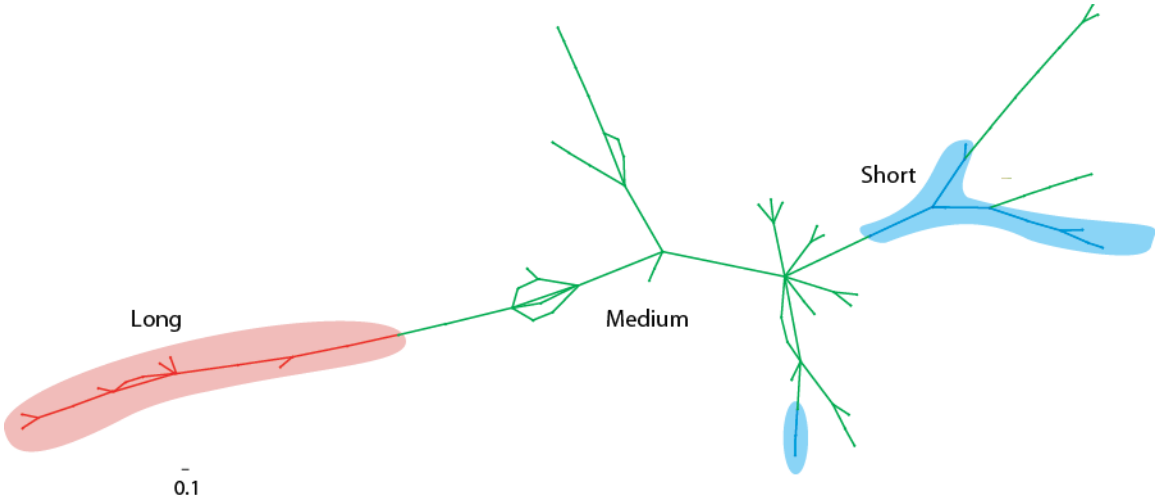


Supplementary Figure 3. Flowering time of T-DNA insertion lines under MDs conditions.

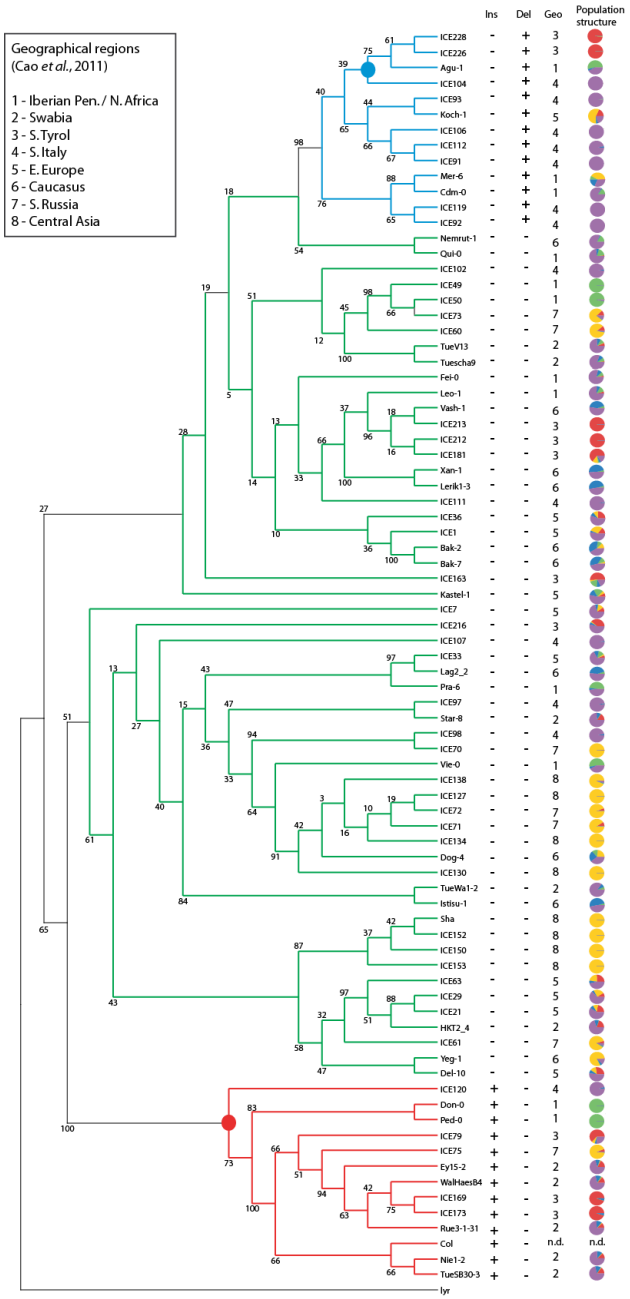
Flowering time was measured in marginal long day (MD) grown Arabidopsis plants carrying T-DNA insertions either upstream (#1) of *Block C* or between *Block C* and the *Block A* (#2-#5) at positions indicated by triangles in Fig.2a. Line #5 is Ws and all other lines in the Col-0 accession. Number of rosette and cauline leaves are shown as mean ± SE. The asterisk (*****) indicates statistically significant differences relative to the control Col or Ws as indicated. Statistical significance was determined using the Student's t-test (p<0.01).



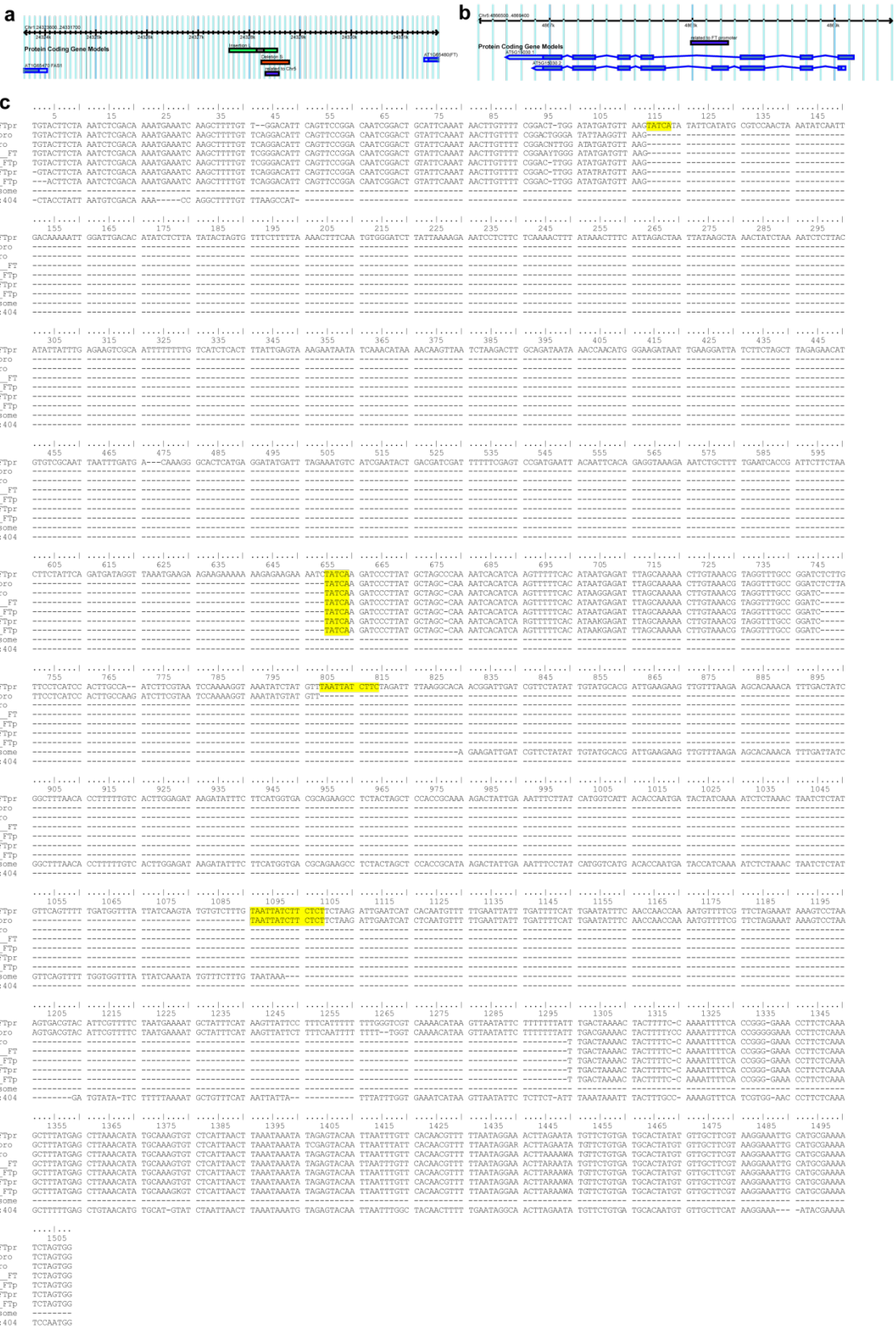
Supplementary Figure 4. Three types of *FT* promoters. DNA was isolated from 13 Arabidopsis accessions and amplified by primer pair 12_forward and 13_reverse. The expected product 1.95 kb, 1.10 kb and 0.85 kb for "long", "medium" and "short" promoters, respectively. PCR products were loaded on 1% agarose gel.



Supplementary Figure 5. Median-joining network of the *FT* haplotypes. The blue, green and red colour of the branches and shading corresponds, respectively, to the clades of the short, medium and long *FT* promoter regions. Scale bar, 0.1 nucleotide substitutions per site.



Supplementary Figure 6. SNP-based maximum-likelihood phylogeny of the coding and promoter regions of *FT* from 81 Arabidopsis accessions. Blue, green and red colour of the branches indicates clades of the short, medium and long *FT* promoters, respectively. The *FT* sequence from *A. lyrata* was used as outgroup. Bootstrap values (%) are shown at the nodes. Filled ellipse pictograms denote sub-clades with the footprints of selection defined as the retention of the phylogenetically close *FT* haplotypes within the genetically divergent genotypes. "Ins", "Del" and "Geo" columns describe, respectively, distribution of the insertion and deletion in the *FT* promoter region (see Fig. 4a) and geographical origin of the genotypes. Coloured pie charts illustrate population structure of the genotypes determined by genome-wide analysis[1]. n.d. means no data.



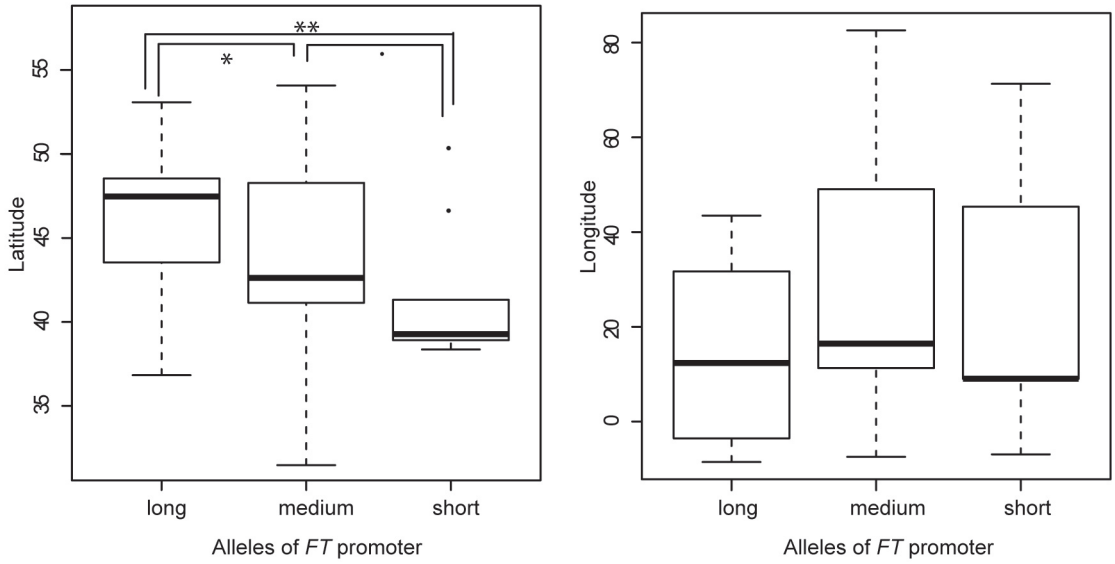
Supplementary Figure 7. Structural variation at the *FT* promoter

a, location and structure of "long", "medium" and "short" FT promoter types. The long *FT* promoter type, corresponding to the reference genome Col, contains two insertions (green

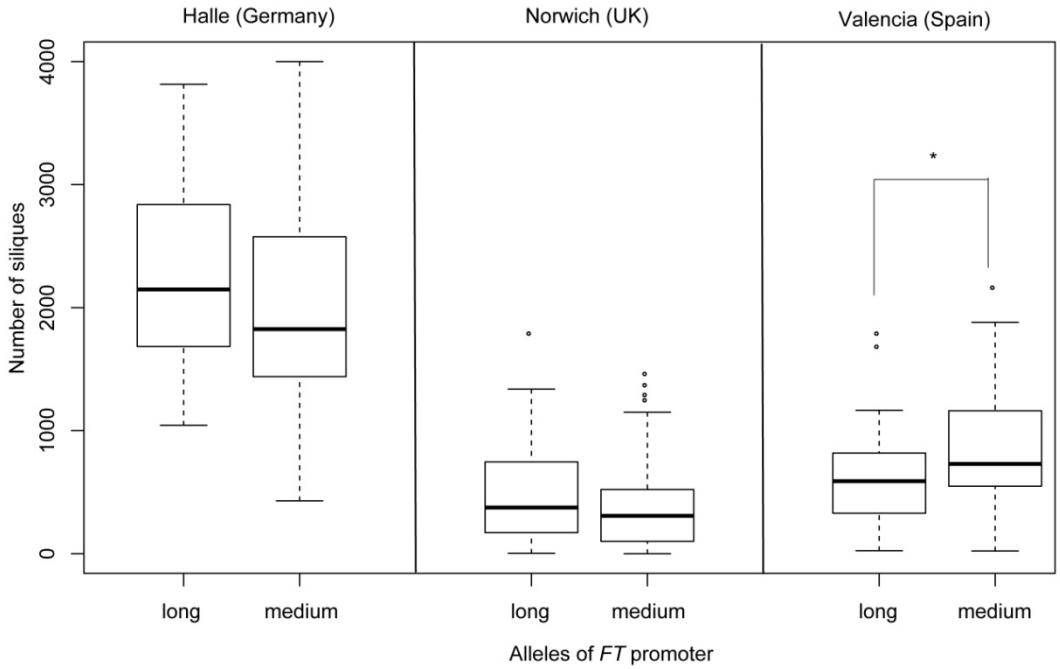
boxes) interrupted by a regions shared with the “medium” type (gray box). The short promoter does not contain any of the additional sequence found in the “long” promoter and carries a further deletion of the sequence shared between “long” and “medium” (red box). The proximal half of the insertion is related to a sequence on chromosome 5 (blue box, 97% sequence identity).

b, location of the duplicated fragment from the *FT* promoter on chromosome 5.

c, alignment of polymorphic region between “long”, “medium” and “short” *FT* promoter types. Yelow boxes indicate the presumed target site insertions in Col-0. The *FT* homologous sequence from chromosome 5 and the Arabidopsis lyrate sequence is included in the alignment.



Supplementary Figure 8. Association of variation in the *FT* promoter with geographic coordinates. Boxplots of the latitudinal and longitudinal distribution of three *FT* promoter types as indicated.[1] Significance codes: p<0.01=**, p<0.05=* p<0.1 =. Long n=12, medium n=55, short n=13. See Supplementary table 2 and Supplementary dataset 1 for statistical tests.



Supplementary Figure 9. Variation in *FT* promoter length explains fitness variation of Arabidopsis accessions grown in Spain but not Halle or Norwich. Common garden experiments performed by Fournier-Level et al. 2011[2] at four different European sites were re-analyzed for 82 accessions for which the *FT* promoter type was confirmed by PCR and for which high density SNP data were available[3]. The medium version of *FT* promoter slightly outperforms the long version at the Spanish site based on the number of siliques produced by plant but not at two other sites (See Supplementary table 5 and Supplementary dataset 3 for statistical tests). Note that the short *FT* promoter type is intentionally removed from the graphs as only three accessions showing this genotype had been included in the study. Data from an experiment performed in Oulu, Finland were excluded from the analysis because most plants did not survive to seed set.

Supplementary Table 1

Analysis of differences in latitudinal and longitudinal distribution considering population structure k=5 determined by STRUCTURE				
## Associations with latitude				
glm(formula = latitude ~ FTprom + IAP1 + IAP2 + IAP3 + IAP4)				
Compare to medium promoter				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	43.055	1.046	41.155	< 2e-16 ***
FT long	2.777	1.284	2.163	0.033829 *
FT short	-2.232	1.325	-1.684	0.096389 .
IAP1	4.735	1.785	2.653	0.009782 **
IAP2	5.585	1.461	3.824	0.000274 ***
IAP3	-7.603	2.089	-3.639	0.000508 ***
IAP4	-4.162	4.248	-0.980	0.330414
Compare to long promoter				
(Intercept)	45.832	1.422	32.229	< 2e-16 ***
FT medium	-2.777	1.284	-2.163	0.033839 *
FT short	-5.010	1.604	-3.123	0.002568 **
IAP1	4.735	1.785	2.653	0.009782 **
IAP2	5.585	1.461	3.824	0.000274 ***
IAP3	-7.603	2.089	-3.639	0.000508 ***
IAP4	-4.162	4.248	-0.980	0.330415
Associations with longitude				
glm(formula = longitude ~ FTprom + IAP1 + IAP2 + IAP3 + IAP4)				
compare to medium promoter				
(Intercept)	6.4171	2.2604	2.839	0.00586 **
FT long	-3.4560	2.7747	-1.246	0.21692
FT short	2.3017	2.8637	0.804	0.42414
IAP1	0.5146	3.8562	0.133	0.89420
IAP2	61.8686	3.1557	19.605	< 2e-16 ***
IAP3	-14.9889	4.5143	-3.320	0.00141 **
IAP4	86.1141	9.1774	9.383	3.57e-14 ***
Compare to long promoter				
(Intercept)	2.9612	3.0725	0.964	0.33835
FT medium	3.4560	2.7747	1.246	0.21692
FT short	5.7577	3.4663	1.661	0.10099
IAP1	0.5146	3.8562	0.133	0.89420
IAP2	61.8686	3.1557	19.605	< 2e-16 ***
IAP3	-14.9889	4.5143	-3.320	0.00141 **
IAP4	86.1141	9.1774	9.383	3.57e-14 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Supplementary Table 1. Differences in longitudinal and latitudinal distribution of *FT* promoter types were tested by calculating a general linear model that included an estimate of the population structure based on k=5 clusters that was previously determined[1]. The data used for the analysis are complied as dataset 1. The model was calculated in R using the function `glm` (see Supplementary dataset 1).

Supplementary Table 2 F1 hybrids for pyro-sequencing

Crosses	SNPs	Promoter types
An-1XCol	GXC	MXL
LerXTs-1	CXG	MXL
ColXC24	CXG	LXS
ColXWs	CXG	LXS
LerXC24	CXG	MXS
LerXBr-0	CXG	MXS
LerXWs	CXG	MXS
An-1XCvi	GXC	MXM
An-1XLer	GXC	MXM
CviXKyo	CXG	MXM
LerXKyo	CXG	MXM
ShaXLer	GXC	MXM
LerXRRS7	CXG	MXM
LerXGot-7	CXG	MXM

Supplementary Table 3

##for the Fallwinter cohort				
glm(formula = log(Fall$Seedmass) ~ Fall$preFTgeno + Fall$FRIG + Fall$PC1 + Fall$PC2 + Fall$PC3 + Fall$PC4 + Fall$PC5 + Fall$PC6 + Fall$PC7 + Fall$PC8 + Fall$PC9 + Fall$PC10)				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.99579	0.82667	7.253	1.45e-09 ***
Fall$preFTgenom	-0.62938	0.24464	-2.573	0.0128 *
Fall$preFTgenos	-0.27941	0.51181	-0.546	0.5873
Fall$FRIGN	0.45137	0.32342	1.396	0.1684
Fall$PC1	-6.26295	7.23199	-0.866	0.3902
Fall$PC2	-1.61701	0.97815	-1.653	0.1040
Fall$PC3	0.37635	1.08943	0.345	0.7311
Fall$PC4	3.01602	1.35699	2.223	0.0304 *
Fall$PC5	1.05982	1.19048	0.890	0.3772
Fall$PC6	0.70891	0.95464	0.743	0.4609
Fall$PC7	0.55376	1.10459	0.501	0.6181
Fall$PC8	1.35789	1.08555	1.251	0.2163
Fall$PC9	-0.22992	0.89966	-0.256	0.7992
Fall$PC10	-0.09429	0.92774	-0.102	0.9194
##for the Spring cohort				
glm(formula = log(Spring$Seedmass) ~ Spring$preFTgeno + Spring$FRIG + Spring$PC1 + Spring$PC2 + Spring$PC3 + Spring$PC4 + Spring$PC5 + Spring$PC6 + Spring$PC7 + Spring$PC8 + Spring$PC9 + Spring$PC10)				
(Intercept)	5.233360	0.428988	12.199	<2e-16 ***
Spring$preFTgenom	0.112398	0.116339	0.966	0.3374
Spring$preFTgenos	0.335794	0.263601	1.274	0.2070
Spring$FRIGN	0.007736	0.154601	0.050	0.9602
Spring$PC1	-6.263512	3.748539	-1.671	0.0993 .
Spring$PC2	0.275107	0.481551	0.571	0.5697
Spring$PC3	0.108806	0.442340	0.246	0.8064
Spring$PC4	0.202776	0.517725	0.392	0.6965
Spring$PC5	-0.137650	0.487851	-0.282	0.7787
Spring$PC6	0.064956	0.430631	0.151	0.8805
Spring$PC7	-0.097047	0.440441	-0.220	0.8263
Spring$PC8	-0.667620	0.432254	-1.545	0.1271
Spring$PC9	-0.538175	0.452742	-1.189	0.2387
Spring$PC10	-0.543723	0.448564	-1.212	0.2297
Signif. codes: 0 ??***?¯ 0.001 ??**?¯ 0.01 ??*?¯ 0.05 ??.?¯ 0.1 ?? ?¯ 1				

Supplementary Table 3. Statistical test of the association of *FT* promoter types with seed mass during two growth seasons. A general linear regression model was applied to the common garden experiments performed by Korves et al. 2007[4] using the `glm` function in `R`. In this model, `Seedmass` is log transformed data taken from Korves et al. 2007[4] after fitting a model that considered average weight of seeds per silique, average number of siliques per plant and survival rate, `proFTgeno` is the FT promoter type (l:long, m:medium and s:small), season was classified as `spring` and `fall`, `FRIG` is the functionality of FRIGIDA (Functional `F` and non-functional `N`), PC1 to 10 are the first 10 principal coordinates, which are used for controlling the population structure based on available high density SNP data[3] (see Supplementary dataset 2).

Supplementary Table 4

##For the Fallwinter season:				
glm(formula = Fall$DTB ~ Fall$preFTgeno + Fall$FRIG + Fall$PC1 + Fall$PC2 + Fall$PC3 + Fall$PC4 + Fall$PC5 + Fall$PC6 + Fall$PC7 + Fall$PC8 + Fall$PC9 + Fall$PC10, family = quasipoisson)				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.136333	0.140026	36.681	<2e-16 ***
Fall$preFTgenom	0.007178	0.043009	0.167	0.8680
Fall$preFTgenos	-0.118274	0.091951	-1.286	0.2035
Fall$FRIGN	-0.016839	0.055719	-0.302	0.7636
Fall$PC1	-0.658018	1.221432	-0.539	0.5922
Fall$PC2	0.424880	0.170754	2.488	0.0158 *
Fall$PC3	-0.168450	0.166629	-1.011	0.3163
Fall$PC4	0.095647	0.190297	0.503	0.6172
Fall$PC5	0.307632	0.190908	1.611	0.1126
Fall$PC6	0.003413	0.151900	0.022	0.9822
Fall$PC7	0.015666	0.166154	0.094	0.9252
Fall$PC8	-0.074107	0.161556	-0.459	0.6482
Fall$PC9	0.022186	0.156163	0.142	0.8875
Fall$PC10	0.104390	0.158945	0.657	0.5140
##For the Spring season:				
glm(formula = Spring$DTB ~ Spring$preFTgeno + Spring$FRIG + Spring$PC1 + Spring$PC2 + Spring$PC3 + Spring$PC4 + Spring$PC5 + Spring$PC6 + Spring$PC7 + Spring$PC8 + Spring$PC9 + Spring$PC10, family = quasipoisson)				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.809868	0.032410	148.407	< 2e-16 ***
Spring$preFTgenom	0.007481	0.008799	0.850	0.398151
Spring$preFTgenos	-0.013654	0.020054	-0.681	0.498280
Spring$FRIGN	-0.040353	0.011640	-3.467	0.000918 ***
Spring$PC1	-0.156569	0.283547	-0.552	0.582637
Spring$PC2	0.184250	0.036533	5.043	3.62e-06 ***
Spring$PC3	-0.059973	0.033450	-1.793	0.077436 .
Spring$PC4	-0.089806	0.039241	-2.289	0.025216 *
Spring$PC5	-0.087145	0.036885	-2.363	0.021013 *
Spring$PC6	0.044995	0.033050	1.361	0.177880
Spring$PC7	-0.113094	0.033545	-3.371	0.001237 **
Spring$PC8	-0.023654	0.032680	-0.724	0.471664
Spring$PC9	-0.013980	0.034163	-0.409	0.683655
Spring$PC10	0.019985	0.033983	0.588	0.558416
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Supplementary Table 4. Statistical test of the association of *FT* promoter types with days to bolting (DTB). Bolting time data from the study of Korves et al., 2007[4] was treated as count data. A `glm` regression model was applied to the data with quasipoisson correction for the overdispersion of the data. Analysis was done following the suggestions of Crawley (2005)[5].

Supplementary Table 5

glm (formula = SilNb ~ proFTgeno + location + FRIG + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10 + proFTgeno:location + location:FRIG, family = quasipoisson).				
	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	7.34536	0.31716	23.160	<2e-16 ***
proFTgenoM	-0.1159	0.10909	-1.063	0.28890
proFTgenoS	-0.08868	0.22065	-0.402	0.68817
locationNorw	-1.40604	0.19307	-7.282	6.33e-12***
locationValen	-1.17363	0.17091	-6.867	7.14e-11***
FRIGN	0.17493	0.10754	1.627	0.10528
PC1	-2.62194	2.80872	-0.933	0.35162
PC2	-1.29656	0.29857	-4.343	2.18e-05***
PC3	-0.37570	0.40119	-0.936	0.35010
PC4	1.09566	0.37976	2.885	0.00432 **
PC5	-0.4752	0.31718	-1.498	0.13553
PC6	0.47380	0.36127	1.311	0.19111
PC7	0.21856	0.35759	0.611	0.54173
PC8	-1.43248	0.35043	-4.088	6.18e-05***
PC9	-0.13306	0.34375	-0.387	0.69908
PC10	0.24461	0.29238	0.837	0.40373
proFTgenoM:locationNorw	-0.07774	0.21824	-0.356	0.72205
proFTgenoS:locationNorw	0.03069	0.49296	0.062	0.95041
proFTgenoM:locationValen	0.42046	0.18407	2.284	0.02335 *
proFTgenoS:locationValen	-0.26176	0.49313	-0.531	0.59609
locationNorw:FRIGN	-0.30175	0.20633	-1.462	0.14509
locationValen:FRIGN	-0.20970	0.16250	-1.290	0.19829
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Supplementary Table 5. Statistical test of the association of *FT* promoter types with silique number. Silique number as measure of fitness was taken from a field study performed Fournier-Level et al 2011 at 3 different European sites[2]. Silique number treated as count data in a `glm` regression model. Quasipoisson correction was applied to account for the overdispersion of the count data. Analysis was done following the suggestion of Crawley (2005)[5]. In this model, `SilNb` is the silique number, `proFTgeno` are the *FT* promoter length types (long (`L`), medium (`M`), small (`S`)), location corresponds to three experimental sites, FRIG is the functionality of FRIGIDA[4], PC1 to 10 are the first 10 principal coordinates according to published high density SNP data[3], which are used for controlling the population structure (see Supplementary dataset 3).

Since removing `location:FRIG` from the model resulted in slightly different results (p=0.09), we kept it as parameter in the final model. Note a significant difference between `proFTgeno` L and M at the Valencia site. This difference was also mildly supported by a non-parametric Wilcoxon rank sum test (p=0.06).

Supplementary Table 6. List of oligonucleotides used in this study.

Purpose	Primer name	Primer name
Plasmid construction	5.2kbFTp-GW-FW	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGATTGTTTTACACTACTTCC
	1kbFTp-GW-RE	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTGATCTTGAACAAACAGGT
	BlockC-GW-FW	GGGGACAAGTTTGTACAAAAAAGCAGGCTCATTTGCTGAACAAAAATCT
	BlockC-GW-RE	GGGGACCACTTTGTACAAGAAAGCTGGGTAAACGTTTGGAAATAGGAAGTATG
	BlockC-overlapA-RE	GTACCGCCAAAAACGTTTGGAAATAGGAAGTATG
	BlockA-overlapC-FW	TTCCAAACGTTTTTGGCGGTACCCTACTTTTT
Genotyping	ft10_GABI_RP	CAGGTTCAAAACAAGCCAAGA
	ft10_GABI_LB	CCCATTTGACGTGAATGTAGACAC
	LBb1.3	ATTTTGCCGATTTCGGAAC
	12_forward	CAACGAGATTTGGGGTTAAG
	13_reverse	GATGCATTGTTTAAGAAAATCAGG
Gene expression analysis	FT_cDNA_RT_fw	GGTGGAGAAGACCTCAGGAA
	FT_cDNA_RT_re	ACCCTGGTGCATACACTGTT
	PP2A-RT-fw	AAATACGCCCAACGAACAAA
	PP2A-RT-re	CAGCAACGAATTGTGTTTGG
Pyro-sequencing	F1	TGTGTGTTACGAAAATCCAAGTCC
	R1	(Bio)AGCCACTCTCCCTCTGACAATT
	S1	GTTTCGACAGCTTGG
3C	Primer1	CCATCTTCCCACTCCCTTCT
	Primer2	TGGAAGTGGAAATGAATGTTAGG
	Primer3	TGTTGGCCAAGATGTCTCAC
	Primer4	TGTTCCCATGTGTGTGTGTG
	Primer5	TTGAATGCAGTCCGATTGTC
	Primer6	AAAAATTGCGACTTCTCAAATAA
	Primer7	TCTGCGTCACCATGAAGAAA
	Primer8	GCCACTGTTCTACACGTCCA
	Primer9	AAAATTCGAAAGCGAAAACG
	Primer10	GGAACAAAGAAAAATCCCAAGA

Supplementary References

1. Cao J*, et al.* Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nat Genet* **43**, 956-963 (2011).

2. Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. A map of local adaptation in Arabidopsis thaliana. *Science* **334**, 86-89 (2011).

3. Atwell S*, et al.* Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**, 627-631 (2010).

4. Korves TM*, et al.* Fitness effects associated with the major flowering time gene FRIGIDA in Arabidopsis thaliana in the field. *Am Nat* **169**, E141-157 (2007).

5. Crawley. *Statistics: an introduction using R*. Wiley (2005).