Explorations of the polygenic genetic architecture of flowering time in the worldwide *Arabidopsis thaliana* population

Yanjun Zan[1], Örjan Carlborg[1*]

Affiliation: [1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

* Corresponding author

Email: orjan.carlborg@imbim.uu.se

Abstract

As a locally adapted complex trait, flowering time in *Arabidopsis thaliana* has attracted much attention in genetics. Most studies have, however, focused on contributions by individual loci rather than the joint contributions by the large number of loci in the genetic architecture of flowering time to local and global adaptation. In an earlier study, we reported 46 loci associated with flowering time variation during growth at 10°C, 16°C or both in the 1,001-genomes collection of *Arabidopsis thaliana* accessions. Here, we explore how these loci together contribute to differences among genetically defined, and geographically divided, subpopulations across the native range of this species. Our approach was to define flowering time as a trait, and the measurements at 10 and 16 °C as two independent measurements of it. This facilitated explorations of the dynamics in the genetic architecture –which loci contribute and their effects– of flowering time across growth temperatures and their potential roles in local and global adaptation. The overall flowering time differences between populations could be explained by subtle changes in allele-frequencies and gradual changes in phenotype due to globally present alleles. More extreme local adaptations were on several occasions due to contributions by regional alleles with relatively large effects. About 2/3 of the 48 evaluated flowering time loci had similar effects on flowering time at 10°C and 16°C, while the remaining 1/3 had different effects in the two temperatures, suggesting an important contribution of gene by temperature interactions to this trait. There are also indications that co-evolution of functionally connected alleles in local populations has been important for local adaptation. Overall, this study provides deeper insights to the polygenic genetic basis of flowering time variation in *Arabidopsis thaliana* across a wide range of ecological habitats.

Author Summary

Many genes can affect flowering time in *Arabidopsis thaliana*, but their contribution to natural flowering time variation in the worldwide population is largely unknown. We explored how 48 loci associated with flowering time, measured at 10°C and 16°C, or their difference, for the same wild collected 1,001-genomes *Arabidopsis thaliana* accessions together contribute to differences among the genetically defined and geographically divided subpopulations from the native species range. The overall flowering time differences among these subpopulations could be explained by the joint small effects of globally present alleles, suggesting an important contribution by polygenic adaptation for this trait. Most alleles with large effects on flowering were present only in some populations, facilitating more extreme local adaptations. Long-range LD was observed between genes in several biological pathways, indicating possible local adaptation via co-evolution of functionally connected polymorphisms. The genetic architecture of flowering time was also found to depend on the growth temperature. Most flowering time loci had similar effects on flowering time measured at 10°C and 16°C, but the effects of about 1/3 of them had effects that varied with temperature. Overall, new insights are provided to how the polygenic architecture of flowering time has facilitated its colonisation of a wide range of ecological habitats.

Introduction

Understanding the genetic basis of adaptation is a central theme in evolutionary genetics [1,2]. Adaptation in nature can rely on a few loci with large individual effects [1,3], and such loci have been identified in natural populations of for example *Arabidopsis thaliana* [4–7] and the Darwin finch [8]. In other cases, populations adapt via selection on many loci with small effect [9], with examples reported for fitness traits in a range of species including mice, drosophila, human and maize[10–12]. In depth characterizations of the genetic mechanisms contributing to adaptation for polygenic quantitative traits is, however, a difficult [13] and despite some successes little is overall known. Adaptations for polygenic traits could be rapid due to selection on available standing variation [14–16] and new adaptive optima could be reached via gradual allele frequency changes across multiple loci [17,18]. The individual effects of the contributing loci might be too low for detection in genome wide scans based on stringent multiple testing thresholds, but they might despite this be necessary for reaching fitness optima [19–21].

Arabidopsis thaliana, a widely used model species in plant biology, has colonised a broad range of ecological habitats around the world. One of the most well studied adaptive traits is flowering time due to its role in ecological adaptation and potential impact on agronomic production in related species. Molecular studies in the laboratory, primarily using the reference accession *Col-0*, have identified many genes with the potential to alter flowering time. Association studies in natural populations, however, rarely reveal more than a handful of the loci contributing to the genetic variation of this trait [22–25]. Hence, even though it is likely that flowering time adaptations across the native range of *A. thaliana* would involve many loci, the polygenic basis of these differences is still largely unexplored.

Previously, we developed and used a new statistical approach to dissect the genetic basis of polygenetic traits in natural populations. The properties of the method was illustrated by exploring the polygenic basis of flowering time in *Arabidopsis thaliana*, using public data from the 1,001 genomes project[26]. In total, 33/29 loci were there mapped for flowering times in 1,004 wild collected *Arabidopsis thaliana* accessions from the 1,001-genomes project grown at 10/16°C[26]. Here we explored how the genetic effects of these loci i) vary within and across loci and growth temperatures and ii) how these variations in allelic effects contribute to the flowering time variation between genetically defined subpopulations that have colonized the native range of the species [22].

Loci with globally segregating alleles made, in general, small individual contributions to flowering but together captured the overall pattern of flowering time differences between the subpopulations, suggesting a role of polygenic adaptation[14–18].The alleles with more restricted geographic distributions had in several instances large effects and made important contributions to local adaptations of some subpopulations. The genetic effects of many loci varied across temperatures, suggesting that genotype by temperature interactions might be important for flowering time adaptation. The long-range LD between polymorphisms in genes of the same biological pathways suggests that co-evolution of different pathways has contributed to local adaptation in different geographic areas. Overall this reanalysis of public data provides new insights to the polygenetic basis of flowering time adaptation in *A. thaliana*.

Results

Flowering time differentiation in *A. thaliana* worldwide population

Flowering time variation of *Arabidopsis thaliana* accessions sampled from its native range is strongly associated with geographic variation in climate [27–29]. To study the genetics of the this geographical differentiation in greater detail, we reanalysed a publicly available dataset where flowering time was measured at 10°C (FT10) and 16°C (FT16) under greenhouse conditions on 1,004 whole-genome re-sequenced *Arabidopsis thaliana* accessions[22]. The genetic population structure and geographic sampling locations for the analysed accessions are illustrated in S1 Fig.

We approached flowering time in this population as a single complex trait measured at two temperatures to reveal the dynamics in its underlying genetic architecture. At these temperatures, flowering times displayed a continuous variation with earlier average flowering times (9.3 days) at 16 °C than at 10 °C (Figure 1A). The variations in flowering times at both growth temperatures were correlated with the location and climate at the sampling sites of the accessions, suggesting its potential role in global, and local, adaptation. For example, there was a significant latitudinal cline of flowering times (Fig 1B/S2A Fig for FT10/FT16; Pearson correlation = 0.37; $P < 2.2 \times 10^{-16}$; df = 963) and a significant correlation between flowering times and mean temperature at the sampling locations (Fig 1C/S2B Fig for FT10/FT16; Pearson correlation = -0.31; $P < 2.2 \times 10^{-16}$; df = 963).

The phenotypic variance increased with temperature, being larger at 16°C than at 10°C (Fig 1A), while the proportion of additive variance estimated by fitting the IBS (Identity By State) kinship matrix were similar at both temperatures (82/83% at 10/16°C). Additive genetic variance is thus released as the temperature increases, which will be explored in more detail in

126 the following section. Regardless of whether flowering times were measured at 10°C or 16°C,
127 there were substantial phenotypic differences between the eight genetically defined and
128 geographically separated subpopulations defined by *Alonso et al* [22] in their admixture
129 analysis of this population (Fig 1D). Increasing the growth temperature leads to earlier
130 flowering times in most of the sub-populations, with the North Swedish population being the
131 only exception with an on average later (4.9 days) flowering time at 16°C.
132
133 **[Fig 1 about here]**
134 **The genetic architecture of flowering time is highly polygenic in the worldwide**
135 ***A.thaliana*** **population**
136 By utilizing prior molecular knowledge of genes regulating flowering time variation in the
137 reference accession (*Col-0*), and results from an expression QTL analysis of those genes in
138 this 1001-genomes collection of natural accessions, we earlier mapped 46 flowering time loci
139 where 33/29 loci were associated with FT10/FT16 using a multi-locus association
140 analysis[26]. Together these mapped loci explain 55/48% of the total genetic variance for
141 FT10/FT16[26]. Here, rather than focusing on the results from the two temperatures
142 separately, the complete genetic architecture of flowering time revealed across them was
143 evaluated. Overall, 16 loci were mapped at both temperatures, and 2 pairs of loci were located
144 within 20 kb from each other, suggesting that they contribute to flowering time regardless of
145 temperature in a relatively stable manner. The remaining 30 loci were significantly associated
146 with flowering times only at one temperature in the earlier analysis[26], and their possible
147 involvement in genotype-by-temperature interactions are further explored here.
148
149 Two additional analyses were performed to explore the possible role of genotype-by-

150 temperature interactions in this population. First a GWA mapping analysis to the flowering
151 time difference at the two evaluated temperatures (FT10-FT16). Two genome-wide
152 significant associations, one at *Chr1: 23,453,207* bp and one at *Chr4: 1,338,596* bp, were
153 detected in this analysis after Bonferroni correction for ~1.4 million SNPs (S3 Fig). Second,
154 the polygenic contribution of the loci that were significant in one temperature, but not the
155 other, to flowering times was evaluated. This by testing if loci associated with FT16, but not
156 FT10, together with the two additional loci affecting FT10-FT16 jointly contributed a
157 significant amount of genetic variation in FT10, and vice versa. Models with or without non-
158 overlapping loci were fitted, and significant increases in the adjusted R^2 were detected at both
159 temperatures (4/8% for FT10/FT16; Likelihood ratio test; P-value= 3.2×10^{-14}/1.3×10^{-33}; S4
160 Fig).
161
162 Linear models including the genotypes of the 48 mapped flowering time loci, but excluding
163 their kinship, explained the majority of the additive genetic variance at both temperatures
164 (79/75% for FT10/FT16). In the following sections, we will explore how these 48 loci
165 together contribute to flowering time across the different genetically and geographically
166 divided sub-populations in this population, to potentially gain deeper insights to how
167 polygenic and genotype-by-temperature interactions might contribute to local and global
168 adaptation in nature.
169
170 **The role of genotype-by-temperature interactions for flowering times in *A. thaliana***
171 It is known that genotype-by-temperature interactions are important for flowering times in
172 *Arabidopsis thaliana* [30–32]. The genetic basis of this phenomenon is, however, largely
173 unknown. Here, we studied this by evaluating the temperature dependence of the genetic

174 effects of 48 mapped flowering time loci. As shown in Fig 1A, increased growth temperature
175 leads to a release of additive genetic variance in flowering time. To identify the genetic basis
176 of this, the genetic effects of the 48 mapped flowering time loci on FT10 and FT16 were
177 estimated by fitting them jointly in a mixed model (Fig 2). The alleles at 42 of the 48 loci
178 have effects in the same direction on FT10 and FT16, i.e they consistently increase or
179 decrease flowering time regardless of temperature. The alleles at the remaining six loci have
180 opposite effects in 10°C and 16°C, but as all these estimates in at least one temperature are
181 close to zero (Fig 2A). This result is likely to at least in part be the outcome of the
182 uncertainty in the estimation of the allelic effects. For ~2/3 of the loci, the sizes of the genetic
183 effects on flowering time are in good general agreement (Fig 2A), suggesting that they affect
184 flowering time in similar way regardless of whether the plants are grown at 10°C or 16°C.
185 The remaining ~1/3 loci, however, have larger effects on flowering time when the plants are
186 grown at 16°C than when they are grown at 10°C (Fig 2A). This is consistent with the
187 observation of an increased additive genetic variance at 16°C, but at the same time
188 unexpected as an increase in temperature generally leads to an overall decrease in flowering
189 time. To evaluate whether these putative genotype-by-temperature were statistically
190 significant, we compared the fits of two linear models for every locus: one including all 48
191 loci and an indicator variable for the two growth temperatures as fixed effects, and another
192 also including an interaction term between growth temperature and each tested locus. In total,
193 18 loci displayed significant genotype-by-temperature interactions after correction for
194 multiple testing (highlighted in Fig 2A; names of loci & P-values from the corresponding
195 Likelihood Ratio Test in Table S1).
196
197 **[Fig 2 about here]**

198

199 **Distribution and effects of flowering time alleles in the worldwide *A. thaliana* population**

200 To study the putative role of the 48 mapped flowering time loci in global and local adaptation,
201 we evaluated how the alleles at these were distributed across the subpopulations identified in
202 this dataset by *Alonso et al* [22]. The assumption in this analysis is that allele frequency
203 differences between subpopulations across these loci, in its most extreme form the complete
204 presence/absence of alleles, at least in part reflects past and ongoing selection leading to
205 global adaptation along latitudinal or geographical clines, and local adaptations in individual
206 subpopulations [33–35]. All 48 mapped loci were identified by having effects on flowering
207 time at 10°C, 16°C, their difference or all, and we therefore consider all to be candidate
208 adaptive loci as the temperature will vary both within and across the sampling sites of the
209 accessions.

210

211 Consistent with the observation that the 48 mapped loci explained the majority of the additive
212 genetic variance in flowering time, the overall pattern of early to late flowering times between
213 the subpopulations in this dataset is captured well by the combined additive effects of these
214 (Fig 3A/S4 Fig). Further, late flowering populations have accumulated both major and minor
215 effect late flowering alleles across many loci (Fig 3B; late flowering alleles in red; effect sizes
216 sorted from top to bottom). The late flowering north Swedish accessions are most extreme,
217 with the highest frequencies at many loci. A visual inspection of the allele frequency patterns
218 of the 48 loci across the populations (Fig 3B) suggests a greater overall similarity between
219 populations from nearby geographic locations. This is analytically supported by a clustering
220 analysis across the subpopulations based on the allele frequencies across these 48 loci (Fig
221 3C), and the pairwise correlations of the allele-frequencies between the populations (S5 Fig).

For example, the accessions from Germany and Central Europe are closely related (Figs 3B and 3C), with an exception being the relatively late flowering southern Swedish accessions that overall were more similar to the earlier flowering Central European/German accessions than the late flowering north Swedish accessions.

[Fig 3 about here]

For some of the 48 loci, both alleles are present (here defined as having MAF > 0.05) in all subpopulations, for example the loci on chromosomes 2 (16,387,379bp) and 4 (1,961,868bp) (Fig 2B; S6A/S6B Fig). For others, one of the alleles is either rare (MAF < 0.05) or absent in one or more of the subpopulations, for example chromosomes 2 (9,312,968bp), 4 (1,728,204bp) and 5 (22,979,827bp) (Fig 2B; S6C/S6E Fig). Both global and local allelic variations are thus likely to contribute to the natural flowering time variation across the worldwide range of *A. thaliana*. To explore this further, the 48 mapped flowering time loci were divided into two groups containing either those that are present across all populations (MAF > 0.05 in all populations; n = 23), or those that are rare or absent in at least one subpopulation (MAF < 0.05 in at least one population; n = 25). The choice of 0.05 as MAF cut-off is arbitrary but nevertheless serves the purpose of representing the spatial distribution of global vs primarily local alleles (Fig 3; Fig 5), and the results are overall similar also for MAF < 0.03 and MAF < 0.10 (results not shown).

Next, the contributions by global and local alleles to flowering time variation in the 8 subpopulations was evaluated. A mixed model, including the IBS kinship and the genotype of all 48 loci, was first fitted to FT10 and FT16 separately. Then, the estimated additive effects

of these 48 alleles were used to calculate the contributions by local and global alleles, as defined above, to flowering time variation.

Contributions by global alleles to flowering time variation: For FT10, the IBS kinship captures the overall early to late flowering time pattern for subpopulations well and explicitly modelling the global alleles improves it further (Fig 4A). This illustrates that the polygenetic effect, together with the gradual allele frequency shifts across the loci with global alleles, are important for the overall pattern of flowering time differentiation between the subpopulations. The majority of the global alleles have moderate to small effects on flowering time (Fig 4A/B), relative to the local alleles (discussed further in section below), but together they contribute 21% of the phenotypic variance of the trait measured at both 10°C and 16°C. This analysis also shows that the global alleles have a larger contribution at 16°C than that to 10°C, again suggesting a role of genotype-by-temperature interaction for this trait (Fig 4A/B). The kinship is estimated to make a negative contribution to flowering times measured at 16°C in some subpopulations (Fig 4B), while the joint contribution by the global alleles is larger than at 10°C. One likely interpretation of this is that some of the global alleles have large overall effects at this temperature. They seems to also differ in effects across subpopulations and this genotype-by-environment interaction is captured by the kinship.

[Fig 4 about here]

Contribution by local alleles to flowering time variation. We next explored the loci with alleles that are rare or absent (MAF < 0.05) in at least one of the populations in more detail (Fig 5). Across these loci, the most apparent enrichment is for late flowering alleles in the

Swedish populations (Fig 3B), where one or both of these populations contain near private late flowering alleles at high frequencies. These include the strongest late-flowering allele of all (chromosome 5 at 23 Mb; Fig 5), suggesting this large effect allele having emerged locally and been under strong selection here. A few other loci with relatively large effects on flowering were also predominantly present in one or a few subpopulations, but none were as striking as the aforementioned locus (Figs 3B and 5; S4 Fig).

The local alleles appears to primarily be contributing more extreme phenotypes in individual subpopulations (Fig 4A). In contrast to the global alleles, there is not only generally larger allele frequency differences between the populations for the local alleles, this group is also enriched for alleles with large effects that individually explain larger proportions of the phenotypic variance (Fig 4B). Together the local alleles contribute 45/42% of the phenotypic variance for FT10/FT16 in this population. The flowering time differences between the subpopulations thus result from shifts in the allele frequencies of global alleles (Fig 3B), with more extreme (for example northern Sweden) or variable (for example Spain) local phenotypes being due to private, or near private, local alleles (Fig 5).

[Fig 5 about here]

Long-range LD between flowering time loci

Long-range LD between loci is a possible sign of co-selection of them. To screen for such signatures, we calculated the pairwise LD (D') between the 48 associated loci (Fig 6). In total, 67 pairs were in significant nominal long-range LD (P-value < 0.05; overall FDR for this set = 12%), where a few loci were in long-range LD with several other loci on either the same or

different chromosomes (Fig 6). When a locus was in long-range LD with more than 4 other loci, the loci were defined as belonging to an LD-cluster.

[Fig 6 about here]

We explored the five identified clusters in more detail. A literature review suggested functional connections between candidate genes in three of the five clusters. One cluster contained several genes in a miRNA mediated flowering time regulation pathway, suggesting a functional connection to the FLC-CO/FT module [36–41] (Table 1). Two clusters contained genes whose function are connected to the *FRI-FLC* module, one of these possibly through vernalization [42–44]. The minor alleles of the five loci that via their long-range LD defined these clusters were enriched in specific geographic regions. The one allele on chromosome 2 (15,044,483bp) was found primarily in Sweden, the one on chromosome 1 (16,471,394bp) in Italy Caucasus, the one on chromosome 3 (7,240,079bp) in Asia, the one on chromosome 4 (17,282,047bp) in Spain and the one on chromosome 5 (22,979,827bp) in north Sweden (Fig 4B). Together this suggests possible contributions from these pathways to local adaptation in these geographic regions.

Table 1. Long-range LD defined clusters of genes with assigned functional classification

Cluster[a]	Locus[b]	Gene	Alias	Functional classification[c]
1	1:16471394*	AT1G03970	GBF4	Unclear
	1:23453207	NA	Unknown	
	2:15044483	AT2G42590	GRF9	
	2:19204119	AT5G08190	NF-YB12	
	3:16321974	AT3G03450	RGL2	
	3:16325481	AT3G33520	SUF3	

	4:8272818	AT2G23080	CKA3	
	5:16541207	AT5G17490	RGL3	
	5:22979827	AT5G57380	VIN3	
	5:26805231	AT5G67180	TOE3	
	2:15044483*	AT2G42590	GRF9	
	2:16387379	AT2G39250	SNZ	
	4:259252	AT4G00650	FRI	
2	4:1338596	NA	Unknown	FLC-CO/FT
	4:7172638	AT2G28190	SOD2	
	4:10999188	NA	TSF	
	4:17282047	NA	Unknown	
	5:3274505	AT4G00650	FLA	
	3:7240079*	AT3G20740	FIS3	
	3:16325481	AT3G33520	SUF3	
3	4:1338596	NA	Unknown	FRI-FLC
	4:8272818	AT2G23080	CKA3	
	4:17282047	NA	Unknown	
	5:5345341	AT5G16320	FRL1	
	4:17282047*	NA		
	2:9048513	AT3G22380	TIC	
	2:15044483	AT2G42590	GRF9	
4	3:7240079	AT3G20740	FIS3	Unclear
	4:259252	AT4G00650	FRI	
	4:1338596	NA	Unknown	
	4:6559815	AT4G10180	DET1	
	5:22979827*	AT5G57380	VIN3	
	1:16471394	AT1G03970	GBF4	
	2:14312485	AT2G33835	FES1	
5	4:259252	AT4G00650	FRI	Vin-FRI-FLC
	4:1338596	NA	Unknown	
	4:7172638	AT2G28190	SOD2	
	5:3274505	AT4G00650	FLA	

	5:5345341	AT5G16320	FRL1

[a]Cluster as defined by the high long-range LD to the first locus (underlined) in the group; [b]Locus (chromosome: position in bp) associated with flowering time at 10°C, 16°C or both in the multi-locus analysis; [c]Functional classification of the cluster based on the likely pathway affected by the polymorphisms based on a literature study of the included genes [36–43,45–47].

Discussion

In an earlier study, we mapped the polygenetic basis of flowering time variation in the latest public release of data from the 1,001 genomes *A. thaliana* project [22,26]. In this analysis, 33/29 loci were reported to be associated with flowering time variation measured at 10/16°C and they together explained 66/58% of the additive genetic variation[26]. These loci were mapped using a two-step polygenic analysis method. First, loci with individual genome-wide significant associations either directly to measured flowering times, or to the expression of experimentally known flowering time genes, were identified as candidate genes for the polygenic analysis in this population. Then, a joint analysis across all these candidate loci was performed for each trait separately using a backward elimination analysis, to infer a polygenic architecture of the trait at a 15% false discovery rate significance threshold. At this significance threshold, a few of the 33/29 loci associated with FT10/FT16 are expected to be false positives. Here, we continued this work by treating flowering time as one trait measured independently for the same genotypes at 10°C and 16 °C. The focus was to evaluate the joint contributions by all mapped flowering time loci to the genetics of this trait and their possible role in global and local adaptation to the different ecological conditions across the worldwide range of this plant. We found that the groups of loci that were individually significant in one growth temperature, but not the other, made significant joint contributions to flowering time

337 in both conditions. Fitting all 48 loci together in a multi-locus linear model, they explain most
338 (79%/75% for FT10/FT16) of the additive genetic variance in flowering times at these growth
339 temperatures. As most of the loci are likely to be true associations, we performed an in-depth
340 evaluation of how this set of loci contribute to the global flowering time variation in *A.*
341 *thaliana* to provide valuable complementary information to earlier studies that have focused
342 primarily on loci with strong individual associations[23,25,48,49].
343
344 It is generally difficult to study the genetic basis of highly polygenic traits in natural
345 populations due to, for example, the small contributions made by individual loci to trait
346 variation and the confounding of adaptive alleles with population structure. This is likely one
347 of the reasons why standard genome-wide association mapping approaches have only been
348 able to reveal a small number of flowering time loci in this or related populations [22–24]. By
349 approaching flowering time as a polygenic complex trait, and making use of its measures in
350 two temperatures, this study could shed lights on how allele frequency changes across many
351 loci, as well as genotype by temperature interactions, contribute to the genetic basis of global
352 and local trait variation.
353
354 **The role of global and local alleles in shaping worldwide flowering time variation**
355 For many of the polymorphic flowering time loci, both alleles were present in all
356 subpopulations of the worldwide population, albeit at different frequencies. Together these
357 contributed to the overall pattern of early to late flowering in the geographically divided
358 subpopulations. Polygenic global adaptation via small allele frequency shifts across many loci
359 thus appears to be a possible contributor to the variation of this trait in *A. thaliana*.
360

361 For a number of the associated loci, one of the alleles was absent or present at low frequency
362 (MAF < 0.05) in one or more of the subpopulations. Several of these local alleles had strong
363 effects on flowering time, and these loci made important contributions to the more extreme
364 flowering time phenotypes of individual subpopulations. Alleles with large effects on
365 adaptive traits have earlier been reported to be present in local populations of *Arabidopsis*
366 *thaliana* [5,7,23,50]. Our finding that many such alleles are present in locally in the
367 worldwide population, is coherent with earlier speculations that locally adaptive alleles for
368 flowering time variation in *Arabidopsis thaliana* are likely to have large effects [51]. Those
369 local populations appear to reach their respective flowering times via the effects of alleles that
370 are primarily present in some populations suggests that different combinations of alleles are
371 likely explaining adaptation to different local environments.
372
373 We evaluated the distribution of many flowering time alleles across the native range of the
374 species. The results are interpreted under the assumption that the alleles associated with
375 flowering time are very likely to have an adaptive value and that the observed patterns of
376 allele frequency differences between populations, to large extent, represent local and global
377 adaptations. Other explanations are, however, also possible as similar patterns could emerge
378 due to, for example, an interplay of drift, historical contingency and selection. A few
379 observations are particularly valuable when interpreting the results. First, the consistency in
380 which loci are mapped, and their patterns of effects when the plants are grown in two
381 different growth temperatures, suggests that the genetic architecture of flowering time is
382 overall stable. Hence, even though the individual effects of many loci are small, making it
383 difficult to assess the selective pressure on each locus in nature, polygenic adaptation acting
384 over time on the many alleles that are present globally in *A.thaliana* appears like a possible

route to adaptation. The finding that stronger local alleles complements the global polygenic adaptation, likely through a more rapid increase in allele-frequencies due to stronger selection for locally favourable phenotypes, is also consistent with earlier reports in which a few flowering time QTL were found to be correlation with latitude [48,52], indicating a potential role of spatial adaptation. That the same process might have been important also for other traits is suggested by the results of another recent study on drought adaptation reporting a polygenic genetic architecture and accompanying geographical enrichment of QTL[53]. Another possible explanation for the finding of, more or less, local alleles in several subpopulations is that the plants are naturally restricted to its native range over their lifespan and that self-pollination in *Arabidopsis thaliana* slows down the long distance spreading of new genetic variants. Hence, newly emerged alleles with large beneficial adaptive effects can quickly sweep to high frequency in local populations, but long distance spreading of those alleles will take longer in this selfing plant, although it is known to happen due to for example human or animal activities [54]. A possible example of such long distance spreading of genetic variants observed in this data, is the large effect allele identified on chromosome 5, 23 Mb. Here, the strong allele is present at high frequency in northern Sweden and in a relict population in Spain (S4F Fig). These populations are geographically distant, and it appears a likely result from lineage mixing between relict and northern populations as indicated in an earlier study of this dataset [22].

Two late-flowering alleles, with intermediate to high effect, are private (or nearly so) to the Spanish population (Fig 4). We note that this population has a significantly higher within-population variation in flowering time ($P = 1.1 \times 10^{-6}$; Brown-Forsythe test) than the other populations (Fig 2A). The reason for this is not known, but a possible explanation could be

409 that these strong alleles have contributed to specific local adaptations in this subpopulation.
410 Further studies of these effects would be valuable to evaluate the possible contributions of
411 these alleles to local adaptations.
412
413 **Possible contribution by co-evolution of functionally related loci in local populations**
414 Four of the flowering time associated loci were in long-range linkage-disequilibrium (D')
415 with several other mapped loci. These alleles were enriched in relatively distinct geographic
416 regions, suggesting that they have emerged and spread locally making a confounding with
417 population structure a possible explanation for this observation. We, however, consider this
418 unlikely as they were detected in a statistical multi-locus analysis accounting for population-
419 structure and hence each of them captures flowering time variation that could not be
420 explained by the other loci. Further, as in three of the five cases the loci involve candidate
421 genes in related biological pathways. An intriguing alternative interpretation is that the LD
422 instead result from co-selection of multiple functionally connected loci. Further work is,
423 however, needed to confirm this hypothesis.
424
425 **Genotype-by-temperature interactions and their possible implications for adaptation**
426 An earlier study reported genotype by environment interactions in the Swedish *A. thaliana*
427 populations that were part of the dataset analysed in this study [30]. Here, we explored the
428 possible role of genotype-by-temperature interactions for in total 48 flowering time loci
429 associated with this trait in the entire 1,001-genomes collection of worldwide accessions
430 grown at both 10°C and 16°C. Our conceptual approach to study these interactions was
431 slightly different to the one used before. First, flowering time was considered as a single
432 complex trait and all loci associated with this trait, regardless of growth temperature, was

433 considered as possible loci involved in genotype-by-temperature interactions. Next, the
434 effects of all the 48 associated loci on the flowering times measured at 10/16°C were
435 estimated. The allelic effects in the two growth temperatures were compared, and possible
436 genotype-by-temperature interactions inferred based on observed differences in effects
437 between the two temperatures. Given the high phenotypic correlation between flowering
438 times of accessions grown in these different temperatures (Pearson correlation=0.88), we
439 expected the effects to be overall similar between environments. However, for ~1/3 of the
440 mapped loci (18 of 48 loci; Fig 6), the effects on flowering time changed with temperature.
441 These results suggest that the genetic effects of many loci in the genetic architecture of this
442 polygenic trait are dependent on one or more environmental factors, as exemplified by
443 temperature here. As studies aiming to detect complex trait loci, for example using a GWAS
444 approach, are generally limited in their ability to assess the effects of loci across multiple
445 environments a significant fraction of loci important in the variable conditions under which
446 plants live in nature might be missed. More studies, for example to identify seasonally
447 sensitive QTL in the field using GWA analysis[48], will therefore be of great value for our
448 understanding of the genetic of such traits and their potential role in adaptation.
449
450 The potential adaptive values of the genotype-by-temperature interactions identified here is
451 difficult to assess. We observed a geographical enrichment of loci involved in such
452 interactions, that have larger effects at 16°C, in the northern Swedish populations, where
453 flowering time is increased at higher temperature (S7 Fig). Further investigations uncovered
454 significant pairwise correlations between the joint effects of these loci on flower time
455 differences between temperatures (FT10-FT16) and within year temperature fluctuation
456 (temperature difference between hottest and coldest day; S8 Fig). It is therefore possible that

these loci are of adaptive value in environments where the temperature fluctuates drastically, and where having alleles that can shorten flowering time in response to cold temperatures would be beneficial. However, further studies are needed to evaluate this hypothesis thoroughly.

Concluding remarks

Our study demonstrates that the variation in flowering times between earlier defined geographically and genetically separated *A. thaliana* subpopulations is likely due to allelic variations in many loci. At some of these loci both alleles are present across the entire range of the species, whereas others are only present in some of the subpopulations. Across these loci, the allele frequency shifts were often more gradual at loci with global present alleles. Larger differences in allele frequencies were observed for several loci with alleles of relatively large effects, and higher allele frequencies were then found in local populations with more extreme phenotypes. This suggests that flowering time adaptation results via a combination of global polygenic adaptation across many loci, complemented by local adaptation via the effects of stronger alleles with a more regional distribution. The results also indicate that local phenotypes might also result from co-evolution of multiple polymorphisms in biological pathways. We identified genotype-by-temperature interactions involving 18 flowering time loci, and propose a hypothesis that these might be adaptive in environments with large temperature fluctuations. However, further investigations are needed to validate this hypothesis. Overall, this study provides new insights to the genetic basis of flowering time in *A. thaliana* and propose possible explanations for how this self-pollinating plant has been able to colonize and adapt to such a wide range of ecological habitats around the world.

Materials and Methods

Data

All phenotype, genome-sequence and subpopulation information are publicly available as part of the *Arabidopsis thaliana* 1,001-genomes project [22,55]. Flowering times measured at 10°C and 16°C in the greenhouse were downloaded from [56,57]. The subpopulation classifications of the accessions were downloaded from http://1001genomes.org/tables/1001genomes-accessions.html. The imputed whole genome SNP data matrix was downloaded from http://1001genomes.org/data/GMIMPI/releases/v3.1/SNP_matrix_imputed_hdf5/1001_SNP_MATRIX.tar.gz. We filtered the SNP data for minor allele frequency and only retained loci with MAF > 0.03. SNP markers were pruned to remove loci in pairwise LD of $r^2 > 0.99$. In total, 1,396,438 SNPs on 1,004 individuals remained. The 33/29 loci associated with flowering times at 10/16°C were identified in a previous study[26]. These loci were inferred using a two-step polygenic mapping method. First, candidate flowering-time loci were identified as those with either i) individual genome-wide significant associations to flowering time or ii) that altered the expression of genes that are experimentally known to affect flowering. Then, these candidate loci were analysed jointly for associations to flowering-times measured at 10°C and 16°C using a backward elimination based multi-locus mapping approach to infer the polygenic architecture of flowering time at a 15% false discovery rate significance level [26].

Detecting loci involved in genotype by environment interactions

Two methods were used to test for QTL by environment interaction. The first was a genome wide scan using the mean flowering time difference between 10°C and 16°C as phenotype. In

the analysis, a forward-selection based approach was employed where a linear mixed model $\bar{y}_{FT10} - \bar{y}_{FT16} = \mathrm{X}\beta + \mathrm{Z}u + \mathrm{e}$ (1) was iteratively fitted to the SNPs across the genome with significant SNPs from earlier rounds added as covariates in the model. This model has earlier been shown to have equivalent power with a bi-variate analysis for detecting loci with increased/decreased effects across environments[58]. Here, $\bar{y}_{FT10} - \bar{y}_{FT16}$ is the mean flowering time difference between 10°C and 16°C for an individual accession. *X* is the design matrix including a column vector of 1's for the population mean and one or more additional column vectors with the indicator regression variables for the homozygous genotypes at the SNP or SNPs included in the model (coded as 0/2, respectively). *Z* is matrix obtained by singular value decomposition of the *G matrix* ($ZZ^{\mathrm{T}} = G$), where *G* is the genomic kinship matrix estimated from the whole genome marker set using the *ibs* function in the *GenABEL* R-package [59]. The significance-threshold used in the analysis was Bonferroni corrected for ~1.4 million tested SNPs (-log10(p-value)=7.33).

Second, the loci associated with flowering-time in the earlier polygenic analysis were explored for genotype by environment interactions. Two models were fitted to the data:

$$\bar{y}_{ij} = \mathrm{u} + E_j + a_1 b_1 + a_2 b_2 + \cdots a_n b_n + \mathrm{e} \quad (2)$$

$$\bar{y}_{ij} = \mathrm{u} + E_j + a_1 b_1 + a_2 b_2 + \cdots a_{\mathrm{n}} b_n + E_j a_x B_x + \mathrm{e} \quad (3)$$

In both models, $\bar{y}_{ij}$ is the mean flowering time among replicates of individual *i* phenotyped in environment *j* where *j* here take the values 1 and 2 for FT10 and FT16, respectively; a_x is the genotype at QTL *x* coded as 0 and 2 for the homozygous minor and major allele genotypes; *b* are the corresponding estimated effect sizes; *u* is the population mean and E_j is the effect of environment on flowering time. Model 3 includes an interaction term between one of the QTL and the environment *E*, and was fitted for each of the QTL in turn. The significance of

each QTL by environment interaction was assessed using a likelihood ratio test between model 2 and model 3. Population structure was accounted for by the simultaneous fitting of the 48 loci. The analyses were performed using customised R scripts [60].

Evaluating how groups of loci contribute to flowering time variation

Several of the mapped loci are only associated with flowering-time in one of the growth temperatures. We tested whether the flowering-time loci that were not significantly associated individually in one of the temperatures made a significant joint contribution by comparing the fit of the following models to the data (models 4 and 5) using a likelihood ratio test.

$\bar{y} = X_1\beta_1 + \mathrm{e}$ (4)

$\bar{y} = X_1\beta_1 + X_2\beta_2 + \mathrm{e}$ (5)

Here, $\bar{y}$, and e are defined as for model 1. Contributions from the individually significant/non-significant loci in the tested temperature were modelled as $X_1\beta_1/X_2\beta_2$, respectively. X_1 is a similar design matrix as X in model 1, which includes a column vector of 1's for the population mean and one or more additional column vectors with the indicator regression variables for the homozygous genotypes at the SNP or SNPs included in the model (coded as 0/2, respectively). X_2 is a second design matix only including the genotype for additional loci unique to FT10 or FT16. β_1/ β_2 are the corresponding effect sizes. A likelihood ratio test was used to compare the fit of the two models using the *lrtest* function in R package *lmtest* [61].

Clustering of subpopulations based on the allele frequencies of flowering time associated loci

Euclidean distances between the 8 subpopulations in the 1,001-genomes dataset defined in [22] were calculated using the *dist* function in R based on the allele frequencies at the 48 loci

553 associated with the flowering time at 10°C or 16°C in this dataset. The subpopulations were
554 then clustered using the hierarchical clustering function *hclust* in R.
555
556 **Estimating the joint, local and global effects of all associated flowering-time loci**
557 The effects of all loci associated with flowering-time in 10°C, 16°C or both temperatures (as
558 in model 4) were estimated simultaneously using the *hglm* function in *hglm* R-package [62].
559 $\bar{y} = \mathrm{X}\beta + \mathrm{Z}u + \mathrm{e}$ (6)
560 All parameters are the same as in model 1, except X in which a column vector of 1s is
561 included to model the population mean and then n genotype columns (where the genotype for
562 an individual at a locus is coded as 0 and 2 as above) for each of the n mapped flowering-time
563 loci. The fitted values from model 6 were then used to model the contributions by global and
564 local alleles to the flowering time. Contributions from the local alleles were estimated by
565 $X_1\beta_1$, where X_1 is the genotype matrix of the local alleles and β_1 is the corresponding
566 estimated effects from model (6). Contributions from the global alleles were estimated by
567 $X_2\beta_2$, where X_2 is the genotype matrix of the local alleles and β_2 is the corresponding
568 estimated effects from model (6). Contributions from the kinship were then estimated by
569 subtracting $X_1\beta_1$ and $X_2\beta_2$, from the modelled phenotype. Then, estimates for each
570 subpopulation were obtained by averaging the individual estimates obtained for the
571 accessions in it.
572
573 **Testing for significant long-range Linkage Disequilibrium**
574 An empirical significance threshold was derived to test for significant long-range LD between
575 the flowering-time loci associated with flowering time at 10°C, 16°C or both. An empirical
576 null distribution was obtained using 1,000 simulations, where in each the same number loci as

detected in the association analysis were simulated with allele frequencies being the same as those of the associated loci. In each simulated dataset, all pairwise LD (D') values were calculated and saved to generate a null distribution for the significance test. We used the 0.99 quantile (corresponding to a D' value of 0.96) as cut-off in our analyses. The FDR for the result was calculated as the ratio between the expected number and observed numbers of significant pairs (chose (n,2)*0.01), where n is the number of associated loci.

Visualization of the results

Figs 3 and 5 were created using the *heatmap.2* function in *gplots* package in R [63]. Fig 6 was created using the *LDheatmap* function in the *LDheatmap* package in R [64]. Fig S3 was made using the R package *maptools* [65]. All other figures were created using custom R-scripts [60].

References

1. Pardo-Diaz C, Salazar C, Jiggins CD. Towards the identification of the loci of adaptive evolution. Methods Ecol Evol. 2015;6: 445–464. doi:10.1111/2041-210X.12324
2. Olson-Manning CF, Wagner MR, Mitchell-Olds T. Adaptive evolution: evaluating empirical support for theoretical predictions. Nat Rev Genet. 2012;13: 867–77. doi:10.1038/nrg3322
3. Rockman M V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution. 2012;66: 1–17. doi:10.1111/j.1558-5646.2011.01486.x
4. Fournier-Level a., Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek a. M. A map of local adaptation in Arabidopsis thaliana. Science. 2011;334: 86–89. doi:10.1126/science.1209271

5. Shen X, De Jonge J, Forsberg SKG, Pettersson ME, Sheng Z, Hennig L, et al. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. PLoS Genet. 2014;10: e1004842. doi:10.1371/journal.pgen.1004842

6. Forsberg SKG, Andreatta ME, Huang XY, Danku J, Salt DE, Carlborg Ö. The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. PLoS Genet. 2015;11: e1005648. doi:10.1371/journal.pgen.1005648

7. Barboza L, Effgen S, Alonso-Blanco C, Kooke R, Keurentjes JJB, Koornneef M, et al. Arabidopsis semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. Proc Natl Acad Sci. 2013;110: 15818–23. doi:10.1073/pnas.1314979110

8. Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015;518: 371–375. doi:10.1038/nature14181

9. Nadeau NJ, Jiggins CD. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet. 2010;26: 484–92. doi:10.1016/j.tig.2010.08.004

10. Flint J, Mackay TFC. Genetic architecture of quantitative traits in mice, flies, and humans. Genome Research. 2009. pp. 723–733. doi:10.1101/gr.086660.108

11. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. Science. 2009;325: 714–718. doi:10.1126/science.1174276

12. Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN. Evidence of widespread selection on standing variation in Europe at height-associated

SNPs. Nat Genet. 2012;44: 1015–1019. doi:10.1038/ng.2368

13. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14: 807–20. doi:10.1038/nrg3522

14. Teotónio H, Chelo IM, Bradić M, Rose MR, Long AD. Experimental evolution reveals natural selection on standing genetic variation. Nat Genet. 2009;41: 251–257. doi:10.1038/ng.289

15. Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol Ecol. 2012;21: 4931–41. doi:10.1111/j.1365-294X.2012.05673.x

16. Burke MK, Liti G, Long AD. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of saccharomyces cerevisiae. Mol Biol Evol. 2014;31: 3228–3239. doi:10.1093/molbev/msu256

17. Pritchard JK, Di Rienzo A. Adaptation - not by sweeps alone. Nat Rev Genet. 2010;11: 665–667. doi:10.1038/nrg2880

18. Graves JL, Hertweck KL, Phillips MA, Han M V, Cabral LG, Barter TT, et al. Genomics of parallel experimental evolution in Drosophila. Mol Biol Evol. 2017;34: 831–842. doi:10.1093/molbev/msw282

19. Collins S, De Meaux J, Acquisti C. Adaptive walks toward a moving optimum. Genetics. 2007;176: 1089–1099. doi:10.1534/genetics.107.072926

20. Kopp M, Hermisson J. The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. Genetics. 2009;182: 233–49. doi:10.1534/genetics.108.099820

21. Yeaman S. Local adaptation by alleles of small effect. Am Nat. 2015;186: S74–S89.

649 doi:10.1086/682405

650 22. Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KMM, et al.
651 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana.
652 Cell. 2016;166: 481–491. doi:10.1016/j.cell.2016.05.063

653 23. Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, et al. Role of
654 *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of
655 Arabidopsis. Plant Physiol. 2005;138: 1163–1173. doi:10.1104/pp.105.061309.1

656 24. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, et al. Genome-
657 wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature.
658 2010;465: 627–631. doi:10.1038/nature08800

659 25. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, et al. Linkage and
660 association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genet.
661 2010;6: 40. doi:10.1371/journal.pgen.1000940

662 26. Zan Y, Carlborg Ö. A multilocus association analysis method integrating phenotype
663 and expression data reveals multiple novel associations to flowering time variation in
664 wild-collected Arabidopsis thaliana. Mol Ecol Resour. 2018; 1–11. doi:10.1111/1755-
665 0998.12757

666 27. Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. Epistatic
667 interaction between Arabidopsis FRI and FLC flowering time genes generates a
668 latitudinal cline in a life history trait. Proc Natl Acad Sci. 2004;101: 15670–15675.
669 doi:10.1073/pnas.0406232101

670 28. Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, et al. A
671 latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering
672 time gene FRIGIDA. Proc Natl Acad Sci. 2004;101: 4712–4717.

673 doi:10.1073/pnas.0306401101

674 29. Samis KE, Heath KD, Stinchcombe JR. Discordant longitudinal clines in flowering
675 time and Phytochrome C in Arabidopsis thaliana. Evolution. 2008;62: 2971–2983.
676 doi:10.1111/j.1558-5646.2008.00484.x

677 30. Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M. Missing G x E Variation Controls
678 Flowering Time in Arabidopsis thaliana. PLOS Genet. 2015;11: e1005597.
679 doi:10.1371/journal.pgen.1005597

680 31. Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO. Genetic variation for life history
681 sensitivity to seasonal warming in Arabidopsis thaliana. Genetics. 2014;196: 569–577.
682 doi:10.1534/genetics.113.157628

683 32. Stratton DA. Reaction norm functions and QTL-environment interactions for flowering
684 time in Arabidopsis thaliana. Heredity. 1998;81: 144–155. doi:10.1038/sj.hdy.6883690

685 33. Allard RW, Babbel GR, Clegg MT, Kahler a L. Evidence for coadaptation in Avena
686 barbata. Proc Natl Acad Sci. 1972;69: 3043–8. doi:10.1073/pnas.69.10.3043

687 34. Mitton JB, Linhart YB, Hamrick JL, Beckman JS. Observations on the genetic
688 structure and mating system of ponderosa pine in the Colorado front range. Theor Appl
689 Genet. 1977;51: 5–13. doi:10.1007/BF00306055

690 35. Barker JSF, Frydenberg J, Sarup P, Loeschcke V. Altitudinal and seasonal variation in
691 microsatellite allele frequencies of Drosophila buzzatii. J Evol Biol. 2011;24: 430–439.
692 doi:10.1111/j.1420-9101.2010.02180.x

693 36. Liang G, He H, Li Y, Wang F, Yu D. Molecular Mechanism of microRNA396
694 Mediating Pistil Development in Arabidopsis. Plant Physiol. 2014;164: 249–258.
695 doi:10.1104/pp.113.225144

696 37. Golembeski GS, Imaizumi T. Photoperiodic Regulation of Florigen Function in

Arabidopsis thaliana. Arab B. 2015;13: e0178. doi:10.1199/tab.0178

38. Hewezi T, Maier TR, Nettleton D, Baum TJ. The Arabidopsis MicroRNA396-GRF1/GRF3 Regulatory Module Acts as a Developmental Regulator in the Reprogramming of Root Cells during Cyst Nematode Infection. PLANT Physiol. 2012;159: 321–335. doi:10.1104/pp.112.193649

39. Lee H, Yoo SJ, Lee JH, Kim W, Yoo SK, Fitzgerald H, et al. Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. Nucleic Acids Res. 2010;38: 3081–3093. doi:10.1093/nar/gkp1240

40. Teotia S, Tang G. To bloom or not to bloom: Role of microrna s in plant flowering. Mol Plant. 2015;8: 359–377. doi:10.1016/j.molp.2014.12.018

41. Wang JW. Regulation of flowering time by the miR156-mediated age pathway. Journal of Experimental Botany. 2014. pp. 4723–4730. doi:10.1093/jxb/eru246

42. Choi K, Kim J, Hwang H-J, Kim S, Park C, Kim SY, et al. The FRIGIDA Complex Activates Transcription of *FLC* , a Strong Flowering Repressor in *Arabidopsis* , by Recruiting Chromatin Modification Factors. Plant Cell. 2011;23: 289–303. doi:10.1105/tpc.110.075911

43. Martin-Trillo M. EARLY IN SHORT DAYS 1 (ESD1) encodes ACTIN-RELATED PROTEIN 6 (AtARP6), a putative component of chromatin remodelling complexes that positively regulates FLC accumulation in Arabidopsis. Development. 2006;133: 1241–1252. doi:10.1242/dev.02301

44. Bond DM, Wilson IW, Dennis ES, Pogson BJ, Jean Finnegan E. Vernalization insensitive 3 (VIN3) is required for the response of Arabidopsis thaliana seedlings exposed to low oxygen conditions. Plant J. 2009;59: 576–587. doi:10.1111/j.1365-313X.2009.03891.x

45. Hall A. The TIME FOR COFFEE Gene Maintains the Amplitude and Timing of Arabidopsis Circadian Clocks. PLANT CELL ONLINE. 2003;15: 2719–2729. doi:10.1105/tpc.013730

46. Kim J, Kim Y, Yeom M, Kim J-H, Nam HG. FIONA1 Is Essential for Regulating Period Length in the Arabidopsis Circadian Clock. Plant Cell. 2008;20: 307–319. doi:10.1105/tpc.107.055715

47. Zhang B, Wang L, Zeng L, Zhang C, Ma H. Arabidopsis TOE proteins convey a photoperiodic signal to antagonize CONSTANS and regulate flowering time. Genes Dev. 2015;29: 975–987. doi:10.1101/gad.251520.114.

48. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc Natl Acad Sci. 2010;107: 21199–21204. doi:10.1073/pnas.1007431107

49. Dittmar EL, Oakley CG, Ågren J, Schemske DW. Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. Mol Ecol. 2014;23: 4291–4303. doi:10.1111/mec.12857

50. Rus A, Baxter I, Muthukumar B, Gustin J, Lahner B, Yakubova E, et al. Natural variants of AtHKT1 enhance Na+ accumulation in two wild populations of Arabidopsis. PLoS Genet. 2006;2: 1964–1973. doi:10.1371/journal.pgen.0020210

51. Salomé PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D. Genetic architecture of flowering-time variation in Arabidopsis thaliana. Genetics. 2011;188: 421–433. doi:10.1534/genetics.111.126607

52. Gould BA, Stinchcombe JR. Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. Mol Ecol. 2017;26: 92–106. doi:10.1111/mec.13643

745 53. Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. Genomic
746 basis and evolutionary potential for extreme drought adaptation in Arabidopsis
747 thaliana. Nature Ecology and Evolution. 2017: 1–7. doi:10.1038/s41559-017-0423-0
748 54. Tilman D, Lehman C. Human-caused environmental change : Impacts on plant
749 diversity and evolution. Proc Natl Acad Sci. 1987;98: 5433–5440.
750 doi:10.1073/pnas.091093198
751 55. Kawakatsu T, Huang S shan??Carol, Jupe F, Sasaki E, Schmitz RJ, Urich MA, et al.
752 Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell.
753 2016;166: 492–506. doi:10.1016/j.cell.2016.06.044
754 56. Consortium 1001genomes. FT10. In: 1,135 Genomes Reveal the Global Pattern of
755 Polymorphism in Arabidopsis thaliana. [Internet]. 1 Jan 2016.
756 doi:10.21958/PHENOTYPE:261
757 57. Consortium 1001genomes. FT16. In: 1,135 Genomes Reveal the Global Pattern of
758 Polymorphism in Arabidopsis thaliana. [Internet]. 1 Jan 2016.
759 doi:10.21958/PHENOTYPE:262
760 58. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model
761 approach for genome-wide association studies of correlated traits in structured
762 populations. Nat Genet. 2012;44: 1066–1071. doi:10.1038/ng.2376
763 59. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: An R library for
764 genome-wide association analysis. Bioinformatics. 2007;23: 1294–1296.
765 doi:10.1093/bioinformatics/btm108
766 60. R Core Team. R Core Team, 2015 R: A Language and Environment for Statistical
767 Computing. R Found. Stat. Comput. Vienna Austria URL: https://www.R-project.org/.
768 [Internet]. R Foundation for Statistical Computing Vienna Austria. 2015. doi:ISBN 3-

900051-07-0

61. Zeileis A, Hothorn T. Diagnostic checking in regression relationships. R News. 2002;2: 7–10. Available: http://cran.r-project.org/doc/Rnews/

62. Rönnegård L, Shen X, Alam M. hglm: A Package for Fitting Hierarchical Generalized Linear Models. R J. 2010;2: 20–28. Available: http://journal.r-project.org/archive/2010-2/RJournal_2010-2_Roennegaard~et~al.pdf

63. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. gplots: Various R Programming Tools for Plotting Data. In: R package version 2.17.0. [Internet]. 2015 p. 2015. doi:10.1111/j.0022-3646.1997.00569.x

64. Shin J-H, Blay S, Graham J, McNeney B. LDheatmap : An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. J Stat Softw. 2006;16. doi:10.18637/jss.v016.c03

65. Lewin-Koh N, Bivand R. Maptools: Tools for reading and handling spatial objects [Internet]. R package version …. 2011. p. 74. Available: http://cran.r-project.org/package=maptools

Figures and figure captions

Fig 1. Flowering time measured at 10 and 16 °C, correlation with latitude/temperature, and the subpopulation differentiation. (**A**). Histogram of flowering time measured at 10°C (blue) and 16 °C (tomato). Flowering time under controlled green house condition is significantly correlated both with latitude (**B**) and temperature (**C**) at the sampling sites, and vary amongst the earlier defined subpopulations (**D**), of the 1,004 analysed accessions from the 1,001-genomes *A. thaliana* collection [22].

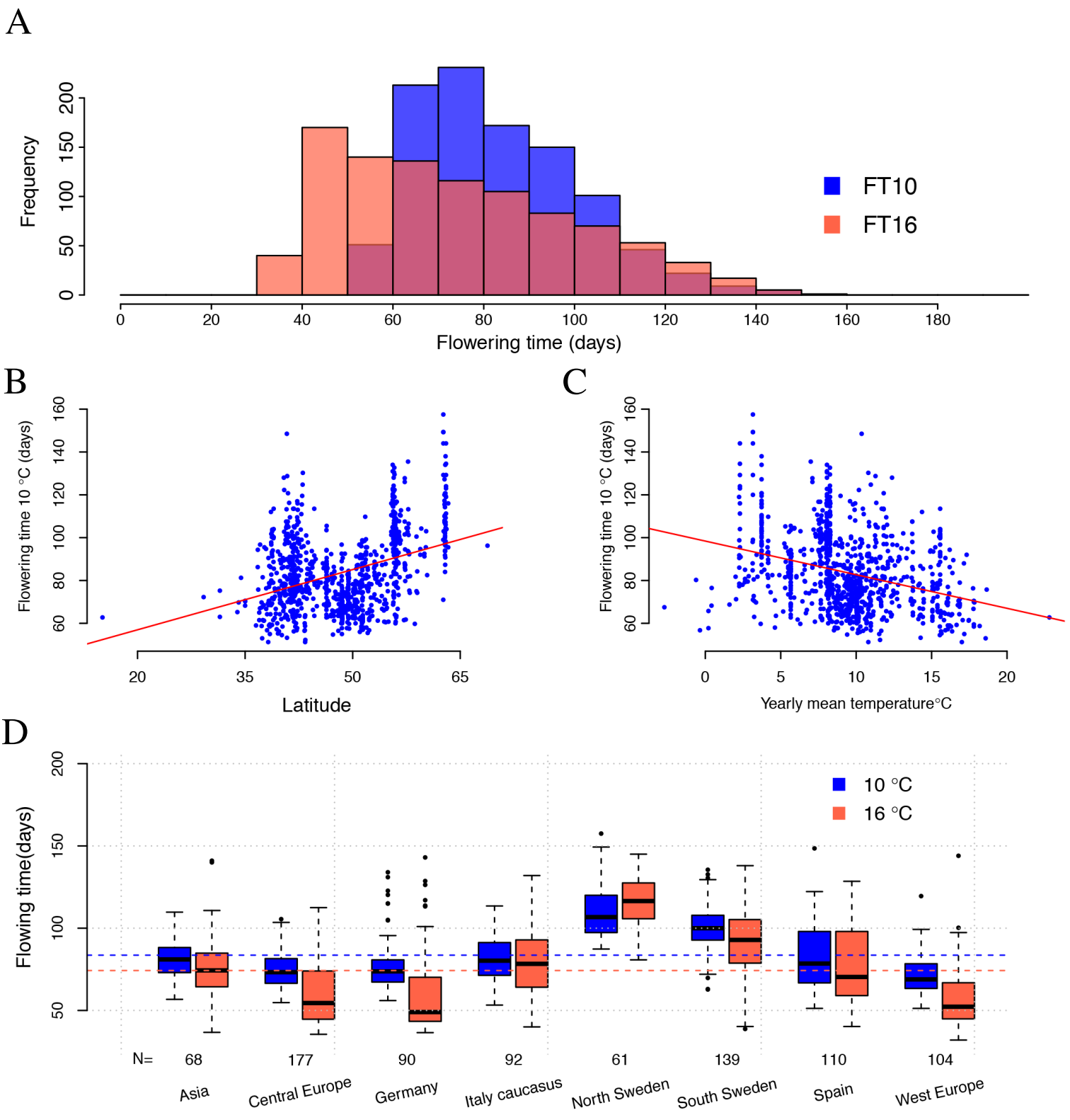
Fig 2. Additive effects of 48 loci on flowering time measured at 10 and 16 °C. (**A**) Additive genetic effects of 48 flowering time loci on FT10 (blue) and FT16 (tomato). The y-axis indicates the additive effects on flowering time and the x-axis the respective loci (SNPs) sorted by the differences in additive effects (FT16-FT10) from left to right. The estimates were obtained by fitting all 48 loci jointly in a mixed model. The bars represent standard errors of the effect estimates. The 18 loci with significant genotype by temperature interactions are indicated with a black star. (**B**) Correlation between the additive genetic effects for the 48 loci estimated for accessions grown at 10 and 16 °C. The colour of the dots indicates whether the locus is significant i) only for FT10 (blue), only for FT16 (tomato), for both FT10 and FT16 (purple), or the difference between the two (FT10-FT16; black).

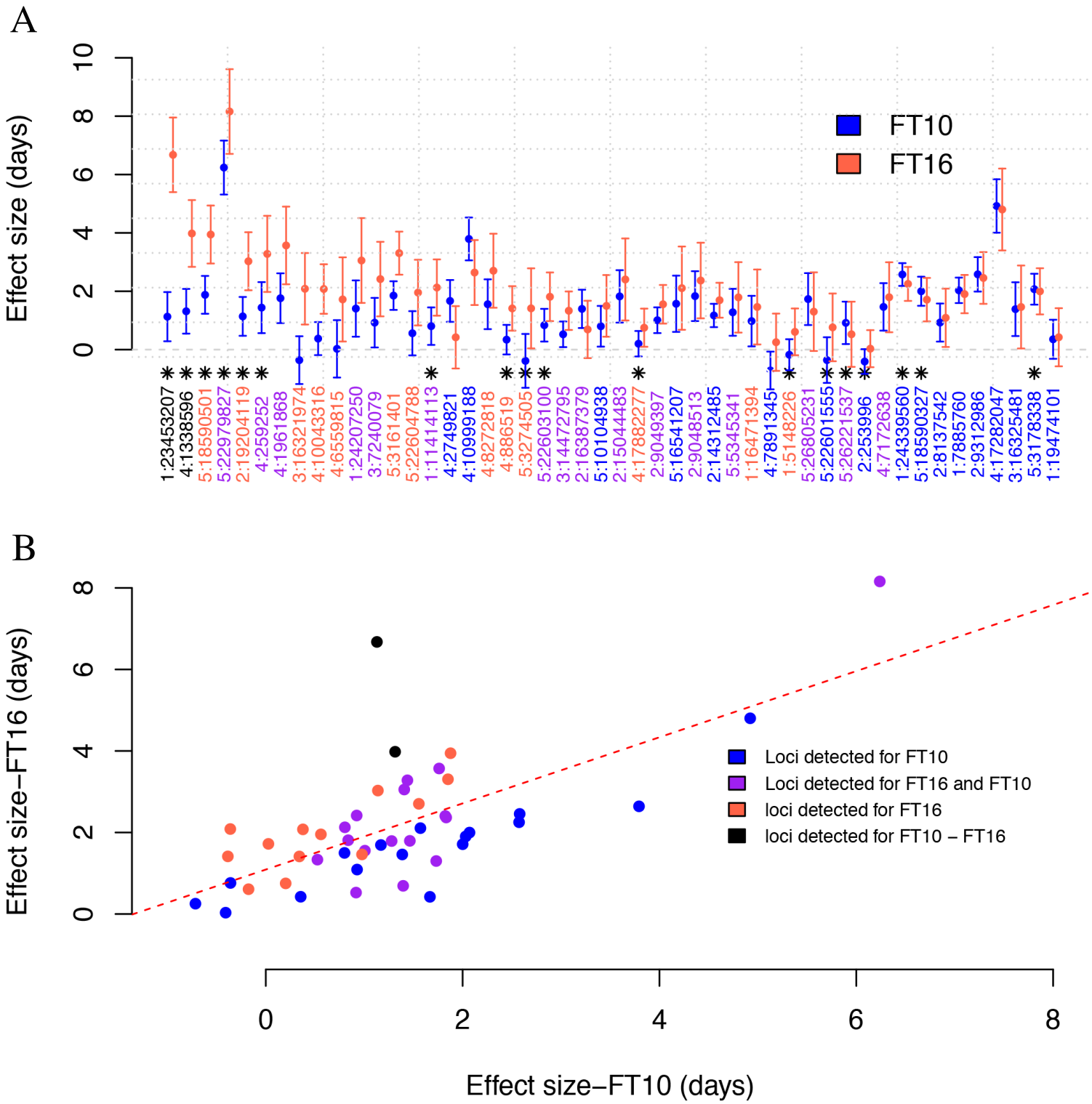
Fig 3. Flowering times, allele frequencies of flowering time associated loci and similarities of these between eight subpopulations of the *A. thaliana* 1,001-genomes collection. (**A**) Flowering times for the 8 subpopulations in this data defined by [22] are modelled well by the additive effects of the 48 flowering time loci and the kinship. (**B**) Distribution of allele frequencies for the 48 loci in the 8 subpopulations. Intensity of

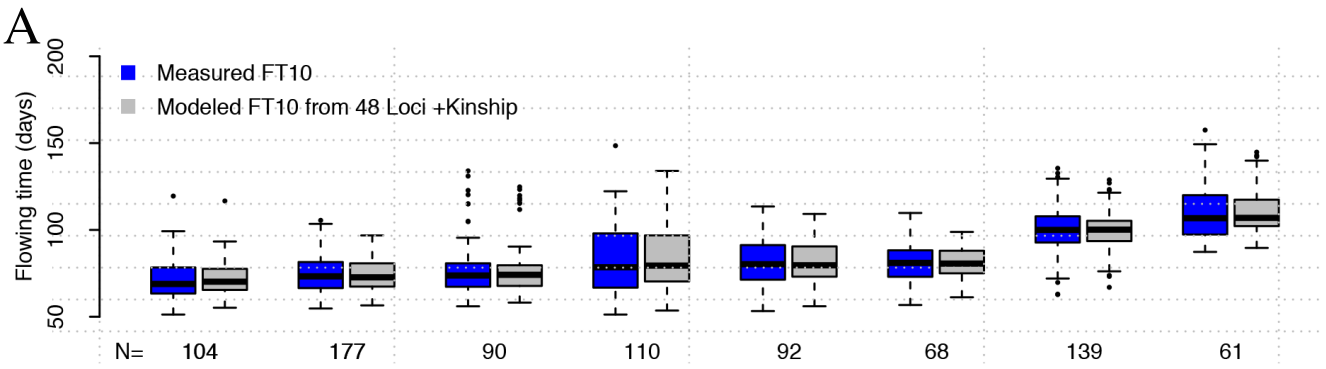
blue/yellow illustrates frequency of early/late-flowering alleles, respectively. The subpopulations are sorted from left to right by increasing average flowering time (columns) and the loci from top to bottom by decreasing additive effects on flowering time (rows). X:YYYYY-ZZZ represent a locus as Chromosome:Position-Type, where Position is in bp and Type is either "QTL", "Diff" or "locus-id" indicating its detection by direct association to FT10, FT16 or (FT10-FT16) in a GWAS ("QTL"/"Diff") or via a genetically regulated expression of the locus "locus-id" in the eQTL analysis. (**C**) Dendrogram showing how the 8 subpopulations cluster based on the allele frequency spectrum across the 48 associated loci.

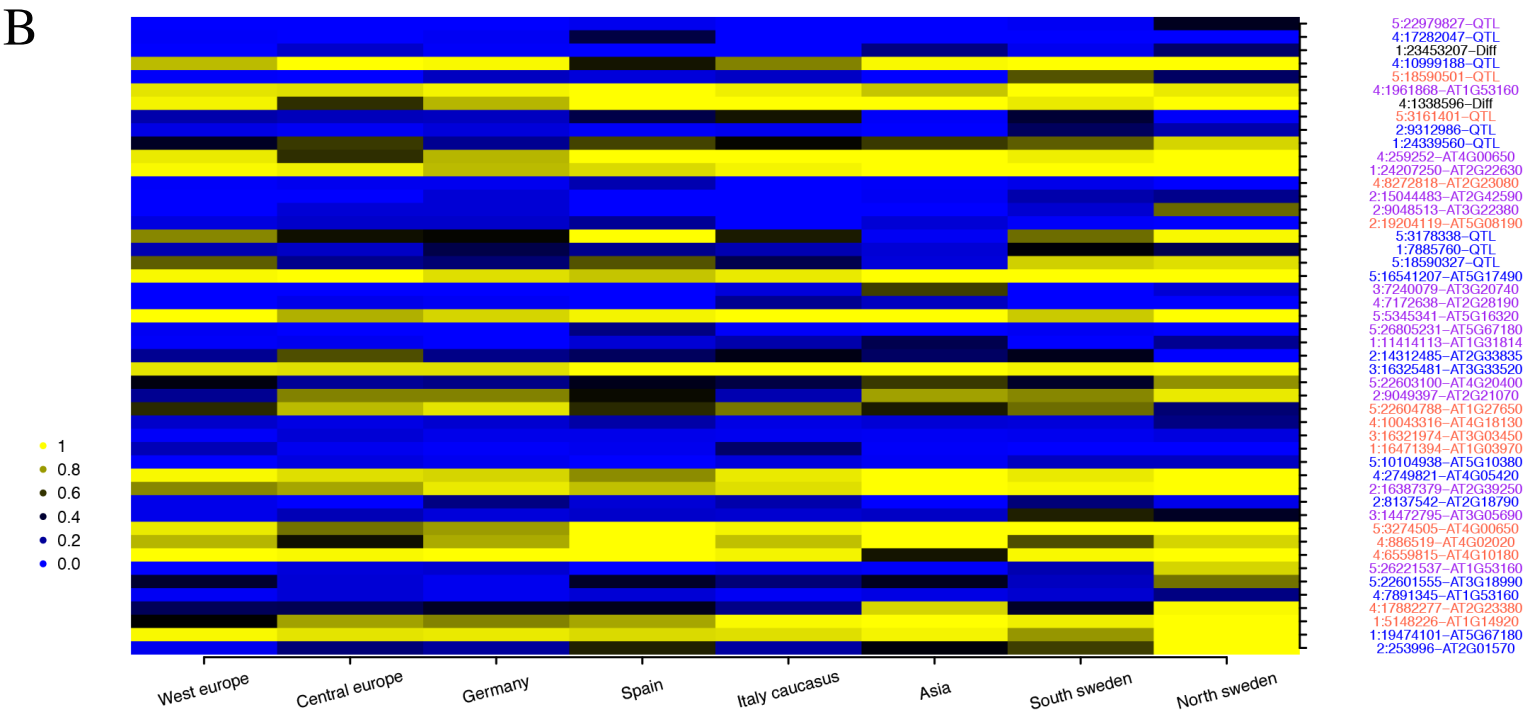
Fig 4. Modelling of the flowering times of the eight subpopulations in the 1,001-genomes *A. thaliana* collection and how the individual associated loci contribute. Experimentally measured and modelled flowering times in the 8 genetically defined subpopulations of the *A. thaliana* 1,001-genomes collection of accessions [22] grown at 10°C (**A**) and 16 °C (**B**). The height of the bars represents the experimentally observed mean flowering times for the accessions belonging to the sub population. The kinship (purple) and additive effects of globally segregating alleles (grey) capture the overall pattern of early to late flowering times between the sub populations well. The additive effects of the local alleles (blue) improve the fit to more extreme adaptations of some sub populations, in particular in Spain and Sweden. The differences between measured and modelled flowering times (tomato) are small at both temperatures. The modelled contributions by the global alleles are larger at 16°C than at 10°C due to their genotype-by-temperature interactions (Fig. 2A). The relationship between the variance contributed by each individual locus and its allele frequency/additive effect is illustrated for FT10 (**C**) and FT16 (**D**). The y-axis gives the minor allele frequency and the x-axis the estimated additive effect on flowering time in days. The size of each dot is

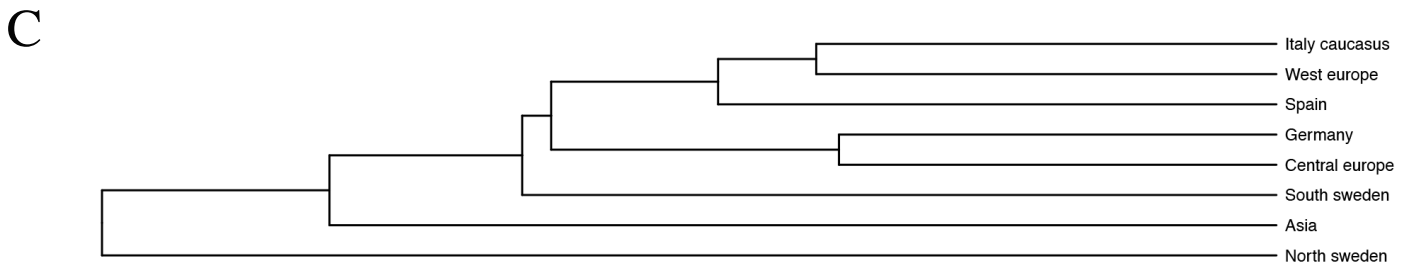
834 proportional to the phenotypic variance explained by the locus. The colour of the dot indicates
835 whether the minor allele is defined as a global (tomato) or local (blue) allele in the population.
836
837 **Fig 5. Allele frequencies for loci where one variant is rare (MAF < 0.05) in at least one of**
838 **the eight 1,001-genomes *A. thaliana* subpopulations**. Blue/yellow indicates high allele
839 frequency for the early/late flowering alleles, respectively. The populations are clustered
840 based on the overall similarity of the allele frequency pattern (left to right) and the loci based
841 on the similarity of the allele frequencies in the subpopulations (top to bottom). X:YYYYY-
842 ZZZ represent a locus as chromosome: position-type, where Position is in bp and Type is
843 either "QTL", "Diff" or "locus-id" indicating its detection by direct association to FT10,
844 FT16 or (FT10-FT16) in a GWAS ("QTL"/"Diff") or via a genetically regulated expression
845 of the locus "locus-id" in the eQTL analysis.
846
847 **Fig 6. Pairwise linkage disequilibrium (LD) between the 48 flowering time-associated**
848 **loci across the genome (D').** Values above the 0.99 quantile of a simulated null distribution
849 are highlighted as yellow. The loci are labelled as chromosome: position on chromosome in
850 bp.



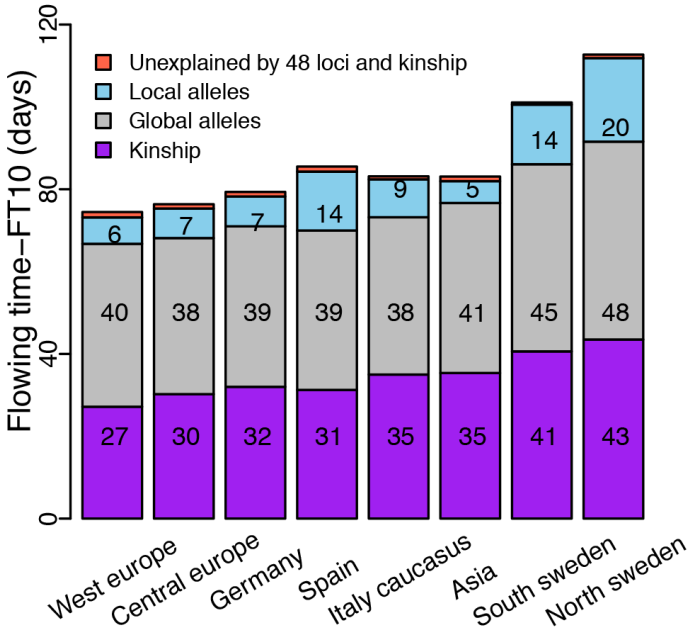




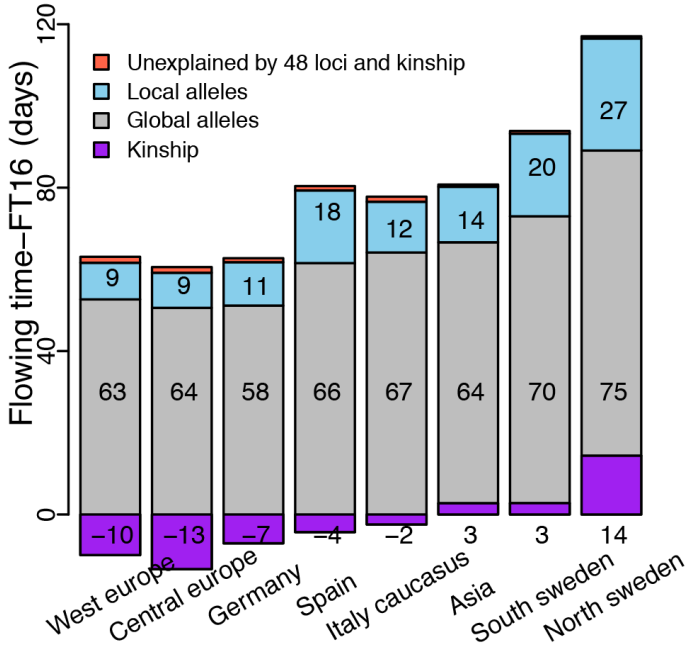




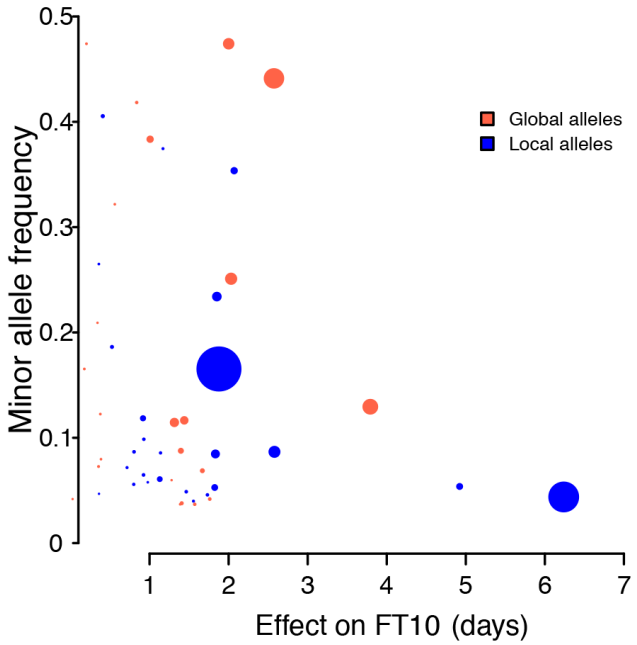
A



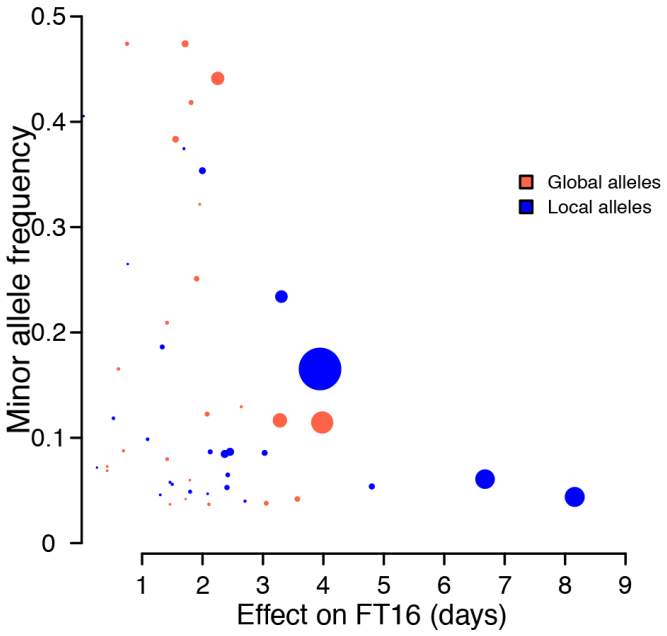
B

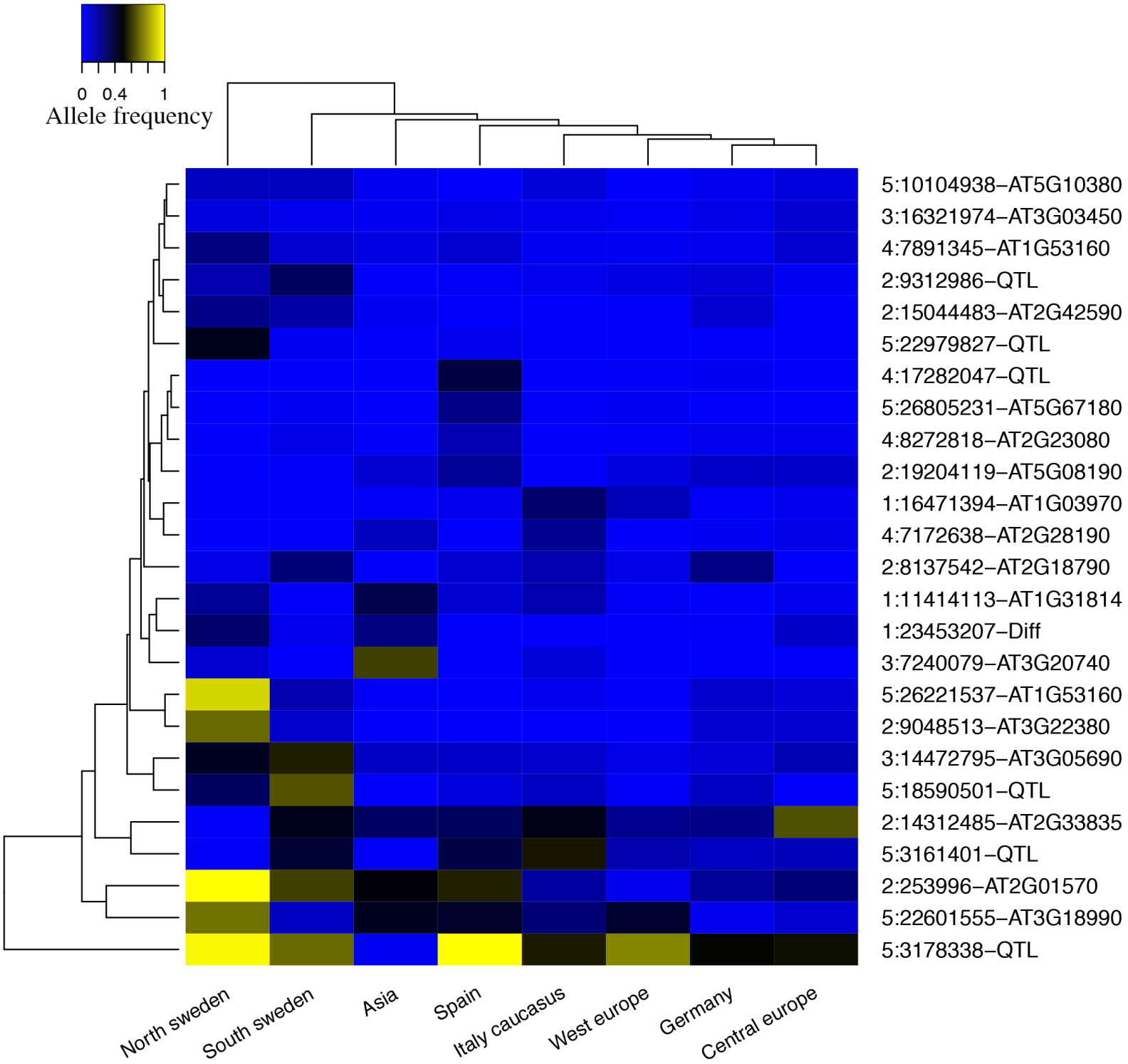


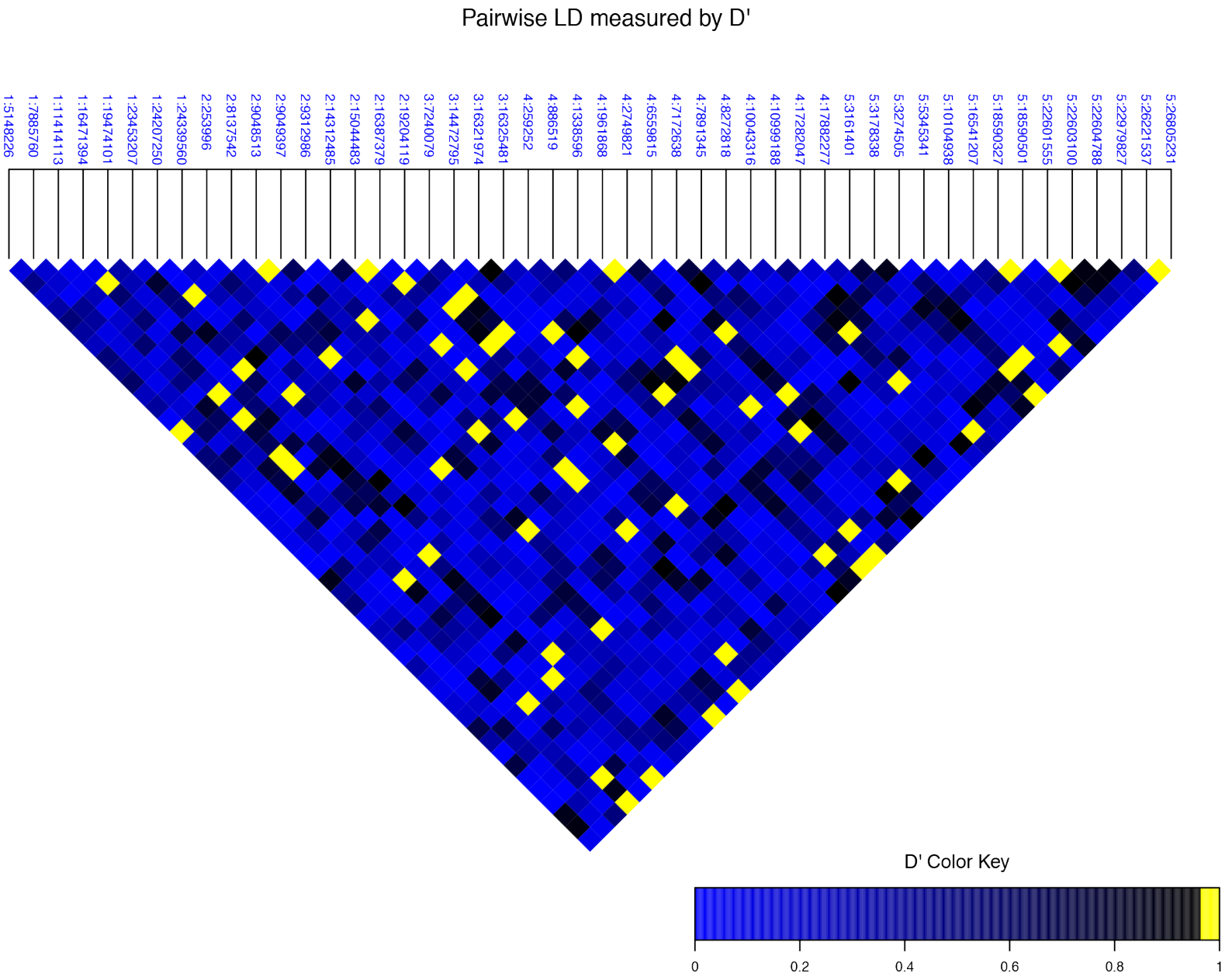
C



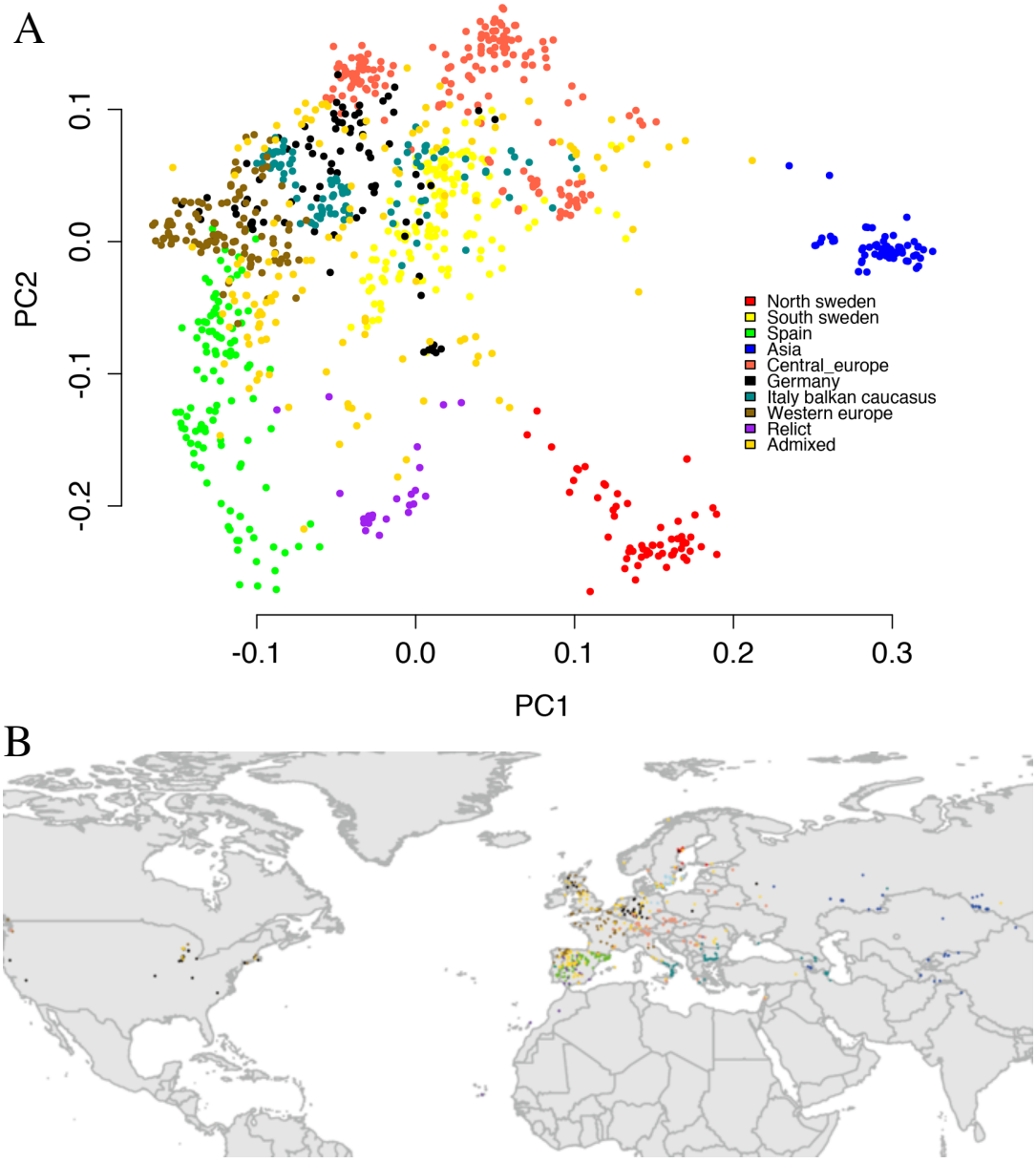
D



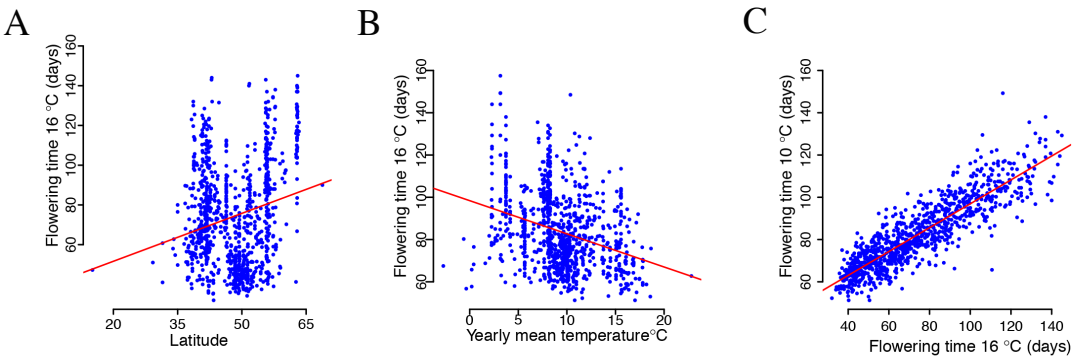




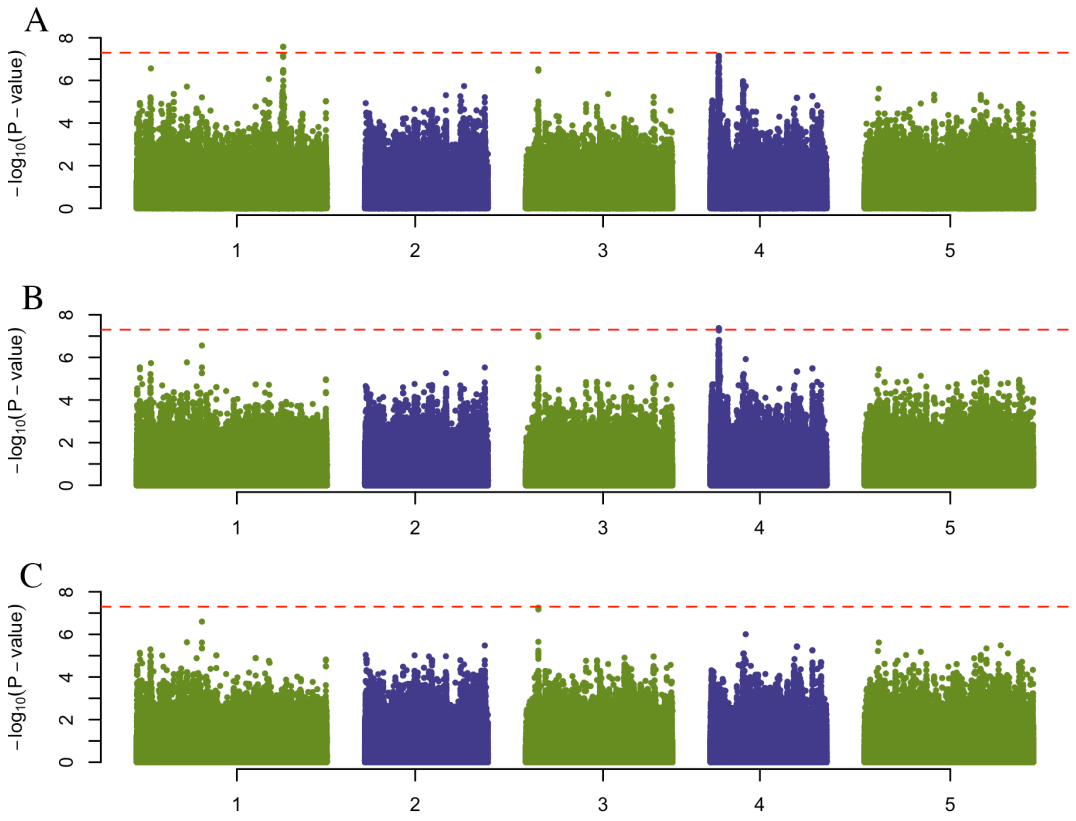
Supplementary information for "**Explorations of the polygenic genetic architecture of flowering time in the worldwide *Arabidopsis thaliana* population**" by Zan and Carlborg.



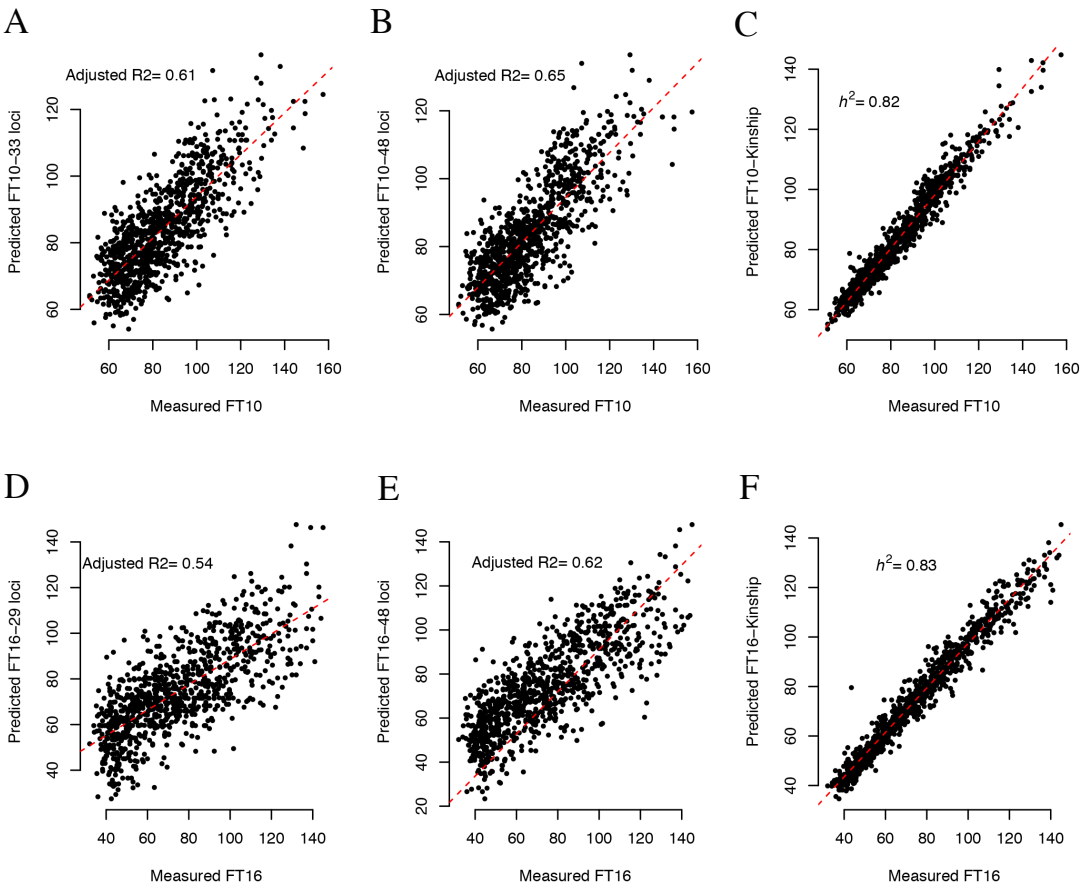
S1 Fig. **Population structure and sampling locations for the accessions in the 1,001-genomes *A. thaliana* collection.** (**A**) Population structure indicated by the first two principal components of the IBS (Identity By State) kinship matrix with the subpopulations defined by [22] highlighted in colour. (**B**) Sampling locations of the analysed accessions.



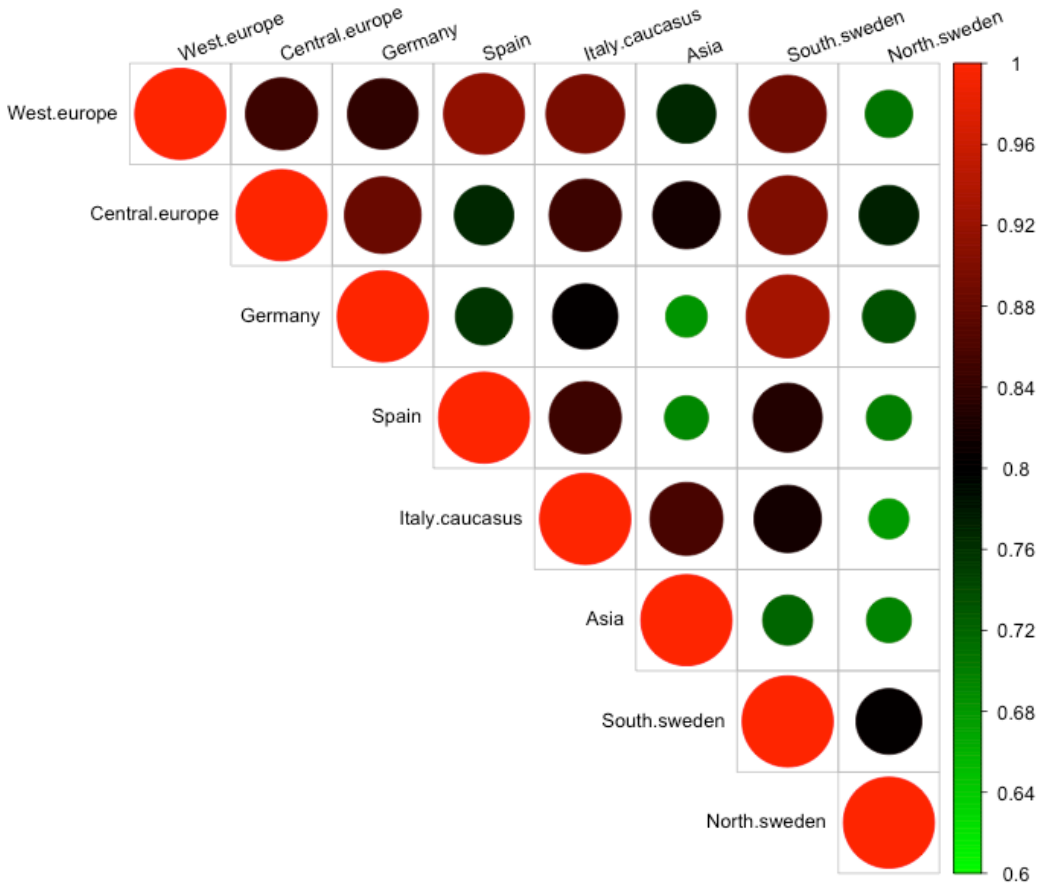
S2 Fig. **Geographical differentiation, and correlations, of the flowering times for the accessions of the 1,001-genomes *A. thaliana* collection grown at 10 and 16°C.** Flowering times measured for accessions grown at 16°C in the green house are significantly correlated both with latitude (**A**) and temperature (**B**) at the sampling sites for the 1,004 analysed accessions from the 1,001-genomes *A. thaliana* collection [22]. (**C**) There is a high correlation between flowing times measured for accessions grown at 10 and 16°C.



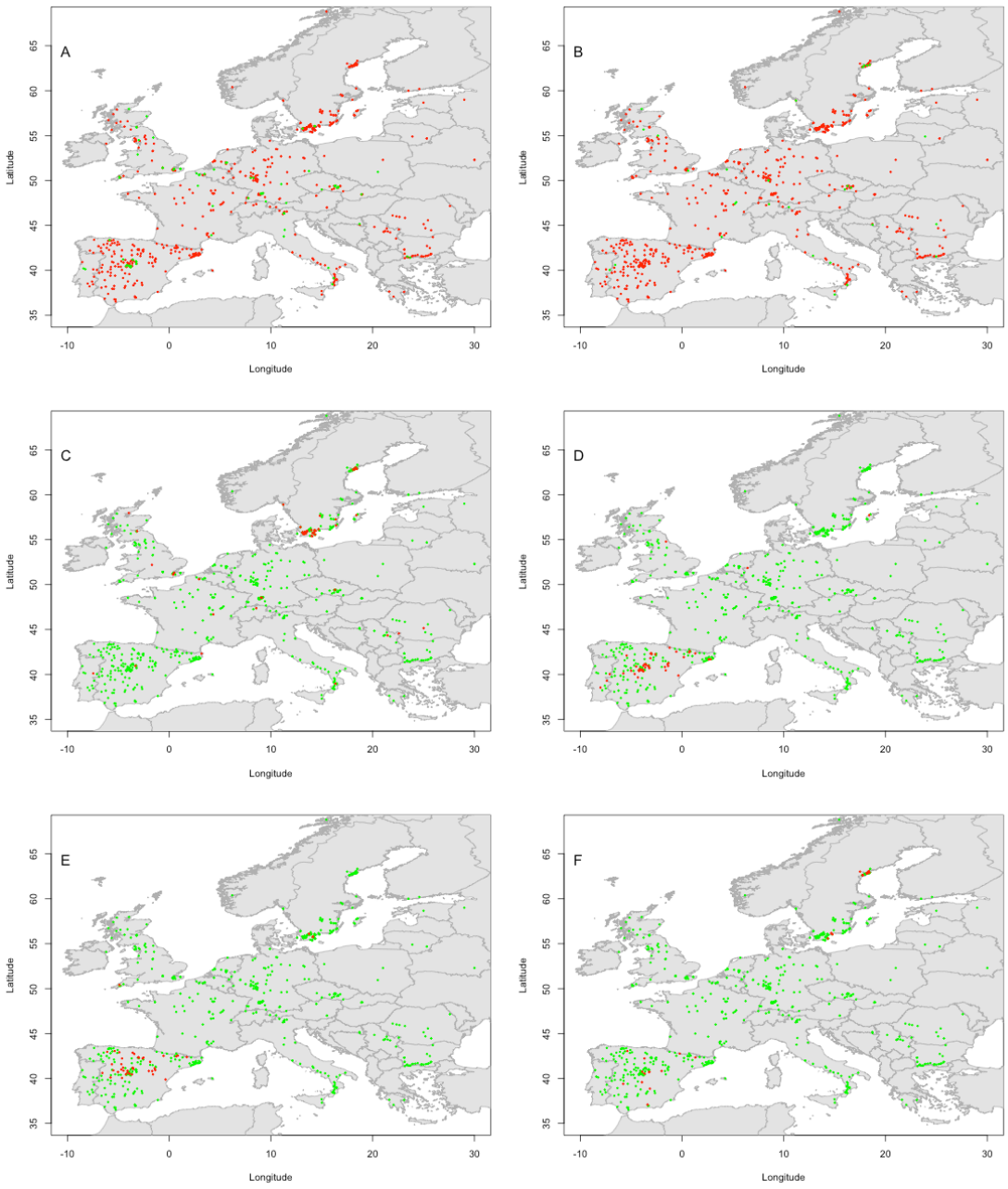
S3 Fig. Results from the GWAS analysis using the differences in flowering times at 10 and 16°C for the accessions of the 1,001-genomes *A. thaliana* collection. (A) Manhattan plot for a GWAS using FT10-FT16 as phenotype where one locus on Chromosome 1 at 23,453,207 bp reached genome wide significance. (**B**) Manhattan plot for a GWAS using FT10-FT16 as phenotype with the locus identified in (**A**) as covariate, where a second locus on Chromosome 4 at 1,338,596 bp reached genome wide significance. (**C**) Manhattan plot for a GWAS using FT10-FT16 as phenotype with the 2 loci identified in (**A**) and (**B**) as covariates, where no additional locus was detected.



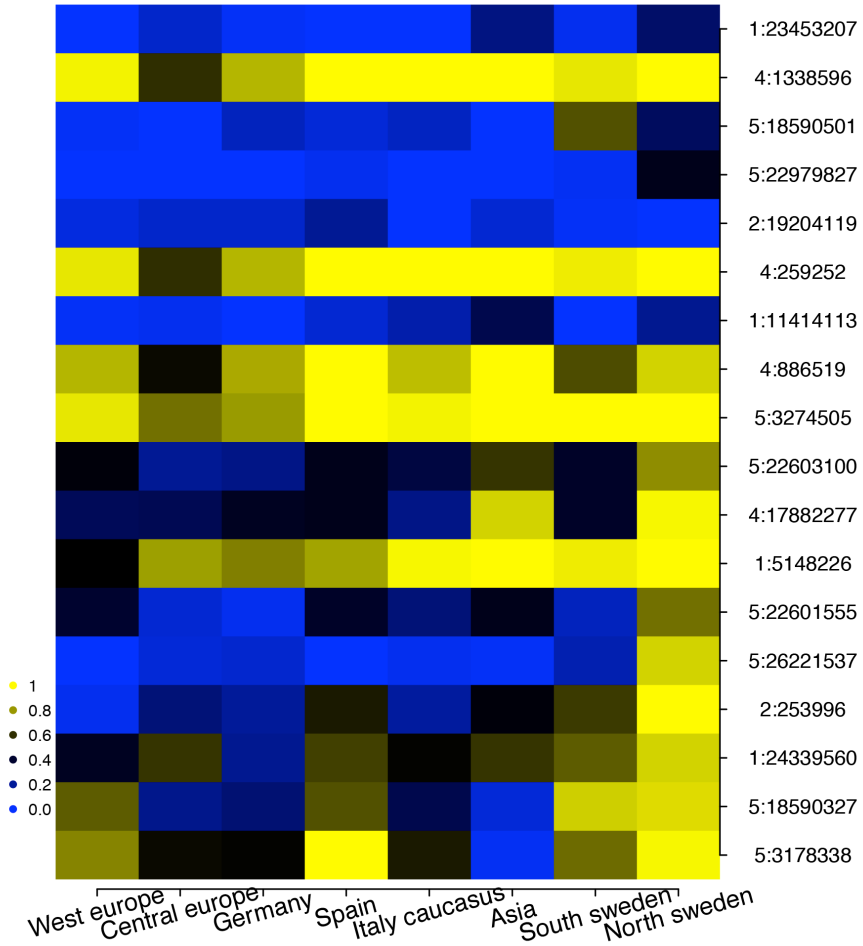
S4 Fig. Modelled and experimentally measured flowering times for the accessions of the 1,001-genomes *A. thaliana* collection. Modelled phenotypes for the 1,001-genomes accessions using the 33 (29) loci detected in the independent analyses of FT10/FT16 and the experimentally measured phenotypes are illustrated in (**A**) and (**D**). Modelled flowering times using all 48 flowering time loci and measured phenotypes are illustrated in (**B**) and (**E**). Modelled flowering times from kinship and measured phenotypes for FT10/FT16 are illustrated in (**C**) and (**F**).



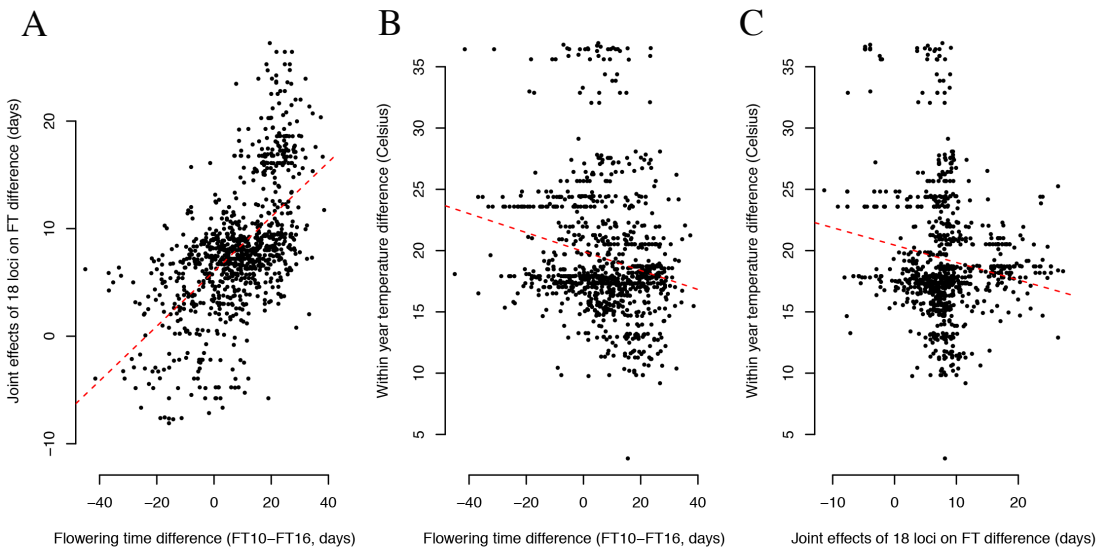
S5 Fig. Illustrations of the overall similarities in the allele frequency spectrum across the sub populations as Spearman correlation between the allele-frequencies across the 48 associated loci



S6 Fig. The geographical distribution of alleles at 6 selected loci associated with flowering-time at 10°C (FT10). The five panels illustrates the geographical distributions for the six associated loci at Chromosome 2:16,387,379 bp (**A**), Chromosome 4:1,961,868 bp (**B**), Chromosome 2: 9,312,968 bp (**C**), Chromosome 4:1,728,204 bp (**D**), Chromosome 5:26,805,231bp (**E**) and Chromosome 5:2,297,9827 bp (**F**). Red/green colours indicate the allele that increases/decreases FT10, respectively.



S7 Fig. Allele frequencies for the 18 loci showing significant genotype by temperature interactions across the 8 sub populations of the 1,001-genomes collection. Blue/yellow indicates high allele frequency for early/late flowering alleles, respectively. The populations are sorted based on early to late flowering (left to right) and large to small additive effects on the flowering time difference (FT10-FT16; top to bottom). Genotype by temperature interacting alleles that have a larger effects on FT16 than on FT10 are enriched in the North Sweden sub population.



S8 Fig. Correlations between observed and modelled flowering time differences for accessions grown at 10°C and 16°C and the within year temperature fluctuations at their sampling locations. **(A)** The correlation between the measured flowering time difference (FT10-FT16) and the flowering time difference modelled by the joint effects the 18 loci with significant genotype by temperature interactions (Adjusted R-square = 0.34; P-value = 1.3×10^{-90}; Fig 2A). **(B)** There is a significant correlation (Adjusted R-squared = 0.04; P-value = 1.7×10^{-10}) between the within year temperature fluctuation (temperature difference between hottest and coldest day) at the sampling locations of the accessions [22] and the measured flowering time difference (FT10-FT16). **(C)** There is also a significant correlation (Adjusted R-squared = 0.03; P-value = 1.2×10^{-7}) between the within year temperature fluctuation (temperature difference between hottest and coldest day) [22] and the modelled flowering time differences for the accessions from the joint effects of the 18 loci with significant genotype by temperature interactions (Fig 2A).