Local adaptation of *Arabidopsis thaliana* in a small geographic region with mild environmental clines

Raúl Y. Wijfjes[1,6,*], René Boesten[2*], Frank F. M. Becker[2], Tom P. J. M. Theeuwen[2], Basten L. Snoek[3], Maria Mastoraki[2], Jelle J. Verheijen[2], Nuri Güvencli[2], Lissy-Anne M. Denkers[4], Maarten Koornneef[2], Fred A. van Eeuwijk[5], Sandra Smit[1], Dick de Ridder[1,†], and Mark G.M. Aarts[2,†]

[1] Bioinformatics Group, Wageningen University & Research
[2] Laboratory of Genetics, Wageningen University & Research
[3] Theoretical Biology and Bioinformatics, Utrecht University
[4] Department of Plant Physiology, Green Life Science Research Cluster, Swammerdam Institute for Life Sciences, University of Amsterdam
[5] Biometris, Wageningen University & Research
[6] Current affiliation: Faculty of Biology, Ludwig Maximilian University of Munich
[*] These authors contributed equally to this work.

Abstract

Natural populations of *Arabidopsis thaliana* provide powerful systems to study adaptation of wild plant species. Previous research has predominantly focused on global populations or accessions collected from regions with diverse climates. However, little is known about the genetics underlying adaptation in regions with mild environmental clines. We have examined a diversity panel consisting of 192 *A. thaliana* accessions collected from the Netherlands, a region with limited climatic variation. Despite the relatively uniform climate, we identified compelling evidence of local adaptation within this population. Notably, semidwarf accessions occur at a relatively high frequency near the coast and these displayed enhanced tolerance to high wind velocities. Additionally, we evaluated the performance of the population under iron deficiency conditions and found that allelic variation in the *FSD3* gene affects tolerance to low iron levels. Moreover, we explored patterns of local adaptation to environmental clines in temperature and precipitation, observing that allelic variation at *LARP1C* likely affects drought tolerance. Not only is the genetic variation observed in a diversity panel of *A. thaliana* collected in a region with mild environmental clines comparable to that in collections sampled over larger geographic ranges, it is also sufficiently rich to elucidate the genetic and environmental factors underlying natural plant adaptation.

Introduction

Adaptation is defined as the process through which a population attains higher fitness in its native environment than non-native populations sampled from different sites (Kawecki & Ebert, 2004). Unraveling the genomic and physiological basis of adaptation in plants is a central question in modern plant biology. From an evolutionary perspective, linking plant genotypes to adaptive traits helps to clarify how environmental gradients drive natural selection of both standing and novel variation, which ultimately may result in speciation (Sobel et al., 2010). It also enables us to model whether natural populations are able to adapt to future environmental conditions (Bay et al., 2017; Exposito-Alonso et al., 2019) and provides insights on processes or even genes to focus on when breeding crops for enhanced abiotic stress tolerance (Huang & Han, 2014).

The model plant species *Arabidopsis thaliana* provides an excellent system to study natural plant adaptation (Weigel & Nordborg, 2015). Recent studies of this species have shifted from the global HapMap population (Y. Li et al., 2010), traditionally used for genome-wide association studies (GWAS), to populations sampled from regions containing a range of contrasting climates, including Sweden (Long et al., 2013), the Iberian Peninsula (Tabas-Madrid et al., 2018), and the south-west of France (Frachon et al., 2018), as these local populations are expected to minimize genetic heterogeneity and thus provide more statistical power for detecting quantitative trait loci (QTLs) (Brachi et al., 2013; Korte & Farlow, 2013). While studies on such regional populations have successfully identified several adaptive loci (Frachon et al., 2018; Tabas-Madrid et al., 2018), we know comparatively little about the loci driving adaptation in regions having a milder diversity in environmental conditions.

The Netherlands, a region which has remained largely under-sampled so far, provides an excellent opportunity to address this issue. This region covers an area comparable in size to the south-west of France, but much smaller than Sweden and the Iberian Peninsula. Compared to the other three regions, the Netherlands has only mild climatic clines (https://www.knmi.nl/klimaat-viewer), but past work indicated that local adaptation of *A. thaliana* may still be expected (Barboza et al., 2013). The Netherlands contains a moderate frequency of accessions with short inflorescences, mediated by the same loss-of-function allele of the *GA5* gene that managed to spread over more than 100 kilometres in the western part of the country (Barboza et al., 2013). There is reason to believe that more such signatures of local adaptation can be found, for instance based on the small differences in average temperature and annual precipitation, or the soil type on which *A. thaliana* grows in the Netherlands (De Vries et al., 2003). While it can be found on the clay and peat regions in the west and north of the country, it mainly grows on sandy soils, which largely occur in the east and south, and in the coastal dunes. It is often found in roadsides and gardens in which substantial movement of soils and seeds from other regions has occurred in the past.

To comprehensively address the contribution of genetic variation to plant adaptation in the Netherlands, we generated the Dutch *Arabidopsis thaliana* Map (DartMap) panel, a collection of 192 *A. thaliana* accessions sampled from different sites in the country. We show that this panel contains a surprisingly high level of genetic diversity and provide evidence of local adaptation to mild climatic clines. Moreover, we present strong evidence that variation at two loci is involved in mediating adaptation to wind tolerance and iron deficiency. Our work demonstrates that plant populations sampled in small geographic regions with a low diversity in environmental conditions can contain a level of genetic variation that is sufficiently rich to facilitate local adaptation.

Results

The degree of genetic diversity in the Netherlands is typical of that of most European regions

The global *A. thaliana* collection of the 1001 Genomes (1001G) Project (Alonso-Blanco et al., 2016) includes only 11 lines from the Netherlands, which likely represents only a small portion of the region's genetic variation due to limited dispersal of minor alleles (Tabas-Madrid et al., 2018). Therefore, we used the DartMap panel and the Dutch 1001G accessions to characterize the overall degree of genetic diversity of *A. thaliana* in the Netherlands. In total, we identified 2,712,612 SNPs and 353,974 indels in the DartMap panel and the Dutch 1001G accessions (collectively referred to as DartMap + 1001G panel for brevity) relative to the *A. thaliana* Col-0 nuclear genome reference sequence. Moreover, we detected 487 SNPs and 209 indels in the chloroplast sequence, and 137 SNPs and 15 indels in the mitochondrial sequence.

Furthermore, we examined copy number variation in the DartMap panel, excluding the Dutch 1001G accessions because they were sequenced with a different sequencing platform. We detected 29,155 copy number variants (CNVs) relative to the Col-0 reference genome. Most of these are less than 500 bp (Figure S1a) and are deletions (Figure S1b), which is in line with the distribution of structural variants found in an earlier study on the full 1001G collection (Göktay et al., 2021). We found that 448 genes overlap with CNVs predicted to disrupt gene function and present at moderate frequencies within the DartMap panel (Table S1) (considering deletions only, as we could confirm that these were called with high precision (Figure S2)). The 448 genes are significantly enriched for genes involved in disease resistance ($P = 4.39\times10^{-6}$).

To contextualize the degree of genetic diversity of the DartMap + 1001G panel, we compared it to similarly sized collections from Sweden (Long et al., 2013) and the Iberian peninsula (Tabas-Madrid et al., 2018), using the same SNP and indel calling approach. Considering only SNPs and indels that are polymorphic in the DartMap + 1001G panel (i.e. at least one accession contains a reference allele), the Swedish collection contains a similar number of variants, while the Iberian one is more diverse (Table 1). This disparity can be attributed to the presence of "relict" accessions in the Iberian collection, which originated from *A. thaliana* populations in Europe before the last ice age and are genetically distinct from the more recent "non-relict" group that repopulated Europe thereafter (Alonso-Blanco et al., 2016).

Most SNPs and indels detected in the DartMap + 1001G panel are shared among multiple accessions and the distribution of their minor allele frequencies is similar to that of the Swedish collection (Figure 1a). In contrast, the Iberian collection contains a relatively larger fraction of rare variants (Figure 1a), possibly because the mountain systems in this region strongly limit the spread of minor alleles (Tabas-Madrid et al., 2018).

To assess the genetic diversity of the DartMap + 1001G panel, we calculated pairwise genetic distances between individual accessions. Most pairs of accessions differ at approximately 320,000-340,000 variant sites, similar to the Swedish collection and most pairs of the Iberian collection (Figure 1b). However, the Iberian collection shows a right-skewed distribution of pairwise genetic distances, indicating the presence of accessions that are genetically more distant from each other compared to the majority (Figure 1b). This tail in the distribution aligns with previous findings of genetic distance between Iberian accessions and attributed to the presence of the relict accessions in this

region (Alonso-Blanco et al., 2016). Taken together, our analyses indicate that genomic variation of *A. thaliana* in the Netherlands predominantly originated from the post-glacial expansion event from which most *A. thaliana* accessions in Europe descended (C.-R. Lee et al., 2017).

Genetic diversity of *A. thaliana* in the Netherlands was shaped by ancient and contemporary forces

We explored the demographic and evolutionary history of the DartMap + 1001G panel using population structure analyses based on the genetic distance between individual accessions, which have proven useful for this purpose in previous studies of *A. thaliana* (Alonso-Blanco et al., 2016; Tabas-Madrid et al., 2018). We computed population structure of the DartMap + 1001G panel through hierarchical clustering based on pairwise organellar (Figure 2a) and nuclear (Figure 2b) genome-wide similarity, including all accessions of the 1001G collection sampled in Belgium, France, Germany, and the United Kingdom to evaluate relatedness between Dutch *A. thaliana* accessions and those of nearby countries. The dendrograms based on chloroplast and mitochondrial variation consist of three nearly equidistant groups (Figure S3) with identical spatial distributions across the Netherlands (Figure 2a), as expected for organellar sequences that are both predominantly maternally inherited.

A previous study showed that European accessions contain several highly diverged chloroplast haplogroups of which the estimated time of divergence considerably predates the post-glacial expansion event (Hsu et al., 2019). A principal component analysis based on chloroplast variation between individual accessions of the DartMap + 1001G panel and of a global collection of *A. thaliana* separate Dutch accessions in three different groups (Figure 2c) that correspond to different subsets of chloroplast haplogroups identified in that same study (Figure 2d). This implies that the distinct organellar groups of the DartMap + 1001G panel represent ancient variation, while the nuclear groups reflect more recent genomic differentiation, explaining the limited overlap in spatial distribution between the two (Figures 2a and b).

The spatial distribution of the nuclear groups provides insights into the factors that shaped genomic variation in the DartMap + 1001G panel following the postglacial expansion event. Based on nuclear variation, the panel can be divided into two main groups, predominantly separated by the central Rhine and Waal rivers (group 1 to the north and group 2 to the south) (Figure 2b). The rivers likely acted as natural dispersion barriers (Pico et al., 2008), restricting genetic flow between the two groups. It is worth noting that *A. thaliana* is generally not found in clay soil accompanying the rivers, except in residential areas where it has been introduced through human-induced transport of soil from other regions. Group 1 can be further divided into three subgroups (Figure 2b), of which two contain accessions that are closely related to each other (subgroups 1a and 1b) and less related to the rest of the population (Figure S4). Subgroup 1a contains nine accessions with short inflorescences, referred to as “semidwarfs”, mainly concentrated in the western part of the Netherlands (Figure 2b). Subgroup 1b is predominantly found on the Dutch islands in the north, implying that these islands were colonized by a limited number of founder accessions, mostly from group 1. Group 2 consists of accessions that are genetically similar to accessions from Germany (subgroup 2a) respectively France and the United Kingdom (subgroup 2b) (Figure S5), suggesting that these are part of a large genetic group spread over Northwestern Europe.

Besides large clusters, we identified 27 pairs of accessions that are near duplicates in terms of nuclear similarity (Table S2), with differences of fewer than 50,000 sites between them. This level of similarity is significantly lower than the median pairwise similarity of 354,388 sites (Figure 1a). Strikingly, they include 8 out of 9 pairs of the semidwarf accessions of group 1a, which are separated by more than 100 km (Table S2). The widespread geographical distribution of highly identical genotypes that all have the same semidwarf phenotype suggests there may be a selective advantage to this trait, similar to the near-identical and widespread atrazine-resistant genotypes distributed along the United Kingdom railway system, which we identified previously (Flood, Van Heerwaarden, et al., 2016).

Semidwarf *ga5* mutants display tolerance to windy conditions

A previous study indicated a higher frequency of semidwarfs in the Netherlands compared to the global estimated frequency of at least 1% (Barboza et al., 2013). In our analysis, we identified ten semidwarf accessions in the DartMap panel, three of which overlapped with those from the previous study (Barboza et al., 2013). Almost all examined semidwarf accessions are loss-of-function mutants at the *GA5* locus, which encodes the *GIBBERELLIN 20-OXIDASE 1* (*GA20OX1*) gene (Barboza et al., 2013). Further investigation of the genomic sequences of *GA5* revealed that eight semidwarf accessions are genotypically nearly identical and share a previously described *ga5* splice site mutation specific to Dutch semidwarfs (Barboza et al., 2013). Of the remaining two accessions, one contains an undescribed 52 bp deletion in the last exon of *GA5*, while the other contains no mutations in the *GA5* coding region.

The eight near-identical semidwarfs that share the same splice site mutation are part of nuclear subgroup 1a (Figure 2b), whereas the semidwarf accession without *GA5* coding sequence mutations is genetically distinct and part of subgroup 1b (Figure 2b). Subgroup 1b consists of five accessions found exclusively on the Dutch islands in the North, with four of them displaying relatively short inflorescence lengths (among the top 15% shortest inflorescence length in the DartMap). These four accessions share a unique 25 bp insertion at 557 bp upstream of the *GA5* transcription start, while the fifth accession in this subgroup has an average inflorescence length and lacks this mutation. Both subgroups (1a and 1b) are genetically most related to each other and both are found in the West of the Netherlands (Figure 2b). The appearance of two distinct *ga5* mutations in accessions with short inflorescences dispersed across the Netherlands suggests that semidwarfism provides an adaptive advantage.

Dwarfism in plants is a common adaptation to high altitude (Billings & Mooney, 1968; Grace, 1988; Luo et al., 2015), although it remains unclear which aspect of high altitude causes this. The majority of the Dutch semidwarfs are collected within 15 km of the coast, often only a few meters above sea level, so there is no association with high altitude. However, these semidwarfs share a characteristic with their original collection locations, namely the presence of high wind speeds, particularly between December and March (Figure 3a). Given that high wind speeds are also a characteristic of high-altitude environments, we aimed to investigate whether the presence of high wind speeds could potentially create a selective environment favoring semidwarf *A. thaliana*.

We grew six Dutch semidwarf accessions and five tall accessions, in a climate chamber equipped to provide controlled wind speeds typically experienced in coastal areas of the Netherlands in March/April. Moreover, we included Col-0, Ler and the *ga5* loss-of-function mutants in Col-0 and Ler backgrounds for a direct comparison of near isogenic lines differing only at the *GA5* locus. Despite our preselection for similar

flowering time (Table S3), based on greenhouse data, the semidwarfs flowered significantly later than the taller accessions (Figure 3b), and the wind treatment therefore affected the semidwarfs for a shorter period while bolting and flowering than the tall accessions. This likely had little effect on our final results though, as all accessions were exposed to windy conditions for the same time and all plants started to flower before the end of the experiment. The tall accessions are strongly affected by the wind, as evidenced by a significant reduction in stem length (approximately 30% on average) and fresh weight (approximately 36% on average) (Figure 3c and 3d). In contrast, the semidwarfs also showed a significant reduction in stem length, but the average reduction was only around 10% and no significant decrease in fresh weight was observed (Figure 3d). Similar effects were observed for Col-0 and Ler, where the wind treatment significantly decreased their stem length and fresh weight, whereas the respective *ga5* loss-of-function mutants remained unaffected (Figure 3c and d). Collectively, these results strongly suggest that the presence of the *ga5* mutation contributes to wind tolerance.

Flowering time variants display a latitudinal cline

The transition from vegetative to reproductive stage is a key life history trait in *A. thaliana*, influenced by seasonal cues like day length and temperature (Andrés & Coupland, 2012). Previous GWA analyses of flowering time yielded few significant associations, mainly due to allelic heterogeneity (Atwell et al., 2010; Tabas-Madrid et al., 2018; L. Zhang & Jiménez-Gómez, 2020). We expect this should be less of a problem in a regional population with more closely related individuals. Therefore, we performed GWA analysis of flowering time, with and without vernalization, using the DartMap panel. We discovered a highly significant association between flowering time and a 16 bp indel near the end of the first exon of the *FRIGIDA* (*FRI*) gene (position 269960 on chromosome 4 of Col-0), irrespective of vernalization treatment (Figure 4a). This indel is known to cause a premature stop codon in the *FRI* coding sequence, leading to a non-functional *FRI* allele (Andrés & Coupland, 2012; Johanson et al., 2000; Koornneef et al., 1994).

FRI is a key regulator of flowering time in *A. thaliana* and loss-of-function alleles are widespread in nature (L. Zhang & Jiménez-Gómez, 2020). As the variant we picked up by GWA analysis is unlikely to be the only loss-of-function mutation to occur in the Netherlands, we examined all polymorphisms in the *FRI* coding sequences (CDS) in the DartMap and found a total of 41 *FRI* alleles. When compared to the *FRI-H51* allele, that encodes a fully functional protein and is regarded as the ancestral *FRI* allele based on sequence comparisons with *A. lyrata* (Le Corre et al., 2002; L. Zhang & Jiménez-Gómez, 2020), another 11 loss-of-function *fri* alleles are identified across 88 accessions. Nearly all of the accessions with a loss-of-function allele are early flowering (Figure 4b). Although the Netherlands has a small latitudinal range, we examined if there is a latitudinal cline in the diversity panel regarding flowering time, consistent with previous findings in Northern European and Mediterranean accessions (Stinchcombe et al., 2004). While we did not observe such a cline when considering the entire population, excluding accessions with *fri* loss-of-function alleles revealed a latitudinal cline (Figure 4c), suggesting the involvement of other loci, apart from *FRI*, in controlling flowering time. These analyses collectively demonstrate the suitability of the DartMap panel for conducting GWA analyses on phenotypic variation relevant to local adaptation.

Local adaptation to the Dutch climate

The Dutch climate is characterized as a temperate maritime climate, which is relatively uniform across the Netherlands. However, there are mild and gradual clines in environmental variables, with slightly higher temperatures and less precipitation in inland areas compared to coastal areas. One advantage of having a densely and uniformly sampled population is the potential to detect whether local adaptation to these gradual environmental clines has influenced genetic diversity. To test this hypothesis, we conducted a genome-environment association (GEA) analysis on 19 quantitative climatic variables related to temperature and precipitation (Table S4 and Fig. 5a as example). We identified multiple QTLs for these variables, but the most notable QTL is located on Chr. 4 (Fig. 5b), which we named *qDPT* (QTL for the Dutch Precipitation and Temperature climatic variables). This QTL is associated with seven climatic variables (Table S4).

Further examination of *qDPT* reveals that it appears to consist of two separate QTLs, approximately 230,000 base pairs apart (Figure S6a). The SNPs that are found most frequently with the highest association in each QTL are SNP $4^{16770900}$ and SNP $4^{17000635}$ with minor allele frequencies (MAF) of 14.2% and 17.5%, respectively. To determine whether the -$\log(P)$ peaks represent distinct QTLs or a single one, we calculated pairwise r^2 values between SNP $4^{16770900}$ and all other biallelic SNPs in the selected genomic region (16,700,000–17,200,000 bp). This analysis revealed only weak evidence of linkage disequilibrium (LD) (pairwise $r^2 = 0.52$) between the two SNPs most strongly associated with the trait (Figure S6b). Hence, we consider these to be two separate QTLs. For convenience, we will refer to them as '*qDPT-A*' and '*qDPT-B*', with their corresponding SNPs labelled as A/a for SNP $4^{16770900}$ and B/b for SNP $4^{17000635}$.

The GEA analyses provide insights into whether specific alleles are correlated with particular climatic variables. A strong association at a specific locus can indicate that allelic variation at that locus is important for local adaptation (Ferrero-Serrano & Assmann, 2019). However, it is important to note that a strong correlation could also arise if the climatic variable aligns with the population structure, potentially resulting in a statistical artifact. We confirmed that this is not the case. First of all, kinship correction is used to perform the GEA analyses, and this should drastically reduce the chance of finding loci that strongly link to the population structure. Secondly, we observe that the geographical distribution of the population structure appears to be mostly affected by the river systems (Figure 2b), which can be regarded as a South – North cline. The climatic variables that map at *qDPT* follow a mostly East – West cline. Moreover, the minor allele of *qDPT-A* is found in five out of six nuclear subclusters (the common allele is found in all subclusters), and for *qDPT-B* both alleles are found all subclusters. In conclusion, identification of these QTLs is not an artifact due to the general population structure and is likely to represent a locus associated with plant adaptation to the environment.

It is not immediately apparent to which specific climatic variable(s) this locus contributes. In the Netherlands, the climate is primarily influenced by the North Sea, which affects various environmental factors such as temperature, wind speed and precipitation patterns and variability. Therefore, it is not surprising that the same peak is consistently observed, given the high correlation between climatic variables associated with *qDPT*. Consequently, unravelling the particular climatic variable or combination of variables to which this specific locus may be adaptive is challenging. However, it is worth noting that the climatic variables associated to *qDPT* are all connected to seasonality, while annual means in precipitation and temperature are not associated. This suggests that the adaptation may be related to conditions that prevail during certain parts of the year, such as one or two seasons, or to adaptations that enable the plant to effectively

tolerate greater variability. To explore which climatic factor may be involved in this regard, we tested if allelic variation at *qDPT* resulted in a differential response to a gradual decrease in water availability.

As *qDPT-A* and *qDPT-B* are in close proximity, we selected a set of accessions with combinations of alleles for these QTLs (13 accessions with 'AB', 7 with 'Ab', 1 with 'aB' and 7 with 'ab'; Table S5) aiming to disentangle the individual contributions of the QTLs. Initially, all accessions were grown under well-watered conditions. Subsequently, after two weeks, half of the replicates for each accession were no longer watered for eleven days. We regularly measured the response in projected leaf area (PLA) during this period. Due to an uneven distribution of accessions across the genotypic combinations for these QTLs, we compared accessions with 'Ab' genotypes against those with 'ab' genotypes (7 accessions per group) to assess the effect of *qDPT-A*. For *qDPT-B*, all 28 accessions (14 accessions per *qDPT-B/b* allele) were included in the analysis. Our results indicate a significant effect of *qDPT-B* on the PLA response to drought (two-way ANOVA with *qDPT-B/b* allele and days post drought treatment as variables; $P = 0.003$) (Figure 5c and 5d). However, we did not observe a significant effect for *qDPT-A*. Subsequently, we examined the distribution of *qDPT-B* alleles across the Netherlands and the 1001G panel. We found that SNP $4^{17000635}$ (*qDPT-B*) has a much higher MAF in the Netherlands (MAF = 17.5%) (Figure 5e) compared to 1001G panel (MAF = 3.1%) (Figure 5f). The association scores at *qDPT-B* varied considerably depending on the specific climatic factor being analysed.

We propose that the region encompassing the genomic locus with a LOD > 6, which is consistently observed across different mappings, is the most likely location for the causal gene(s). This region comprises 13 genes spanning from At4g35820 to At4g35940 (Table S6). Within this region, there are a few genetic variants that are in linkage ($r^2 > 0.4$) with SNP $4^{17000635}$ and potentially impact the protein. Particularly noteworthy are two non-synonymous variants in *OSCA4.1* (At4g35870), also known as early-responsive to dehydration stress protein (*ERD4*). The first variant constitutes a change from methionine (neutral-polar) to isoleucine (hydrophobic) at the 10th position, while the second variant changes an isoleucine to leucine (both hydrophobic) at the 641st position. Presently, it is unknown whether these changes affect functionality of this gene. We tested T-DNA insertion lines for 11 of these candidate genes in response to control and drought treatments and measured F_v/F_m as a proxy of plant stress (Maxwell & Johnson, 2000) (Figure 5g). For the remaining two genes, At4g35930 and At4g35940, no T-DNA insertion lines were available. The mutant of *LARP1C* (At4g35890) has significantly lower F_v/F_m values in the drought treatment compared to Col-0, while it does not differ from Col-0 in the control treatment (Figure 5g). Additionally, the mutant performance does not differ significantly from Col-0 in PLA under the control treatment, but is significantly impaired under the drought treatment (Figure 5h). These results pinpoint *LARP1C* as the most prominent candidate gene to underly the difference in response to drought.

Natural allelic variation for *FSD3* affects iron deficiency tolerance

Although the Netherlands is relatively uniform in climate, the soil type composition differs substantially across the country, with clay, peat and sand forming the major types (Hartemink & Sonneveld, 2013). This could be an important driver for local adaptation in the Netherlands, as nutrient composition and particularly nutrient availability (Veer, 2006), differs between soil types. Iron (Fe) in particular is known to vary in availability and shortage of Fe will severely limit plant growth (Hell & Stephan, 2003). We determined the light-use efficiency of photosystem II (PSII) electron transport (Φ_{PSII}) of the DartMap

diversity panel grown under control and Fe-deficient conditions, through chlorophyll fluorescence measurements. Low Φ_{PSII} can serve as an early indicator of Fe deficiency (Terry, 1980), as Fe is an essential component of ironsulphur, and heme proteins that play pivotal roles in photosynthesis and respiration (Hänsch & Mendel, 2009).

The low Fe supply leads to only mild visible symptoms, such as reduced plant growth and mild leaf chlorosis, but causes extensive variation for Φ_{PSII}, much more than in the control condition (Figure 6a). GWA mapping identified a single associated region, on chromosome 5, with the most significantly associated SNP located at position 7,856,132 (Figure 6b). The same locus, and the same SNP, is also found when the average Φ_{PSII} values in the Fe-deficient condition, and the ratio between Φ_{PSII} in Fe deficiency versus control, are used as traits for GWA mapping (Figure S7). To identify candidate genes for the causal allelic variant, we limited the associated region to 22 kb, based on the genome positions of SNPs that are in linkage disequilibrium (LD) ($r^2 > 0.55$) with the most significantly associated SNP. This region is predicted to contain 10 genes (Figure 6c).

We identified three common haplotypes for this genomic region in the DartMap panel. Accessions that share the haplotype with the reference variant associated with $SNP5^{7856132}$ will be referred to as haplotype 1, whereas accessions containing the non-reference variant can be further distinguished into two haplotypes that will be referred to as haplotypes 2 and 3 (Table S7). Haplotype 1 associates with higher Φ_{PSII} in Fe-deficient conditions relative to the non-reference haplotypes, which do not significantly differ in Φ_{PSII} (Figure 6d). The higher Φ_{PSII} of accessions with haplotype 1 correlates with a significantly higher projected leaf area in Fe-deficient conditions than haplotypes 2 and 3 (Figure 6e), while both contain no significant differences in projected leaf area between each other.

In total, 40 sequence variants are in LD with $SNP5^{7856132}$ at a cut-off of $r^2 > 0.55$, of which 27 are intergenic, 7 intronic and 6 exonic (Table S8). Only two of the exonic variants encode a non-synonymous amino acid substitution, but in both cases the substitutions are from a leucine to an isoleucine or vice versa. These are very similar amino acids, and neither of these substitutions, one in *RIBF1* (AT5G23330), the other in *DUF789* (AT5G23380), change the domains predicted by InterPro (Mitchell et al., 2019).

We thus hypothesized that non-coding variants that affect either levels or timing of gene expression should be causing the observed QTL. We measured expression of each candidate gene by means of RT-qPCR in both control and Fe-deficient conditions in shoots of six accessions of each haplotype (Table S8). Two known marker genes for Fe deficiency in leaves, *FSD1* and *FRO3* (Waters et al., 2012), indeed showed the expected lower (*FSD1*) or higher (*FRO3*) expression under the Fe deficiency treatment compared to control (Figure S8a). Two of the 10 candidate genes, *FSD3* and *STE14A* (AT5G23320), are about 2-fold higher expressed in accessions with haplotype 1 compared to accessions with either of the other two haplotypes (Figure 6f and S8b). The Φ_{PSII} in Fe deficiency correlates well with expression of these genes (Figure 6g). None of the eight other candidate genes showed such response of expression between haplotypes and correlation with the Φ_{PSII} in Fe deficiency (Figure S8b).

Since the absolute expression levels of *STE14A* are very low, we focused on *FSD3* as the most likely candidate to underlie the QTL. Two distinct splice variants are described for this gene that influence functionality regarding chloroplast development (S. Lee et al., 2019). We measured the relative gene expression of each splice variant independently, but found highly similar splice variant ratios between the different haplotypes (Figure S8c), indicating that natural variation affects gene expression of *FSD3* only quantitatively and not through altering differential splicing.

Analysis of the non-coding sequences surrounding *FSD3* identified a variable $(TTC)_n/(GAA)_n$ microsatellite repeat directly downstream of the *FSD3* 3' UTR, which lacks either two or four of such repeats in haplotypes 2 and 3 relative to haplotype 1. As there are no differences in either Φ_{PSII} or *FSD3* expression between haplotype 2 and 3 accessions we distinguished two alleles for the *FSD3* locus based on their functionality in Fe deficiency: an Fe deficiency tolerant allele, which has the Col-0 reference sequence, and a sensitive allele lacking two or four $(TTC)_n/(GAA)_n$ repeats. We subsequently confirmed the causal nature of the allelic variation at *FSD3* to explain the QTL using quantitative complementation, in which we test if the difference in Φ_{PSII} between F_1 progeny derived from crosses between accessions with a tolerant allele crossed with both Col-0 and the *fsd3* mutant, and F_1 progeny derived from similar crosses with accessions carrying the sensitive allele. In Fe-deficient conditions, the F_1 progeny derived from the *fsd3* mutant and accessions with a tolerant *FSD3* allele significantly outperform the F_1 progeny derived from accessions with a sensitive *FSD3* allele ($P = 0.046$) (Figure 6h), but not under control conditions (Figure 6i).

Discussion

Here we present the DartMap panel, consisting of 192 Dutch *A. thaliana* accessions and demonstrate its potential for detecting local adaptation. The genetic variation observed in this panel largely falls within the reported diversity of the 1001G collection, which is expected for accessions collected in Northwestern Europe outside the European relict regions (Alonso-Blanco et al., 2016). However, the genetic diversity is surprisingly high considering the relatively small geographical area without remarkable environmental clines. It illustrates the high level of standing genetic variation that exists among local populations of this predominantly self-pollinating species.

Despite, or perhaps due to, the absence of strong environmental clines, the DartMap panel is very suitable for conducting GWAS. The genetic architecture underlying natural adaptation is strongly dependent on the local environment (Brachi et al., 2013; Fournier-Level et al., 2013). Consequently, adaptive loci may exhibit genetic or allelic heterogeneity in regions with contrasting climates, which hampers their detection through GWAS. A milder diversity in environmental conditions can help to alleviate this limitation, as exemplified by the successful identification of the well-known *FRI* gene, which controls flowering time variation. This gene was not previously detected by GWAS in other collections such as the Iberian collection (Tabas-Madrid et al., 2018) or the large 1001G panel (Alonso-Blanco et al., 2016), primarily due to allelic heterogeneity. As such, the DartMap should be regarded as a valuable complement to collections sampled from regions with contrasting climates, and not a replacement. With the availability of several regional diversity panels, it will be easier to identify such adaptive loci when taking into account that each region may favour its own adaptive variation.

Our finding that genes involved in disease resistance are overrepresented among those impacted by CNV is consistent with previous work. A similar phenomenon was found for defense-response genes in a global set of 1,301 natural *A. thaliana* accessions (Göktay et al., 2021) and this same category of genes was found to be enriched in regions highly rearranged between seven diverse *A. thaliana* genomes (Jiao & Schneeberger, 2020). These observations suggest that CNVs may act as a genomic response to biotic stress. The next step would be to investigate whether such structural variants correlate

to variation in the prevalence and abundance of known pathogens of *A. thaliana* (Bartoli et al., 2018).

The prevalence of plants with short inflorescences and the predominance of a near-isogenic genotype at various dispersed locations, strongly indicate an adaptive advantage for *ga5* mutants in the coastal regions of the Netherlands. Allelic heterogeneity and strong selection of one successful, near-isogenic genotype represented multiple times in the DartMap panel was the reason we could not identify the *GA5* locus through GWAS, but only based on prior knowledge on possible causal loci determining plant semidwarfism (Barboza et al., 2013). Signatures of selection were previously identified in *GA5* among *A. thaliana* accessions collected global (Barboza et al., 2013), but the environmental cue driving such selection remained elusive. Later, Luo et al., (2015) identified several *ga5* loss-of-function plants in two Alpine populations of *A. thaliana* and and noted a decrease in plant height with increasing altitude although the adaptive advantage to high altitude environments specifically was not conclusively demonstrated. Our findings demonstrate that high-wind treatment significantly impacts the growth vigour of tall Dutch *A. thaliana* accessions, whereas it has no significant effect on the vigour of predominantly coastally located semidwarf plants.

Arguably, these are plants growing under optimal water and nutrient supply. Under natural circumstances, with additional growth-limiting factors and probably periods with much stronger winds, this reduction in vigour in the tall accessions is likely to be more detrimental, and also likely to affect their fitness (Younginger et al., 2017), which we could not establish under our artificial windy conditions. The observed phenotypic responses in terms of biomass and stem length to windy conditions are similar between the natural Dutch accessions and the Col-0 and Ler reference strains, as well as their corresponding *ga5* loss-of-function lines. Considering that wind speed generally increases with altitude in mountainous areas (https://map.neweuropeanwindatlas.eu), the selective advantage of the Alpine semidwarfs is also likely to be due to tolerance to wind speed. The selective pressure seems high, given the many independent *ga5* alleles found in the Netherlands, and elsewhere in the world (Barboza et al., 2013; Luo et al., 2015). It is remarkable that the near-isogenic genotype collected at several Dutch sites spread over such large distances, without any obvious means of transport. This is in contrast to a previous case of whole-genome hitchhiking, due to selection on herbicide tolerance, where the spread was clearly linked to transportation along railway tracks (Flood, Van Heerwaarden, et al., 2016).

The presence of a North-South cline in flowering time and the association between windy conditions and inflorescence height both highlight the importance of relatively mild environmental clines for local adaptation. These patterns can be detected in a uniformly and densely sampled population like the DartMap. This is further exemplified by the strong association on chromosome 4 related to the mild cline in seasonality, where the alternative allele of *qDPT-B* is only found at a high local abundance in the West of the Netherlands. However, it should be noted that such associations do not necessarily provide evidence of causality. Demographic processes or genetic drift can produce similar patterns to those created by selection (Rellstab et al., 2015). To further investigate whether this association represents a case of local adaptation, common garden experiments or reciprocal transplanting experiments can be conducted (de Villemereuil et al., 2016; Hancock et al., 2011; Rellstab et al., 2015; Terés et al., 2019). Nonetheless, the high local relative abundance observed in the West of the Netherlands compared to rest of the world, along with the differential sensitivity to drought, suggests the presence of an adaptive locus. However, it seems unlikely that this adaptation is solely attributed to the drought response measured in our experiments, as similar climate characteristics

in other regions would also be expected to exhibit a higher relative abundance. Therefore, it is more plausible that a combination of different climatic factors and/or other unconsidered factors such as urbanization contribute to this adaptation.

Identifying the causal gene in genotype-environment association analysis can be challenging because allelic variation may affect tolerance to an experimentally measured trait that may not necessarily reflect the trait under natural selection. However, in our study we found that the mutant of *La-RELATED PROTEIN 1c* (*LARP1c*) is more sensitive to drought compared to other tested mutants and Col-0. *LARP1c* encodes a protein with a conserved La-motif domain, which is associated with RNA recognition and is widely found among eukaryotes (Bayfield et al., 2010). In *A. thaliana, LARP1c* is known to be involved in seed germination and the regulation of leaf senescence, which can be induced by various environmental stresses (Yan et al., 2023; B. Zhang et al., 2012). However, the specific functions of *LARP1c* in plants are still relatively understudied, and there may be other roles yet to be explored (Yan et al., 2023). *LARP1c* may thus also be involved in the response to drought and possibly also other environmental factors.

In addition to the mild and gradual environmental clines, we also examined the panel for sensitivity to low iron availability, as we expected to find variation for this trait given the distinct differences in soil type across the Netherlands. This identified non-coding variation at *FSD3* to cause a major QTL affecting sensitivity to low iron supply. *FSD3* is long known to encode an essential chloroplast-localized iron superoxide dismutase. Its function is to protect developing chloroplasts against excessive damage by superoxide (Myouga et al., 2008). However, natural genetic variation in *FSD3* has not been previously observed. The QTL affecting sensitivity to low iron supply is influenced by subtle allelic variation in *FSD3*, specifically related to the number of repeats in a microsatellite region directly downstream of the transcribed region. This variation results in a two-fold difference in gene expression between contrasting alleles, which strongly correlates with photosynthesis efficiency under iron-deficient conditions. When plants are supplied with sufficient iron, there is no phenotypic difference between contrasting haplotypes. FSD3 is one of two chloroplast-localized iron superoxide dismutases, the other being FSD2 (Myouga et al., 2008). While both can form heterocomplexes and *FSD3* overexpression can partly complement the *fsd2* mutant phenotype (Gallie & Chen, 2019), they are not genetically redundant. The *fsd2* loss-of-function mutant displays a compact rosette phenotype, with pale green leaves, and the *fsd3* loss-of-function mutant is barely viable, with a very small and pale rosette (Myouga et al., 2008). Since iron is an essential co-factor for iron superoxide dismutases, a decrease in iron supply is likely to affect their function. Under the mild iron deficiency conditions we tested, this decrease in function was apparently just enough to affect photosynthesis efficiency significantly more in the genotypes with lower *FSD3* expression.

We found no reason to suggest strong selection for either of the two alleles. Since FSD3 function is essential for proper chloroplast development and tolerance to reactive oxygen species generated by photosynthesis, lower *FSD3* expression is likely disadvantageous. The impact of lower *FSD3* expression may even extent beyond chloroplast development and photosynthesis efficiency. For instance, hydrogen peroxide, which is produced when superoxides are converted by superoxide dismutases, can function as a signaling molecule in the systemic response to iron deficiency (Le et al., 2016). Therefore, plants with lower *FSD3* expression may have a delayed response in detecting iron deficiency. Nevertheless, a large ionome survey of 1135 global *A. thaliana* accessions did not reveal any clear indications of local adaptation to iron deficiency (Campos et al., 2021). This suggests that establishing a direct link between allelic variation and local adaptation to iron deficiency may be challenging. Factors such seasonal

variation in iron availability depending on precipitation or limited distribution and heterogeneity of the allelic variation could contribute to the difficulty in establishing such a link.

In conclusion, our work demonstrates that regions with relatively mild environmental clines can still harbor a wide range of adaptive genetic variants. This highlights the value of establishing regional collections as excellent tools for studying local adaptation of *A. thaliana*. Such a collection does not need to be limited to regions with strong environmental clines as were put forward in previous studies (Brachi et al., 2013; Frachon et al., 2018). Even a diversity panel from a relatively small geographic region with mild environmental clines, such as the Netherlands, provides enough variation to pick up small differences in flowering time and sensitivity to iron deficiency. Given the decreasing costs of whole genome sequencing, it becomes increasingly easier to characterize the genetic variation in such populations in great detail. Our study may serve as a blueprint of this approach, supporting the view that collections of wild plant species can be powerful systems to determine the genomic basis of natural adaptation (Monnahan et al., 2019; Sollars et al., 2017; Weigel & Nordborg, 2015).

Materials and Methods

Plant material

A set of 192 lines was selected from a large collection of around 2,000 natural accessions sampled throughout the Netherlands, with the aim to obtain a uniform geographical spread of accessions and generate a local diversity panel, the DartMap panel, suitable for GWAS (Table S9). While accessions were selected from a larger set of about 2,000 collected individuals, the selected panel has a relatively lower density of accessions in the north of the Netherlands and in the very south, due to limited availability or collector capacity. Seeds collected from individual plants in the field were used to propagate each accession as a line for two generations through single-seed-descent upon self-fertilization in a greenhouse, before DNA isolation.

DNA isolation

Genomic DNA was extracted from one or more inflorescences (open flowers and above) from a single plant per accession. The material was frozen in liquid nitrogen, ground to a fine dust and incubated in $300\mu L$ 2x CTAB buffer (2% CTAB, $1.4m$ NaCl, $100mm$ Tris, $20mm$ EDTA, pH 8) for 30 minutes at 65 °C. An equal volume of chloroform was added, mixed, centrifuged (3250 rpm for 15 min) and the supernatant was collected. The DNA was precipitated by adding an equal volume of ice-cold isopropanol, incubated overnight at -20 °C and centrifuged (3250 rpm for 15 min). The precipitate was washed twice with 70% ethanol and air dried. DNA was dissolved in milliQ water and treated with RNase (Promega) for 30 minutes at 37 °C.

Genomic DNA sequence data

Genomic DNA of the DartMap panel was sequenced at GenomeScan B.V., Leiden. Libraries were prepared using the NEBNext Ultra DNA Library Prep kit for Illumina, according to described procedures. 500-700 bp insert size fragments were sequenced using a Illumina Hiseq 4000 device, aiming for at least 30x coverage of paired-end reads (2x151 bp) of each accession, and used to generate the alignment files necessary for calling genomic variants (see Supplementary Methods for details). We additionally generated alignments for the DartMap panel combined with 11 Dutch accessions

included in the 1,001 Genomes (1001G) Project (Alonso-Blanco et al., 2016) (Sequence Read Archive ID: SRP056687). This combined set of samples will be referred to as the "DartMap + 1001G panel", to clearly distinguish it from the DartMap panel containing accessions collected in this study only.

Single nucleotide polymorphism (SNP) and indel calling

Genetic variants were called using a workflow based on the Genome Analysis Toolkit (GATK) Best Practices (Van der Auwera et al., 2013). Base quality scores of aligned reads were recalibrated using GATK (version 4.0.2.1) BaseRecalibrator with default parameters, using a set of variants called in a world-wide panel of 1135 *A. thaliana* accessions as known sites (Alonso-Blanco et al., 2016) (obtained from ftp.ensemblgenomes.org/pub/plants/release-37/vcf/arabidopsis_ thaliana). SNPs and short insertions/deletions (indels) were called in each sample using GATK HaplotypeCaller, allowing a maximum of three alternate alleles at each site. Samples were jointly genotyped using GATK GenomicsDBImport and GATK GenotypeGVCFs with default parameters. This last step generated three different VCF files containing calls for the nuclear genome, the mitochondrial genome, and the chloroplast genome respectively. We filtered these sets to remove putative false positive calls (see Supplementary Methods for details).

Copy number variation (CNV) calling

We called CNVs in the DartMap panel, defined here as deletions, duplications, and insertions of at least 50 bp, by applying the Hecaton workflow (Wijfjes et al., 2019) to the alignment files of each accession. Accessions of the 1001G were excluded from this analysis, as they were sequenced using a different sequencing platform, which can strongly affect CNV calling: a linear model of genetic distance between accessions based on CNVs as a function of differences in coverage, sequencing technology, insert size, and read length (see Supplementary Methods for details) explained 30% of the variation in the number of CNVs found in each sample ($R^2 = 0.3$). CNVs called in each accession were genotyped and merged to generate a single VCF file as output (see Supplementary Methods for details).

Characterizing the impact of deletions

We detected deletions overlapping protein-coding genes (by at least 1 bp) using bedtools intersect. Genes were obtained from the TAIR10 genomic annotation of Ensembl Plants (release 40) and from the annotation of the improved *A. thaliana* mitochondrial assembly (BK010421.1). Gene function was considered disrupted if the associated gene overlapped with a deletion and at least 76 bp (more than half the length of a read) of the coding sequence lacked read coverage. Deletions were defined as having a moderate frequency if they were present in at least 48 samples (25%) of the DartMap panel. We defined disease resistance genes by first collecting all Uniprot entries linked to all *A. thaliana* reference genes using the Retrieve/ID mapping tool (https://www.uniprot.org/uploadlists/), excluding a small number of genes (53) for which no corresponding entries could be found. Genes were labeled as involved in disease resistance if the name of their protein product contained the terms "disease resistance" or "Disease resistance". Enrichment tests were performed using Fisher's exact test as implemented in the Python SciPy library (Virtanen et al., 2020) (version 1.3.1).

Population structure analysis

We clustered samples based on bi-allelic sites missing calls in fewer than 10% of the samples and having a non-reference allele in at least 10 of them, to ensure that accessions are grouped on variation at the population level, rather than rare variants found in few individuals. For clusters based on mitochondrial and chloroplast variants, this criterium was relaxed to a minimum of two samples, as a relatively large number of sites would have been discarded otherwise. To generate a distance metric, we computed the Hamming distance between samples represented by binary vectors with 1s for homozygous variants and 0 for all other cases, using the dist.gene function of the R package ape (Paradis & Schliep, 2019) (v5.3). Samples were clustered with the hclust function in R (R Core Team, 2013), using complete linkage as the agglomeration method, similar to analyses on the Swedish and Iberian collection (Long et al., 2013; Tabas-Madrid et al., 2018) (Figure S9). We cut trees into different numbers of groups using the R function cutree, to explore the geographical distribution of clusters formed at various hierarchical levels. Geographical distance between accessions was computed using the distance function of the geopy library (https://github.com/geopy/geopy) (v2.0.0) with default parameters and their geographical locations were plotted using the basemap library (https://matplotlib.org/basemap/) (v1.2.2).

Principal component analysis

Principal component analysis of chloroplast variants called in the Dutch and global collection was performed using the SeqArray (Zheng et al., 2017) (v1.28.0) and SNPrelate (Zheng et al., 2012) (v1.22.0) libraries in R. To remove putative false positive calls, variants were excluded if they had a quality-by-depth (QD) score lower than 25, leaving a total of 4095 sites. Furthermore, we removed 10 accessions in which at least 50% of the variants were called as heterozygous, suggesting DNA contamination of the samples used for sequencing. Results were visualized using ggplot2 (Wickham, 2016) (v3.3.2). Three outlier accessions were excluded from the final figures to improve the visualization of Dutch accessions with respect to the global collection. Accessions were coloured based on their chloroplast haplogroup assigned in a previous study (Hsu et al., 2019). Haplogroups of accessions that were not included in that paper (i.e. the DartMap panel) were predicted with a 1-nearest neighbour classifier on the accessions that were, using the first two principal components as features.

Phenotyping flowering time and stem length

For all experiments, seeds were sown on wet filter paper and stratified for 3 days at 4 °C before they were germinated. Seedlings were grown hydroponically on rockwool blocks (Grodan Rockwool Group, 40 x 40 mm in size) pre-soaked in a nutrient solution designed for *A. thaliana* (1.7mm NH_4^+, 4.5mm K^+, 0.4mm Na^+, 2.3mm Ca^{2+}, 1.5mm Mg^{2+}, 4.4mm NO_3^-, 0.2mm Cl^-, 3.5mm SO_4^{2-}, 0.6mm HCO_3^-, 1.12mm PO_4^{3-}, 0.23mm SiO_3^{2-}, 21µm Fe^{2+} (chelated with 3% diethylene triaminopentaacetic acid), 3.4µm Mn^{2+}, 4.7µm Zn^{2+}, 14µm BO_3^{3-} and 6.9µm Cu^{2+} at pH 7, EC 1.4 mS cm^{-1}). Plants were grown in a greenhouse until 97 days after sowing. For each treatment, plants were equally divided over 20 plastic trays, with 76-78 plants per tray. Eight replicates per genotype per treatment were used, randomly assigned over four separate trays with two replicates per tray per genotype. Each tray consisted of a unique, randomly generated combination of genotypes. Trays were placed in a random order in the greenhouse, and were moved once every two weeks to a new position to minimize possible spatial effects in the greenhouse. A vernalisation

treatment of six weeks was performed in a climate-controlled growth chamber set at 4 °C, with 12h/12h light/dark period, 70µm m^{-2} s^{-1} irradiance and 70% humidity starting at day 20 after sowing. After vernalisation, plants were returned to the greenhouse. Flowering time and height up to the first silique were scored in the greenhouse, either with or without a vernalisation period of six weeks. Flowering time was determined as the day when the first flower had opened. Stem length up to the first silique was measured one week after the first flower had opened, as this trait remains stable thereafter (Barboza et al., 2013). To obtain the set of phenotypes for further analysis, we calculated the average phenotypic values per accession per treatment.

Wind simulation treatment

In a separate experiment, plants of six semidwarf accessions and five tall accessions of the DartMap panel, with similar flowering time, as well as the accessions Col-0 and Ler, and their respective loss-of-function *ga5* mutants (Table S3) were grown for the first 14 days after sowing in a climate-controlled growth chamber, set at 20/18 °C (day/night) with 10 h light (from 08:00 - 18:00) and 70% humidity at an irradiance of 200µm m^{-2} s^{-1} at ambient CO_2 levels. The growth chamber contained two treatment basins. The treatment basins have the same design as described in a previous study (Flood, Kruijer, et al., 2016), with the distinction that the two treatment basins are physically separated and are on either side of the climate chamber. Each treatment basin was divided into five equally sized blocks to hold plants. Accessions were grown in a randomized complete block design with six replicates per block. Wind was simulated with two fans that were positioned at either end of one of the treatment basins. The fans were set to provide a wind speed that varied around 5.6 m/s for the majority of the treatment basin. The wind speed varied depending on the distance of the plant to the fan, with plants closest to the fan experiencing wind speeds up to 7 m/s and plants furthest away of the fan and at the edge of the treatment basin experiencing 3.5 m/s wind. The wind speed for the control treatment basin varied between 0.5 and 1.0 m/s. Wind speeds were measured with a TSI Velocalc (model: 8347-M-GB) air velocity monitor. Wind treatment was simulated by alternating patterns of six hours with wind by six hours without wind and was terminated when the first plants in the wind treatment basin had fully ripened siliques.

Extraction of climatic variables

Climatic variables were obtained from google earth engine using R and RGEE (Aybar et al., 2020). All bands from Worldclim BIO data (Hijmans et al., 2005) selected from 250m-by-250m squares using the coordinates from the collection site for each accession as input.

Drought treatment

To test which of the two nearby QTLs at *qDPT* affected drought tolerance, plants were grown in a greenhouse on ceramic-based granular soil. For the first two weeks, all plants were watered three times a week with the nutrient solution described previously. This watering regime was continued for the control treatment while for the drought treatment excess nutrient solution was removed and watering was ceased. Accessions (Table S5) were grown in a randomised complete block design with five replicates per treatment. Photographs were taken with a Nikon D3000 with the same settings, zoom and distance to the plants and were subsequently analysed to obtain the projected leaf area (PLA) with ImageJ (settings Hue 26-114, Saturation 58-218 and Brightness 121-255). The PLA was then normalised per accession by dividing the PLA in the drought treatment by the PLA in

the control treatment to correct for innate differences in size of different accessions. For candidate gene validation T-DNA insertion lines were used (Table S6). Plants were grown in our phenotyping growth chamber, the Phenovator (Flood, Kruijer, et al., 2016), with the same conditions in the growth chamber as described for the wind simulation treatment. The drought treatment was induced in a similar way, by using ceramic-based granular soil. In contrast to the experiment in the greenhouse, the growth substrate was covered with black plastic cover to prevent algal growth that might interfere with high throughput phenotyping. Moreover, we watered plants in the drought treatment only once at the start, as there was less evaporation in this setup compared to the greenhouse. Starting at 10 days after sowing, we measured projected leaf area based on near infrared reflectance (NIR) imaging as described by Flood et al., (2016) four times per day, and F_v/F_m using chlorophyll fluorescence once per day.

Iron deficiency treatment

In a separate experiment, plants were grown as described for the wind simulation experiment, but in this case, seven rather than five blocks per treatment basin were used. Plants in the control treatment were watered with the nutrient solution described above, while those under iron deficiency treatment received an adjusted solution with only 1µm Fe^{2+} instead of 21µm Fe^{2+}. Either three or four replicates were used per accession per treatment, with half of the accessions having four replicates in control condition and three replicates in the iron deficiency treatment and vice versa for the other half of the accessions. On day 15, plants were transferred to our phenotyping growth chamber, the Phenovator (Flood et al., 2016), with the same conditions. We measured the steady state quantum yield of photosystem II electron transport (Φ_{PSII}), as a proxy for photosynthesis efficiency, and the projected leaf area, based on near infrared reflectance (NIR) imaging as described by Flood et al., (2016), at five resp. seven time points per day. We discarded outlier samples with NIR values smaller than the mean minus two standard deviations per treatment. We corrected the raw phenotypic data (Φ_{PSII} and projected leaf area) for block effects by fitting the following linear mixed model in which random terms are underlined:

$$\underline{Y} = Genotype + Treatment + (Genotype \times Treatment) + \underline{Block} + \underline{\epsilon}$$

The model was analysed using the restricted maximum likelihood procedure with the lme4 package in R (Bates et al., 2014). For the GWAS, we calculated the average Φ_{PSII} per accession per treatment. To obtain a single phenotypic value representing a measure of tolerance per accession, we then plotted Φ_{PSII} for each accession under the iron deficient condition versus control condition and fitted a linear regression through the entire population. We calculated the residuals of each accession relative to this linear regression to be used as the phenotypic value for GWA analysis. Additionally, we used the average Φ_{PSII} values in the iron deficient condition and the calculated ratio between Φ_{PSII} in iron deficiency versus control as alternative traits for GWAS.

Plant material for expression analysis

Plants that were used for gene expression analysis were grown in the greenhouse on rockwool blocks that were watered with either the control nutrient solution (21µm Fe^{2+}) or the adjusted low iron solution (1µm Fe^{2+}). For both treatments we used five plastic trays to contain the nutrient solution. Two replicate plants per accession were distributed over trays in a complete randomised block design. Whole shoots of two replicate plants per accession originating from the same tray were pooled and harvested as one sample

for RNA isolation 21 days after germination, while five such samples per accession per treatment were harvested in total. The material was snapfrozen in liquid nitrogen immediately after harvesting and stored at -80 °C until RNA isolation. Accessions were selected for gene expression analysis (Table S7) based on their haplotype for the region with the highest association in the GWAS.

Quantitative complementation

Quantitative complementation was used to test if the difference in Φ_{PSII} between the F_1's of Columbia-0 (Col-0) and the knockout mutant *fsd3* in Col-0 background crossed with accessions homozygous for the tolerant allele was significantly different from the response in Φ_{PSII} for F_1's of Col-0 and *fsd3* crossed with accessions homozygous for the sensitive allele. Five accessions with a tolerant, and nine accessions with a sensitive allele for *FSD3* were crossed to both Col-0 and *fsd3* (Table S7). Col-0 and *fsd3* were used as pollen donor in all crosses. The *fsd3* T-DNA insertion knockout mutant, SALK_103228, was ordered from the Arabidopsis Biological Resource Center and insertion of the T-DNA was validated by PCR as earlier described (Myouga et al., 2008). This T-DNA insertion line has no *FSD3* transcript (Myouga et al., 2008), and we could only grow it beyond the seedling stage on a 0.3% gel-rite medium supplemented with 3% sucrose.
Plants (N = 10 per cross per treatment) were grown as described in the section "Plant material for expression analysis". Plants were phenotyped for Φ_{PSII} from 21 DAS onwards as described by Baker and Oxborough (Baker & Oxborough, 2004) with the PlantScreenTM SC System (Photosystems Instruments). Phenotyping was done for twenty plants per run and lasted for three consecutive days overall. We phenotyped crosses with the tolerant allele and the sensitive allele equally spread out over the three days to prevent possible phenotypic differences caused by the delay between measurements. Moreover, in each measurement of twenty plants we included F_1's from both the cross between a given accession with Col-0 and with *fsd3* in a single measurement. We included one reference tray that was measured each day in the morning, at the start of the afternoon and at the end of the afternoon to test for possible differences in Φ_{PSII} during the day, but no significant differences were found. We corrected the phenotypic data for possible block effects by fitting a linear model, with random terms underlined:

$$\underline{Y} = (FSD3\ allele\ group \times Pollen\ donor) + \underline{Block} + \underline{\varepsilon}$$

The model was analysed using the restricted maximum likelihood procedure with the lme4 package in R (Bates et al., 2014). We tested interaction between the *FSD3* allele group (natural tolerant or sensitive allele) and the pollen donor (Col-0 or the *fsd3* mutant) with a two-way ANOVA, for which a Kenward-Roger approximation for the degrees of freedom was used. Tukey's post-hoc multiple comparison test ($\alpha = 0.05$) was used to assess significance.

RNA isolation and gene expression analysis

Total RNA was extracted with the Direct-zol RNA mini-prep kit (Zymo Research) according to the company's instructions and then subjected to a DNAse (Promega) treatment at 37 °C for one hour. Total removal of DNA was validated by means of a no-reverse-transcriptase PCR reaction on a few mixtures of samples. The RNA quality was assessed for purity (A260/A280) with a spectrophotometer (Thermo scientific Nanodrop 2000) and for possible RNA degradation by means of a visual inspection of the RNA on a 1% agarose gel. cDNA was then synthesized from 1µg total RNA (measured by spectrophotometer)

with the iScript™ cDNA synthesis kit (Bio-RAD) according to the company's instructions. Reverse transcriptase quantitative PCR (RT-qPCR) reactions were performed with the SensiFast SYBR Green Mastermix (Bioline) on a Bio-RAD cfx96 machine. For each RT-qPCR primer set used, we first validated its efficiency by means of a standard curve and only used primer sets with efficiencies ranging between 90% and 110%. Gene expression was normalized, using the delta cycling threshold (ΔCt) method (Livak & Schmittgen, 2001), to three reference genes, *SAND*¸ *YLS8*, and *TIP41-like* as earlier described (Han et al., 2013). Differential expression was assessed by two-sample t-tests on the ΔCt values. All primer sequences can be found in Table S10. We included *FSD1* and *FRO3* in the expression analysis as marker genes for iron deficiency in shoots (Waters et al., 2012).

Genome-wide association (GWA) and genome-environment association (GEA) analyses

GWA and GEA mapping were performed in GEMMA (Zhou & Stephens, 2012) (version 0.98) with a univariate linear mixed model and the minor allele frequency cut-off set at 0.05. A kinship matrix was constructed in GEMMA to correct for population structure. CNVs were included in the analysis in a similar fashion as SNPs and indels by converting them into a reference variant, an alternative variant or heterozygous. To estimate a suitable genome-wide p-value threshold, we performed 1,000 permutations of the phenotype values. For each permutation we performed association analysis and extracted the highest association score. The distribution of these 1,000 highest association scores was then used to determine the empirical threshold value of $p < 0.05$.

Data availability

Raw sequencing data of the DartMap panel can be found on the repository of the National Center for Biotechnology Information (NCBI) (BioProject ID: PRJNA727738).

Acknowledgements

We thank Yanda Zhou, Jeroen van Buren, and Imme Bartels for their contributions to the data collection of phenotyping experiments; the Unifarm staff for taking excellent care of all plants grown during this work; and Korbinian Schneeberger for providing feedback on the manuscript. We are especially grateful to Erik Wijnker and Joost Keurentjes, who called upon the listeners of the Dutch radio program "Vroege Vogels" (https://www.bnnvara.nl/vroegevogels) to collect *A. thaliana* plants and seeds and send them to Wageningen University, and all those people who did. Without them, we would not have been able to get such a well-distributed set of Dutch *A. thaliana* accessions.

Authors' contributions

RYW performed computational analysis of sequencing data and overall patterns of genetic diversity. Wet-lab experiments were conducted by RB, FFMB, MM, JJV, NG and LMD. Subsequent data analysis of wet-lab experiments was performed by RB, MM, JJV, NG and LMD. TPJMT substantially contributed to data collection and analysis of sequencing data of non-Dutch *A. thaliana* accessions. BLS extracted environmental data. RYW, RB, BLS, JJV, MM, NG, LMD, MK, FAvE, SS, DDR, and MGMA were involved in experimental design,

interpretation of results, and preparing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the project grant ALWGR.2015.9 of the Netherlands Organisation of Scientific Research (NWO). We are grateful to the NWO and private partners Rijk Zwaan Breeding, Bejo Zaden, Genetwister Technologies, Averis Seeds, C. Meijer and HZPC Holland for their financial support through this grant. The funding bodies had no role in the design of the study; the collection, analysis, and interpretation of data; and in writing the manuscript.

References

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., Cao, J., Chae, E., Dezwaan, T. M., & Ding, W. (2016). 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell*, *166*(2), 481–491.

Andrés, F., & Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics*, *13*(9), 627–639.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., & Hu, T. T. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, *465*(7298), 627–631.

Aybar, C., Wu, Q., Bautista, L., Yali, R., & Barja, A. (2020). rgee: An R package for interacting with Google Earth Engine. *Journal of Open Source Software*, *5*(51), 2272.

Baker, N. R., & Oxborough, K. (2004). Chlorophyll fluorescence as a probe of photosynthetic productivity. In *Chlorophyll a fluorescence: a signature of photosynthesis* (pp. 65–82). Springer.

Barboza, L., Effgen, S., Alonso-Blanco, C., Kooke, R., Keurentjes, J. J. B., Koornneef, M., & Alcázar, R. (2013). Arabidopsis semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. *Proceedings of the National Academy of Sciences*, *110*(39), 15818–15823.

Bartoli, C., Frachon, L., Barret, M., Rigal, M., Huard-Chauveau, C., Mayjonade, B., Zanchetta, C., Bouchez, O., Roby, D., & Carrère, S. (2018). In situ relationships between microbiota and potential pathobiota in Arabidopsis thaliana. *The ISME Journal*, *12*(8), 2024–2038.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv Preprint ArXiv:1406.5823*.

Bay, R. A., Rose, N., Barrett, R., Bernatchez, L., Ghalambor, C. K., Lasky, J. R., Brem, R. B., Palumbi, S. R., & Ralph, P. (2017). Predicting responses to contemporary environmental change using evolutionary response architectures. *The American Naturalist*, *189*(5), 463–473.

Bayfield, M. A., Yang, R., & Maraia, R. J. (2010). Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1799*(5–6), 365–378.

Billings, W. D., & Mooney, H. A. (1968). The ecology of arctic and alpine plants. *Biological Reviews*, *43*(4), 481–529.

Brachi, B., Villoutreix, R., Faure, N., Hautekèete, N., Piquot, Y., Pauwels, M., Roby, D., Cuguen, J., Bergelson, J., & Roux, F. (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Molecular Ecology*, *22*(16), 4222–4240.

Cameron, D. L., Di Stefano, L., & Papenfuss, A. T. (2019). Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nature Communications*, *10*(1), 3240.

Campos, A. C. A. L., van Dijk, W. F. A., Ramakrishna, P., Giles, T., Korte, P., Douglas, A., Smith, P., & Salt, D. E. (2021). 1,135 ionomes reveal the global pattern of leaf and seed mineral nutrient and trace element diversity in *Arabidopsis thaliana*. *The Plant Journal*, *106*(2), 536–554.

de Villemereuil, P., Gaggiotti, O. E., Mouterde, M., & Till-Bottraud, I. (2016). Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity*, *116*(3), 249–254.

De Vries, F., De Groot, W. J. M., Hoogland, T., & Denneboom, J. (2003). *De Bodemkaart van Nederland digitaal; toelichting bij inhoud, actualiteit en methodiek en korte beschrijving van additionele informatie*. Alterra.

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., & Alonso-Blanco, C. (2017). African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, *114*(20), 5213–5218.

Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*, *573*(7772), 126–129.

Ferrero-Serrano, Á., & Assmann, S. M. (2019). Phenotypic and genome-wide association with the local environment of Arabidopsis. *Nature Ecology & Evolution*, *3*(2), 274–285.

Flood, P. J., Kruijer, W., Schnabel, S. K., van der Schoor, R., Jalink, H., Snel, J. F. H., Harbinson, J., & Aarts, M. G. M. (2016). Phenomics for photosynthesis, growth and reflectance in Arabidopsis thaliana reveals circadian and long-term fluctuations in heritability. *Plant Methods*, *12*(1), 1–14.

Flood, P. J., Van Heerwaarden, J., Becker, F., De Snoo, C. B., Harbinson, J., & Aarts, M. G. M. (2016). Whole-genome hitchhiking on an organelle mutation. *Current Biology*, *26*(10), 1306–1311.

Fournier-Level, A., Wilczek, A. M., Cooper, M. D., Roe, J. L., Anderson, J., Eaton, D., Moyers, B. T., Petipas, R. H., Schaeffer, R. N., & Pieper, B. (2013). Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. *Molecular Ecology*, *22*(13), 3552–3566.

Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., Roby, D., & Roux, F. (2018). A genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. *Frontiers in Plant Science*, *9*, 967.

Fulgione, A., Koornneef, M., Roux, F., Hermisson, J., & Hancock, A. M. (2018). Madeiran *Arabidopsis thaliana* reveals ancient long-range colonization and clarifies demography in Eurasia. *Molecular Biology and Evolution*, *35*(3), 564–574.

Gallie, D. R., & Chen, Z. (2019). Chloroplast-localized iron superoxide dismutases FSD2 and FSD3 are functionally distinct in Arabidopsis. *PloS One*, *14*(7), e0220078.

Göktay, M., Fulgione, A., & Hancock, A. M. (2021). A new catalog of structural variants in 1,301 *A. thaliana* lines from Africa, Eurasia, and North America reveals a signature of

balancing selection at defense response genes. *Molecular Biology and Evolution*, *38*(4), 1498–1511.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1–19.

Grace, J. (1988). 3. Plant response to wind. *Agriculture, Ecosystems & Environment*, *22*, 71–88.

Han, B., Yang, Z., Samma, M. K., Wang, R., & Shen, W. (2013). Systematic validation of candidate reference genes for qRT-PCR normalization under iron deficiency in Arabidopsis. *Biometals*, *26*, 403–413.

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F., & Bergelson, J. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, *334*(6052), 83–86.

Hänsch, R., & Mendel, R. R. (2009). Physiological functions of mineral micronutrients (cu, Zn, Mn, Fe, Ni, Mo, B, cl). *Current Opinion in Plant Biology*, *12*(3), 259–266.

Hartemink, A. E., & Sonneveld, M. P. W. (2013). Soil maps of the Netherlands. *Geoderma*, *204*, 1–9.

Hell, R., & Stephan, U. W. (2003). Iron uptake, trafficking and homeostasis in plants. *Planta*, *216*, 541–551.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *25*(15), 1965–1978.

Hsu, C., Lo, C., & Lee, C. (2019). On the postglacial spread of human commensal *Arabidopsis thaliana*: journey to the East. *New Phytologist*, *222*(3), 1447–1457.

Huang, X., & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, *65*, 531–551.

Jiao, W.-B., & Schneeberger, K. (2020). Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nature Communications*, *11*(1), 989.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., & Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science*, *290*(5490), 344–347.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241.

Koornneef, M., Blankestijn-de Vries, H., Hanhart, C., Soppe, W., & Peeters, T. (1994). The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild-type. *The Plant Journal*, *6*(6), 911–919.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(1), 1–9.

Le, C. T. T., Brumbarova, T., Ivanov, R., Stoof, C., Weber, E., Mohrbacher, J., Fink-Straube, C., & Bauer, P. (2016). Zinc finger of Arabidopsis thaliana12 (ZAT12) interacts with FER-like iron deficiency-induced transcription factor (FIT) linking iron deficiency and oxidative stress responses. *Plant Physiology*, *170*(1), 540–557.

Le Corre, V., Roux, F., & Reboud, X. (2002). DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution*, *19*(8), 1261–1271.

Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-Blanco, C., Weigel, D., & Nordborg, M. (2017). On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nature Communications*, *8*(1), 14458.

Lee, S., Joung, Y. H., Kim, J.-K., Do Choi, Y., & Jang, G. (2019). An isoform of the plastid RNA polymerase-associated protein FSD3 negatively regulates chloroplast development. *BMC Plant Biology*, *19*(1), 1–12.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M., & Borevitz, J. O. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, *107*(49), 21199–21204.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods*, *25*(4), 402–408.

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., Zhang, Q., Vilhjálmsson, B. J., Korte, A., & Nizhynska, V. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, *45*(8), 884–890.

Luo, Y., Dong, X., Yu, T., Shi, X., Li, Z., Yang, W., Widmer, A., & Karrenberg, S. (2015). A single nucleotide deletion in gibberellin20-oxidase1 causes alpine dwarfism in Arabidopsis. *Plant Physiology*, *168*(3), 930–937.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12.

Maxwell, K., & Johnson, G. N. (2000). Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany*, *51*(345), 659–668.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H. Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., … Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), D351–D360.

Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., Laenen, B., Schmickl, R., Paajanen, P., & Šrámková, G. (2019). Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. *Nature Ecology & Evolution*, *3*(3), 457–468.

Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., & Shinozaki, K. (2008). A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in Arabidopsis. *The Plant Cell*, *20*(11), 3148–3162.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.

Pedersen, B. S., & Quinlan, A. R. (2019). Duphold: scalable, depth-based annotation and curation of high-confidence structural variant calls. *Gigascience*, *8*(4), giz040.

Pico, F. X., Méndez-Vigo, B., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2008). Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics*, *180*(2), 1009–1021.

R Core Team, R. (2013). *R: A language and environment for statistical computing*.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370.

Schneeberger, K., Hagmann, J., Ossowski, S., Warthmann, N., Gesing, S., Kohlbacher, O., & Weigel, D. (2009). Simultaneous alignment of short reads against multiple genomes. *Genome Biology*, *10*(9), 1–12.

Sloan, D. B., Wu, Z., & Sharbrough, J. (2018). Correction of persistent errors in Arabidopsis reference mitochondrial genomes. *The Plant Cell*, *30*(3), 525–527.

Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, *64*(2), 295–315.

Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., Kaithakottil, G., Cooper, E. D., Uauy, C., & Havlickova, L. (2017). Genome sequence and genetic diversity of European ash trees. *Nature*, *541*(7636), 212–216.

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., Purugganan, M. D., & Schmitt, J. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences*, *101*(13), 4712–4717.

Tabas-Madrid, D., Méndez-Vigo, B., Arteaga, N., Marcer, A., Pascual-Montano, A., Weigel, D., Xavier Pico, F., & Alonso-Blanco, C. (2018). Genome-wide signatures of flowering adaptation to climate temperature: Regional analyses in a highly diverse native range of *Arabidopsis thaliana*. *Plant, Cell & Environment*, *41*(8), 1806–1820.

Terés, J., Busoms, S., Perez Martín, L., Luís-Villarroya, A., Flis, P., Álvarez-Fernández, A., Tolrà, R., Salt, D. E., & Poschenrieder, C. (2019). Soil carbonate drives local adaptation in *Arabidopsis thaliana*. *Plant, Cell & Environment*, *42*(8), 2384–2398.

Terry, N. (1980). Limiting factors in photosynthesis: I. Use of iron stress to control photochemical capacity in vivo. *Plant Physiology*, *65*(1), 114–120.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., & Thibault, J. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1), 10–11.

Veer, G. van der. (2006). *Geochemical soil survey of the Netherlands. Atlas of major and trace elements in topsoil and parent material; assessment of natural and anthropegenic enrichment factors* (Issue 347). Utrecht University.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Waters, B. M., McInturf, S. A., & Stein, R. J. (2012). Rosette iron deficiency transcript and microRNA profiling reveals links between copper and iron homeostasis in *Arabidopsis thaliana*. *Journal of Experimental Botany*, *63*(16), 5903–5918.

Weigel, D., & Nordborg, M. (2015). Population genomics for understanding adaptation in wild plant species. *Annual Review of Genetics*, *49*, 315–338.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wijfjes, R. Y., Smit, S., & De Ridder, D. (2019). Hecaton: reliably detecting copy number variation in plant genomes using short read sequencing data. *BMC Genomics*, *20*, 1–13.

Yan, Z., Jing, M., Zhang, B., Shi, H., Jin, X., Yan, X., Gao, T., & Han, Y. (2023). The Arabidopsis LARP1s are Involved in Regulation of Seed Germination. *Journal of Plant Growth Regulation*, *42*(3), 1775–1788.

Younginger, B. S., Sirová, D., Cruzan, M. B., & Ballhorn, D. J. (2017). Is biomass a reliable estimate of plant fitness? *Applications in Plant Sciences*, *5*(2), 1600094.

Zhang, B., Jia, J., Yang, M., Yan, C., & Han, Y. (2012). Overexpression of a LAM domain containing RNA-binding protein LARP1c induces precocious leaf senescence in Arabidopsis. *Molecules and Cells*, *34*, 367–374.

Zhang, L., & Jiménez-Gómez, J. M. (2020). Functional analysis of FRIGIDA using naturally occurring variation in Arabidopsis thaliana. *The Plant Journal*, *103*(1), 154–165.

Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C., & Levine, D. (2017). SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics*, *33*(15), 2251–2257.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824.

Zou, Y.-P., Hou, X.-H., Wu, Q., Chen, J.-F., Li, Z.-W., Han, T.-S., Niu, X.-M., Yang, L., Xu, Y.-C., & Zhang, J. (2017). Adaptation of *Arabidopsis thaliana* to the Yangtze River basin. *Genome Biology*, *18*, 1–11.

Table 1: **The number of polymorphic bi-allelic SNPs and indels ("Variants") found in regional *A. thaliana* populations.**

Population	Samples	Variants	Reference
Dutch (DartMap + 1001G)	203	3,057,293	This study, (Alonso-Blanco et al., 2016)
Sweden	243	3,084,964	(Long et al., 2013)
Iberian Peninsula	190	4,482,982	(Tabas-Madrid et al., 2018)

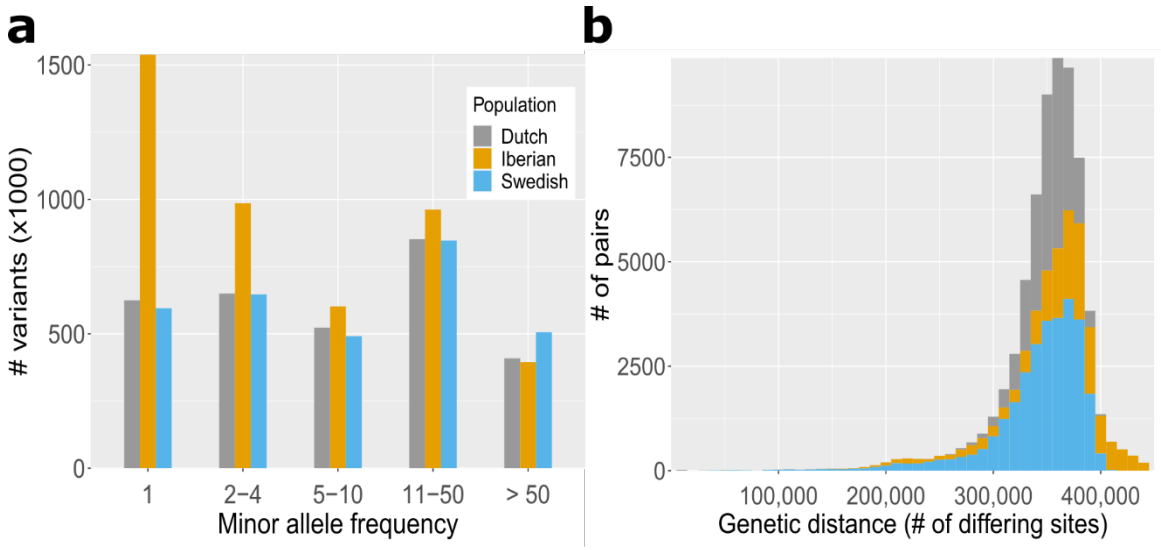

Figure 1: **Comparing the genetic diversity of three regional populations.** (a) Minor allele frequency of bi-allelic SNPs and indels. (b) Genome-wide similarity between pairs of samples (bars are overlapping).

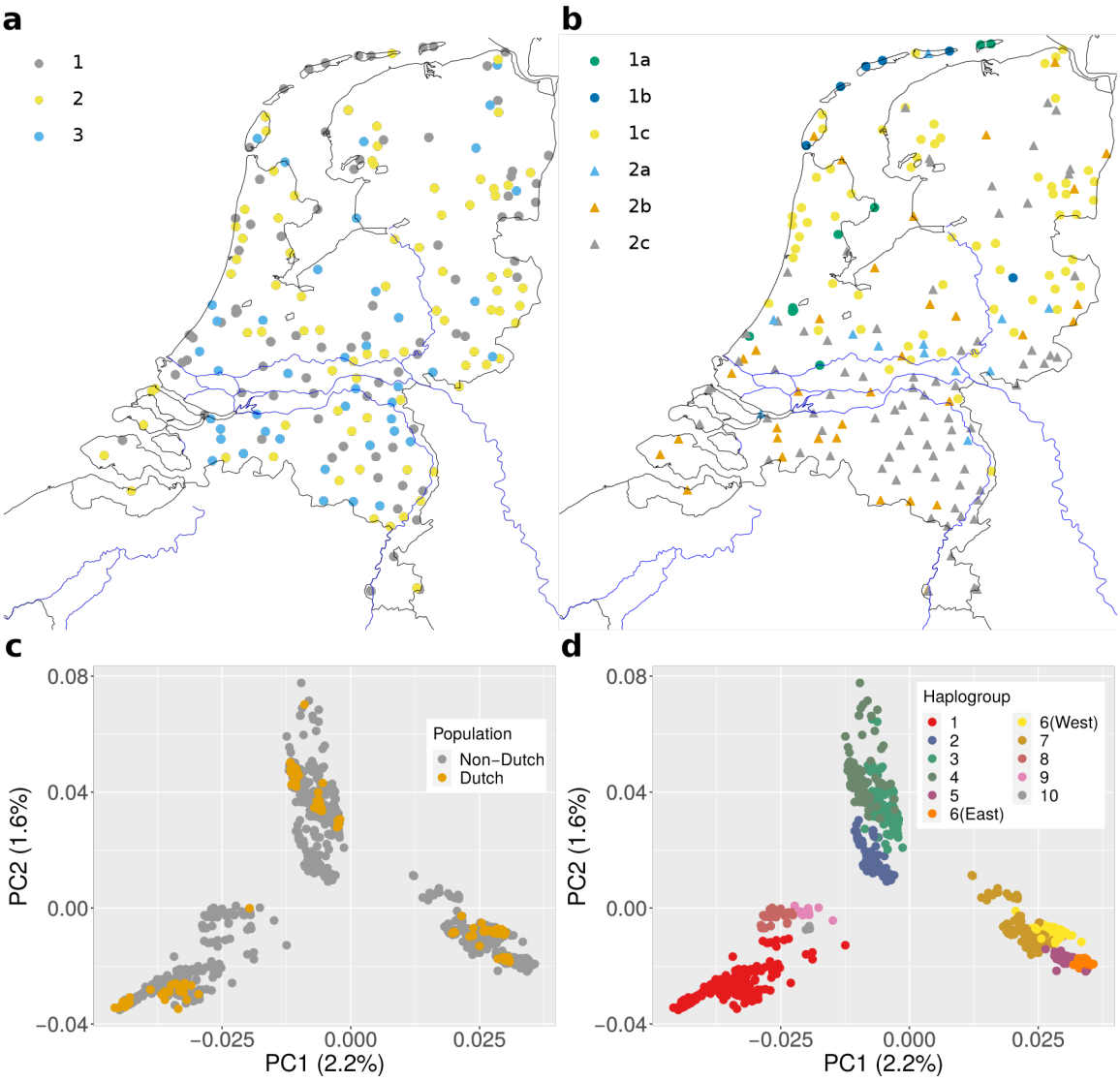

Figure 2: **Population structure of *A. thaliana* in the Netherlands.** (a-b) Geographical location of Dutch accessions clustered based on organellar (a) and nuclear (b) genome-wide similarity. Colours are used to distinguish accessions of different clusters. The main rivers of the Netherlands are depicted by blue lines. (c) Principal component analysis of chloroplast variants in Dutch accessions and the global collection of *A. thaliana* analysed in a previous study (Hsu et al., 2019). (d) Samples in (c) coloured according to the chloroplast haplogroups reported in a previous study (Hsu et al., 2019).

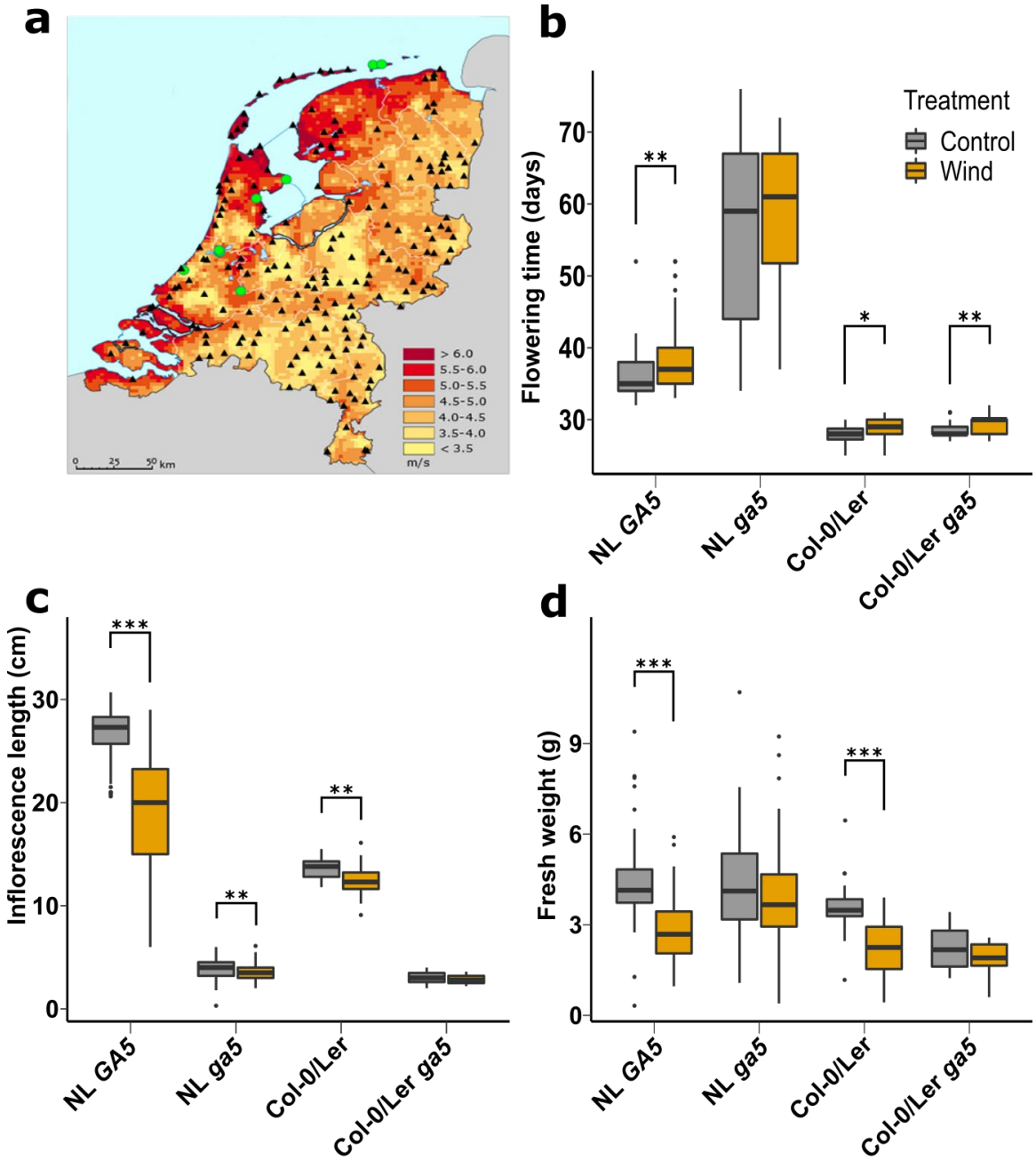

Figure 3: **Phenotypic response to windy conditions.** (a) Average annual wind speed (between 1981-2010) in the Netherlands as measured at ground level. Green circles indicate semidwarf accessions that share the same splice site mutation, black triangles indicate all other DartMap accessions. Figure adapted from https://www.knmi.nl/klimaat-viewer/kaarten/wind/gemiddelde-windsnelheid/jaar/Periode_1981-2010. (b-d) Plant responses to absence or presence of windy conditions (grey and orange respectively) for (b) flowering time, (c) inflorescence length one week after the first flower opened, (d) fresh weight at flowering $*P < 0.05$; $**P < 0.01$; $***P < 0.001$, two-tailed Student's *t*-test.

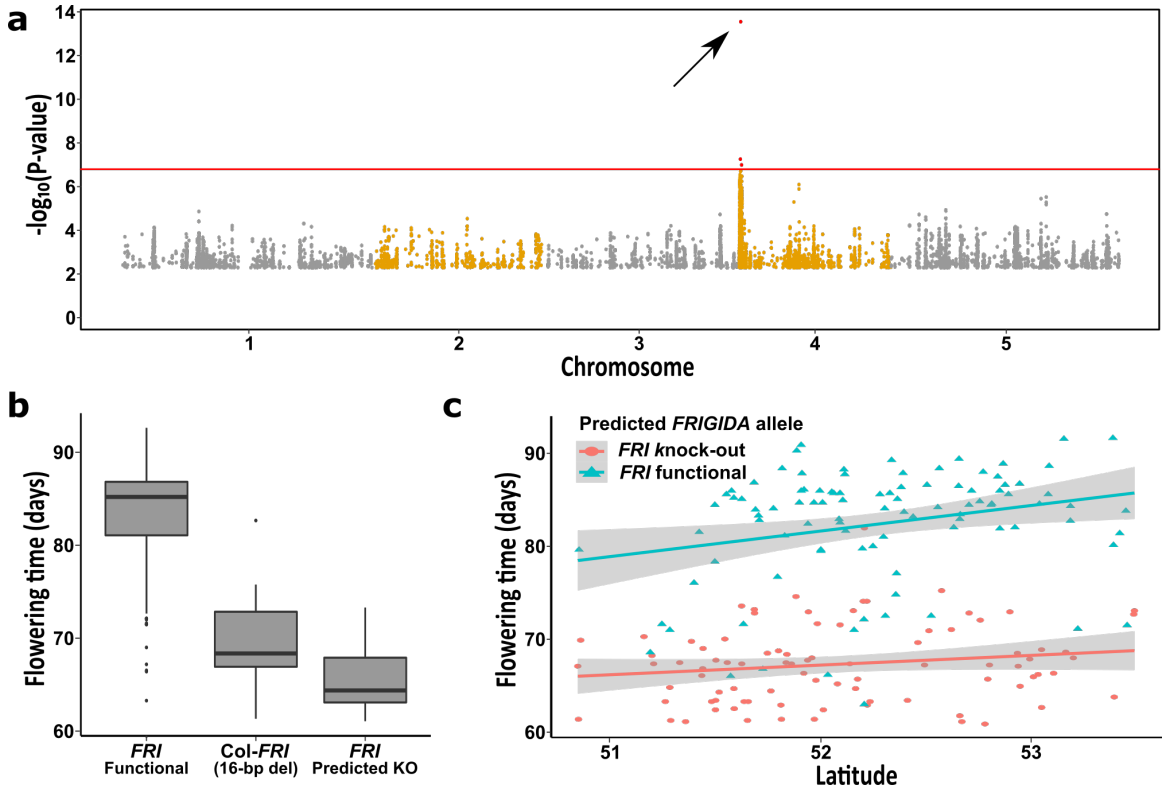

Figure 4: **GWA analysis of flowering time in the DartMap panel upon vernalization.** (a) Manhattan plot for days until flowering after a six-week vernalization treatment, identifying one highly associated polymorphism located in the *FRIGIDA* gene (arrow). The red line corresponds to a permutation-based threshold level of significance at $P < 0.05$. (b) Flowering time distribution of three *FRIGIDA* variants in the DartMap panel. The Col-0 variant corresponds to a 16 bp deletion causing a premature stop codon in the *FRIGIDA* coding sequence. (c) Average days until flowering for DartMap accessions relative to their latitudinal site of origin with linear fit for accessions with a predicted loss-of-function *fri* allele (red, $R^2 = 0.016$, $P = 0.128$) or a functional *FRI* allele (blue, $R^2 = 0.082$, $P = 0.002$).

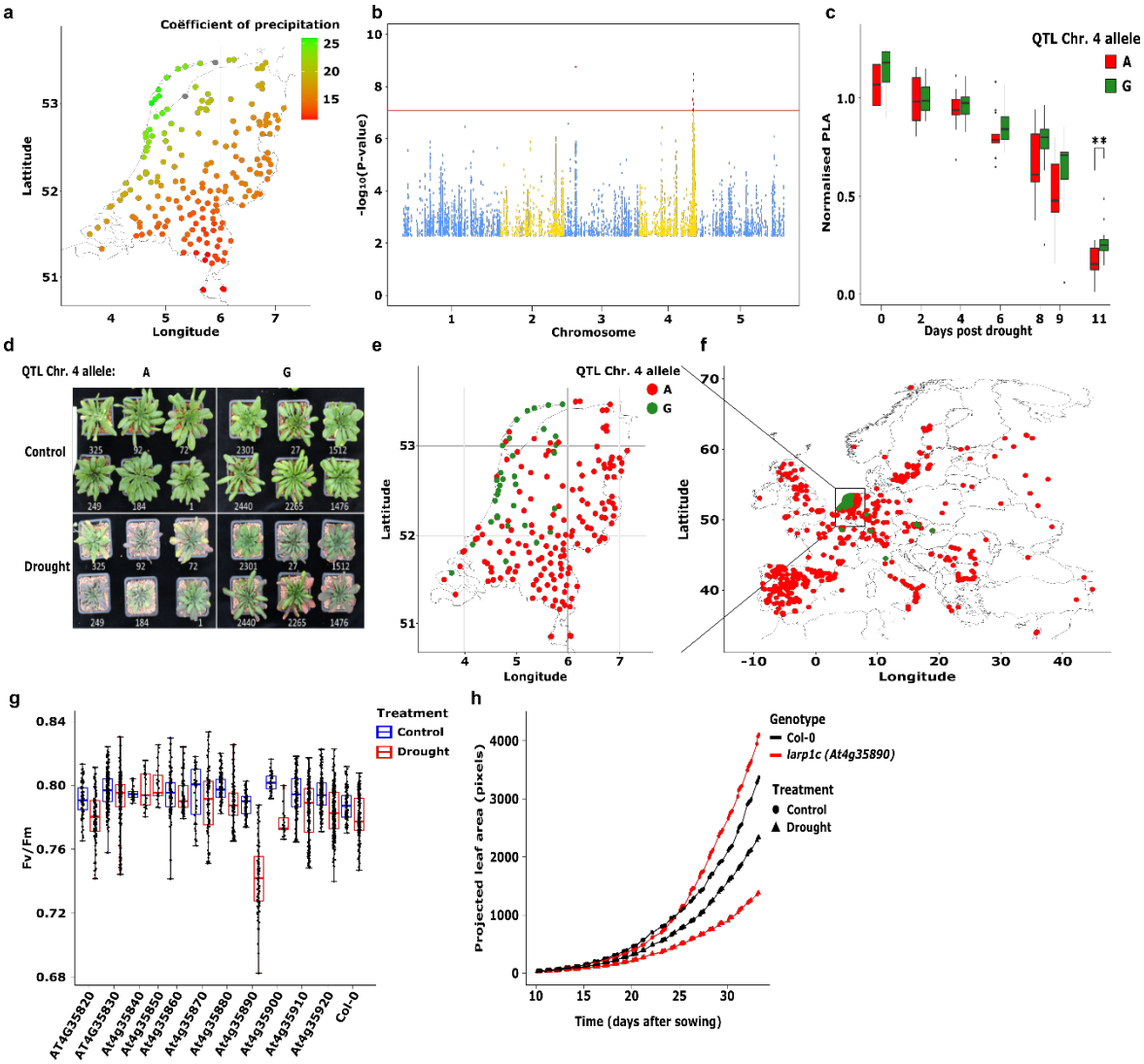

Figure 5: **Genome-environment association (GEA) analysis on climatic variables related to precipitation and temperature.** (a) Geographical distribution of precipitation seasonality (coefficient of variation) expressed in Seasonality Index (SI) units for the DartMap. SI values <0.19 are defined as 'precipitation throughout the year', whereas values between 0.20 and 0.39 are defined as 'precipitation throughout the year, but with a definite wetter season'. (b) Manhattan plot for the GEA related to precipitation seasonality (coefficient of variation). (c) Response to drought treatment for accessions differing at their *qDPT-B* allele. Boxplots represent the normalised projected leaf area (PLA) calculated per accession by dividing the PLA in the drought treatment over the PLA in the control treatment for a set of 14 randomly chosen accessions per allele (depicted as an adenine (A) or guanine (G) nucleotide at SNP $4^{17000635}$). (d) Pictures of six representative accessions per *qDPT-B* allele in the control and drought treatment, taken 11 days after the start of the drought treatment. (e) Geographical distribution of the two alleles ('A' and 'G') across the Netherlands and (f) Europe. (g)) F_v/F_m as a proxy of plant stress of Col-0 and candidate gene T-DNA insertion lines. (f) The response of PLA of Col-0 and the *larp1c* mutant.

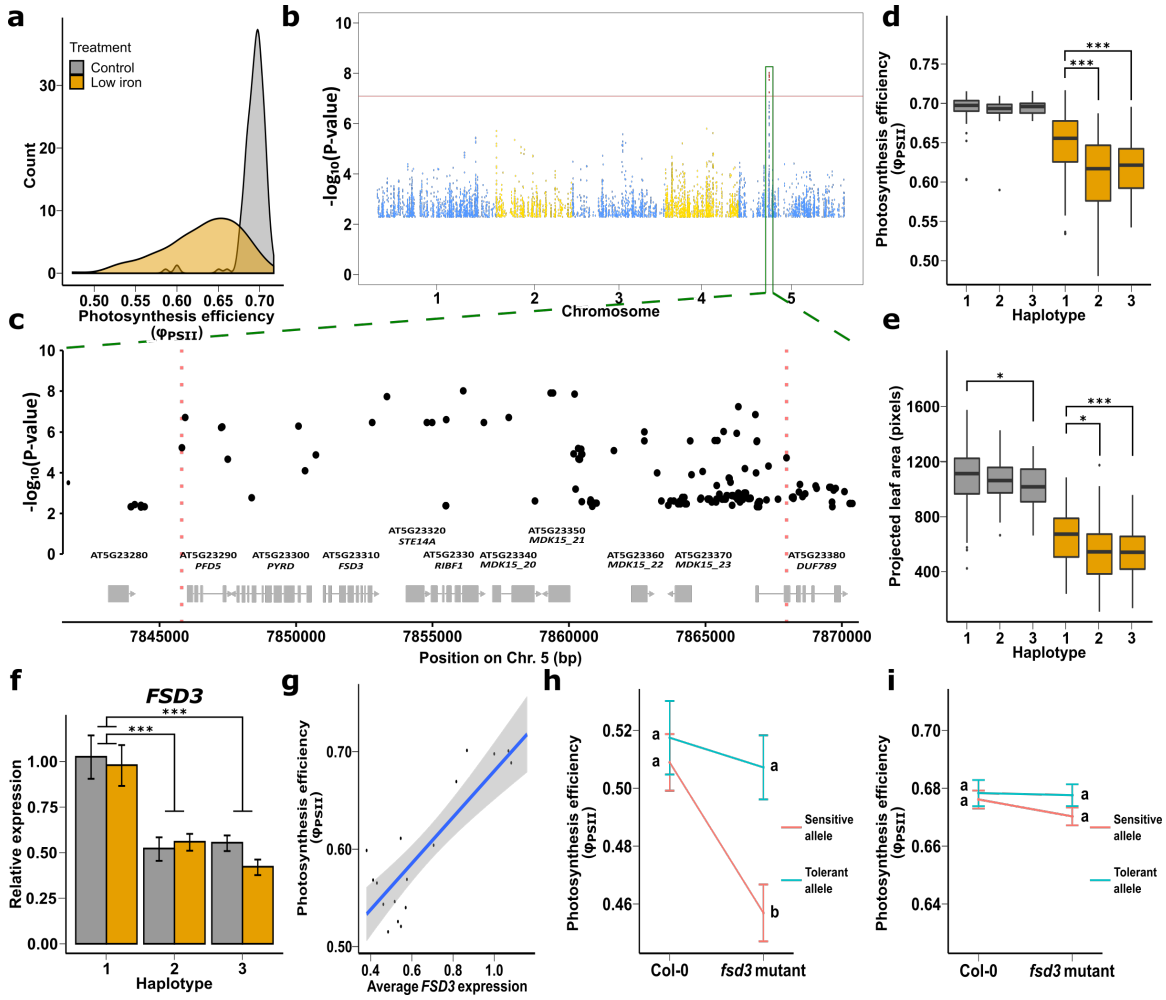

Figure 6: **Genome Wide Association (GWA) analysis of the response of photosynthesis efficiency to iron deficiency.** (a) Distribution of Φ_{PSII}, the light-use efficiency of photosystem II (PSII) electron transport, when grown in control (21µm Fe^{2+}, grey) and low iron supply (1µm Fe^{2+}), orange). (b) Manhattan plot of the SNP variant associations [-log(P) > 2] when using the residual values of the linear regression of Φ_{PSII} in control conditions versus iron deficiency conditions as trait for GWA analysis. The red line indicates a permutation-based threshold level of significance of $P < 0.05$. (c) The genomic region surrounding the significantly associated region on chromosome 5 (green box in (b)). Red dotted lines demarcate the area in linkage disequilibrium with the most significant SNP, containing candidate genes depicted below (in grey). (d-e) Phenotypic distributions for Φ_{PSII} (d) and projected leaf area (e) per haplotype. (f) Relative gene expression levels per haplotype for *FSD3*. (g) Photosynthesis efficiency relative to the average relative *FSD3* expression level per accession with a linear fit (*FSD3*: $R^2 = 0.74$, $P = 2.94 \times 10^{-6}$). (h-i) Φ_{PSII} of F_1 progeny of crosses between Col-0 (tolerant *FSD3* allele) or the *fsd3* knockout mutant with accessions homozygous for the tolerant allele (N = 5) (blue) and similarly with accessions homozygous for the sensitive allele (N = 9) (red) in iron deficiency (h) and control conditions (i). Error bars represent SE. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$, two-tailed Student's *t*-test.

Supplementary Figures

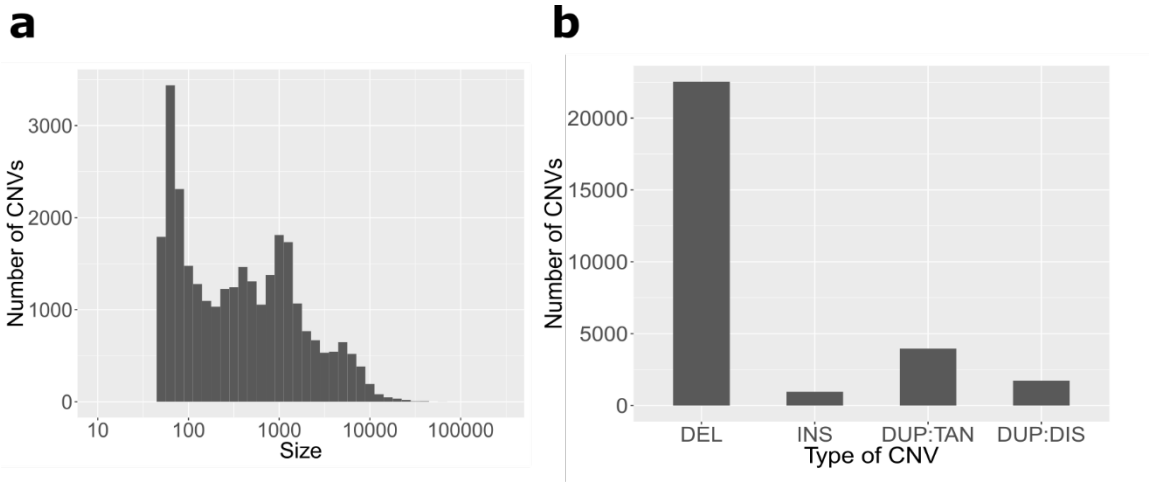

Figure S1: **CNVs in the DartMap panel.** (a) Size distribution of CNVs. CNVs below 50 bp are excluded, as these are considered indels. (b) Type distribution of CNVs: deletions (DEL), insertions (INS), tandem duplications (DUP:TAN), or dispersed duplications (DUP:DIS).

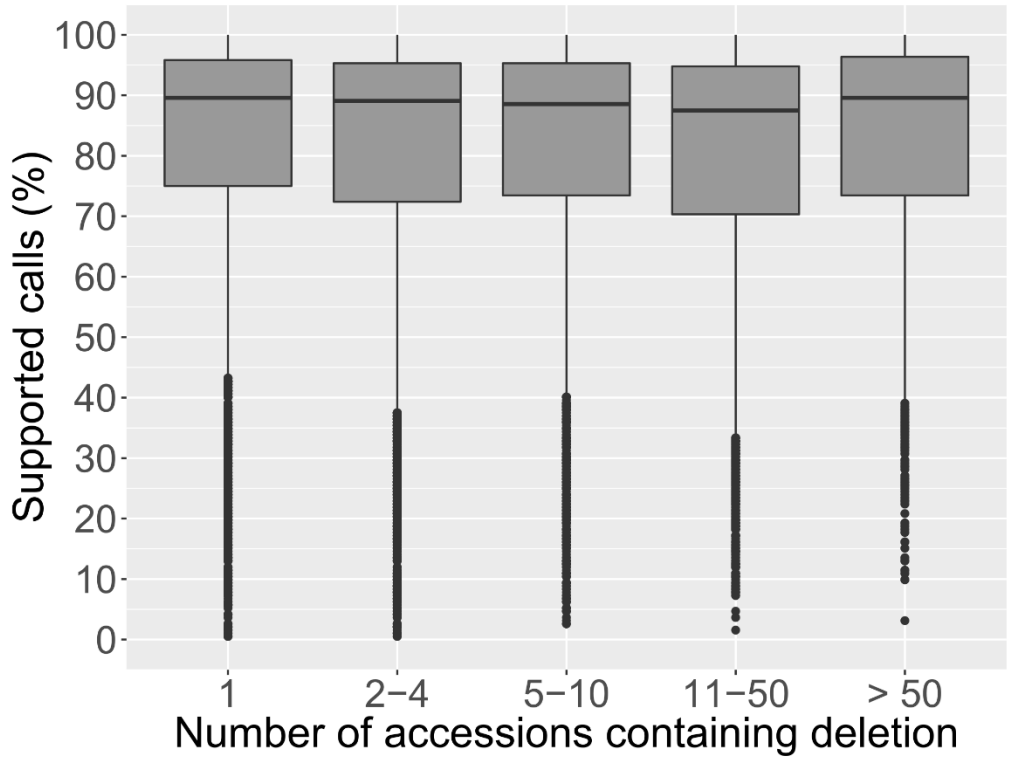

Figure S2: **Deletions in the DartMap panel were called with high precision.** The y-axis shows the percentage of genotype calls supported by a matching decrease (non-reference allele) or no change (reference allele) in read depth. Deletion variant sites are stratified along the x-axis by frequency of the non-reference (deletion) allele.

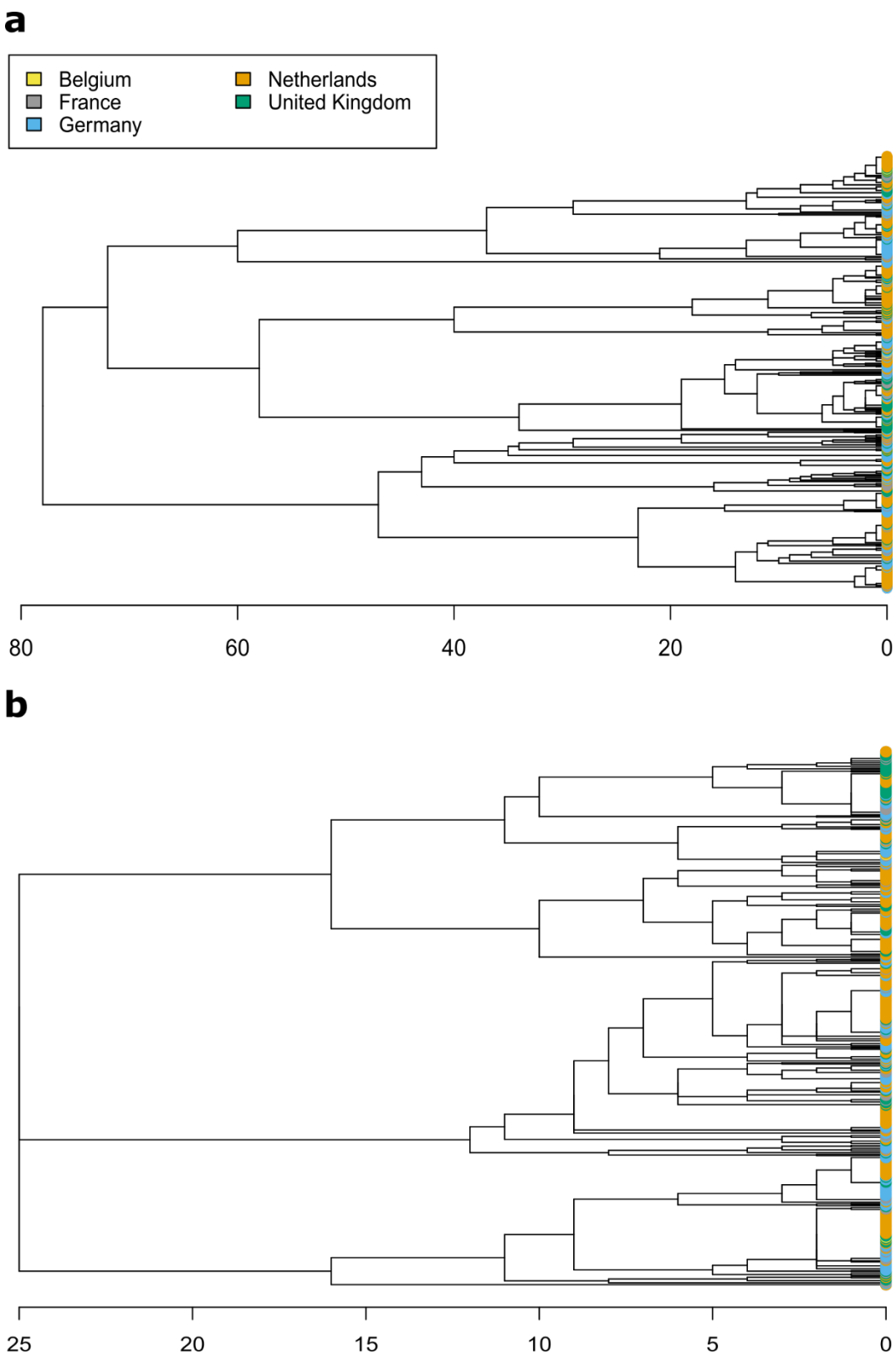

Figure S3: **Dendrograms based on organellar variation.** (a-b) Dendrograms constructed based on pairwise similarity of chloroplast (a) and mitochondrial (b) variants between Dutch accessions and those of nearby countries.

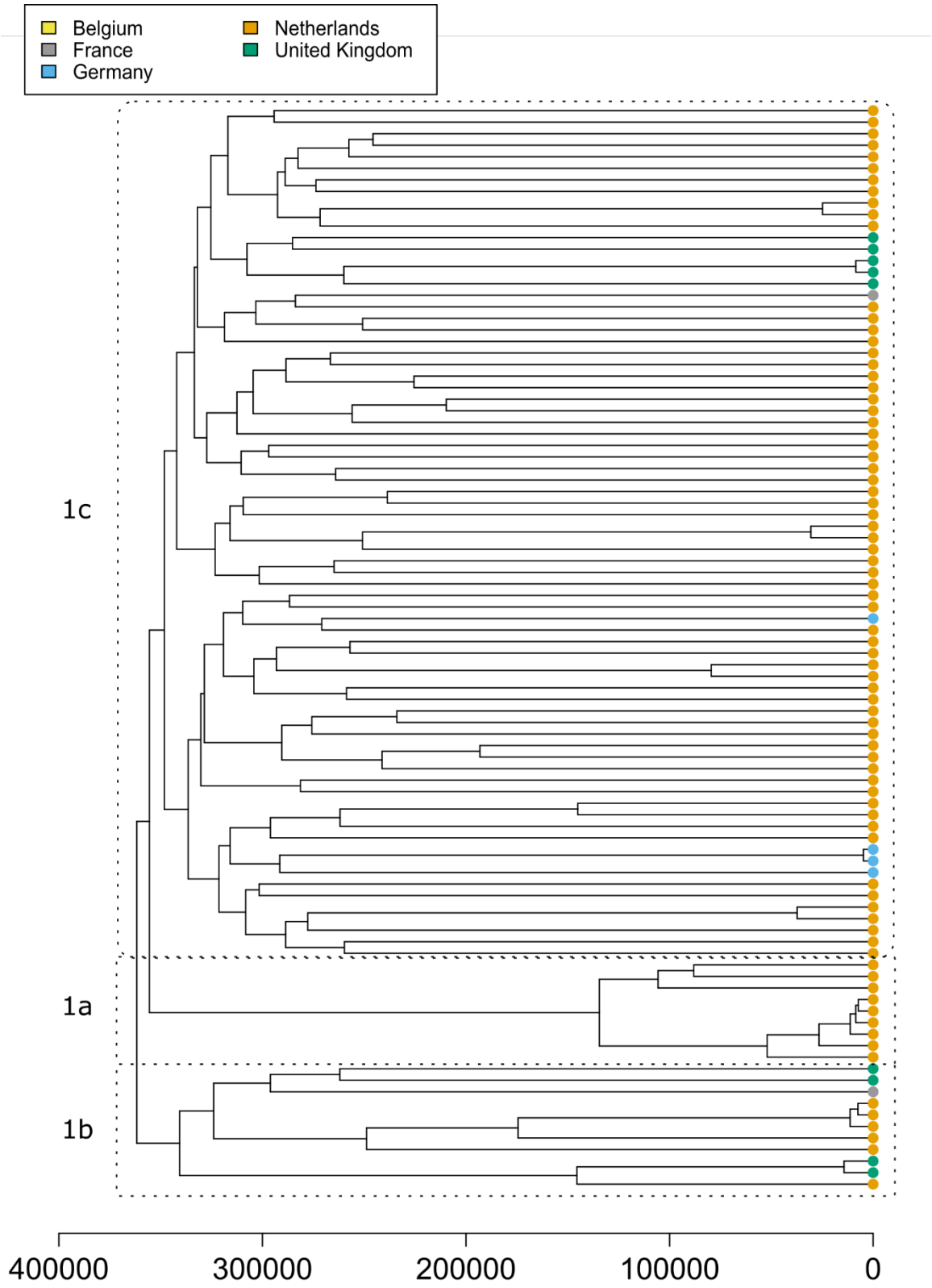

Figure S4: **Dendrogram of group 1 based on nuclear variation.** At its highest level, the dendrogram constructed based on pairwise similarity of nuclear variants between Dutch accessions and those of nearby countries can be split into two groups.
Group 1 and its three subgroups (1a, 1b, and 1c) are depicted here.

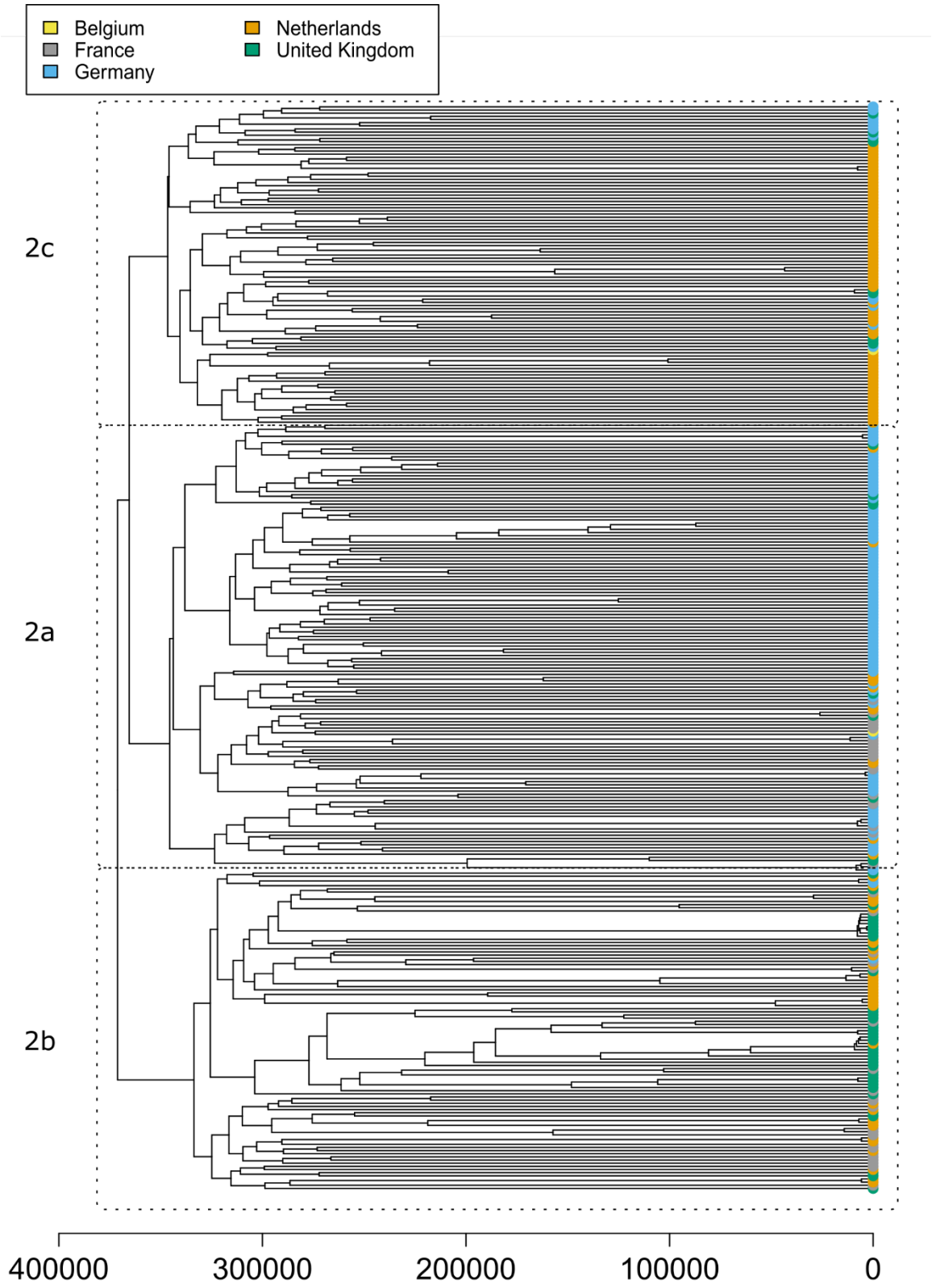

Figure S5: **Dendrogram of group 2 based on nuclear variation.** At its highest level, the dendrogram constructed based on pairwise similarity of nuclear variants between Dutch accessions and those of nearby countries can be split into two groups.
Group 2 and its three subgroups (2a, 2b, and 2c) are depicted here.

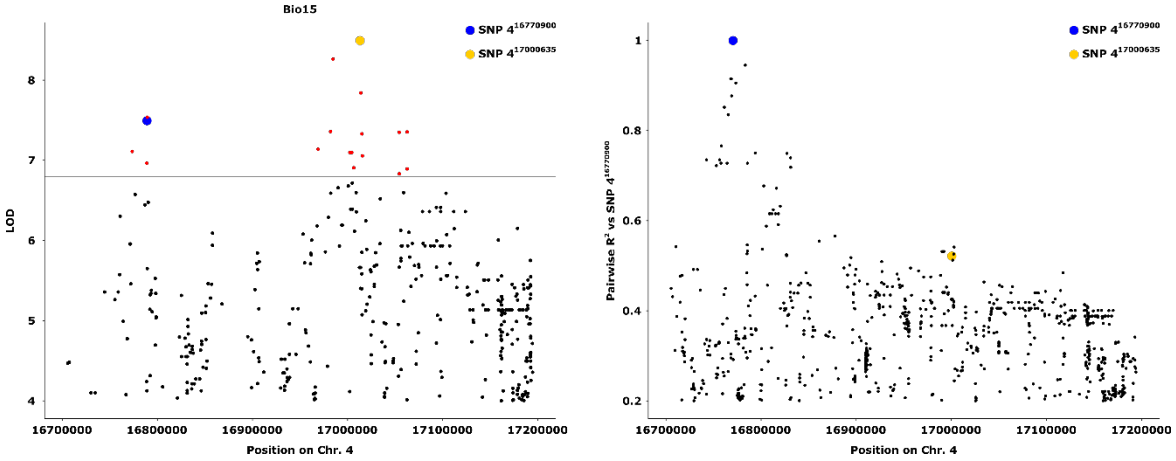

Figure S6: **Genomic region surrounding the chromosome 4 peak that was mapped for multiple climatic variables.** (a) A zoomed in section of the Chr. 4 peak of the mapping of precipitation seasonality (figure 5b) shows that there appear to be two nearby QTLs. The two SNPs that were found most frequently with the highest LOD score among the different mappings are depicted with a blue and yellow dot. b) Linkage analysis (expressed as the pairwise R^2) of SNPs in the Chr. 4 peak area relative to SNP $4^{16770900}$ (blue dot) reveals that the two QTLs are not in strong LD.

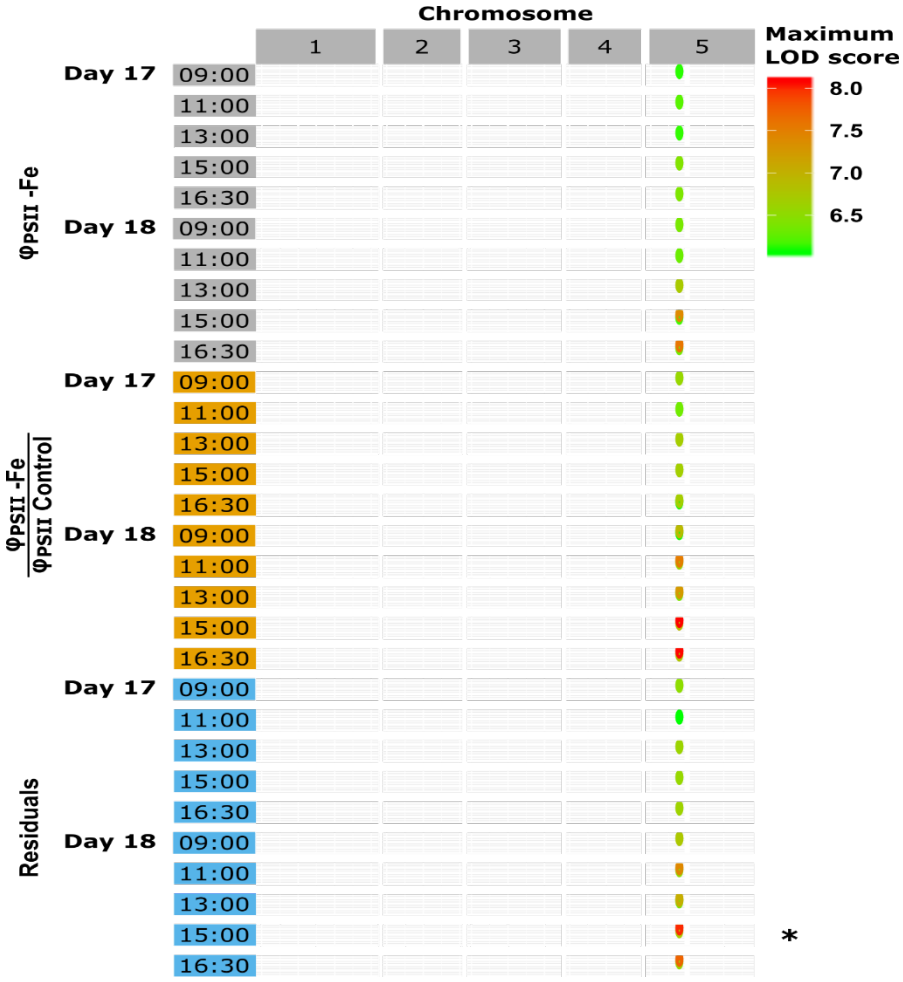

Figure S7: **Iron deficiency GWAS time series.** Summary data of the GWA results over 10 time points, measured during days 17 and 18 after sowing for three different phenotypes: [1] the average Φ_{PSII} per accession in iron deficient conditions (grey), [2] the ratio between Φ_{PSII} in iron deficient conditions and Φ_{PSII} in control conditions (orange), and [3] the residuals (blue). The dots on the graph represent the highest LOD scores in a 25 kb window size above a LOD threshold of 6. The * represents the data from the Manhattan plot in Figure 6b.

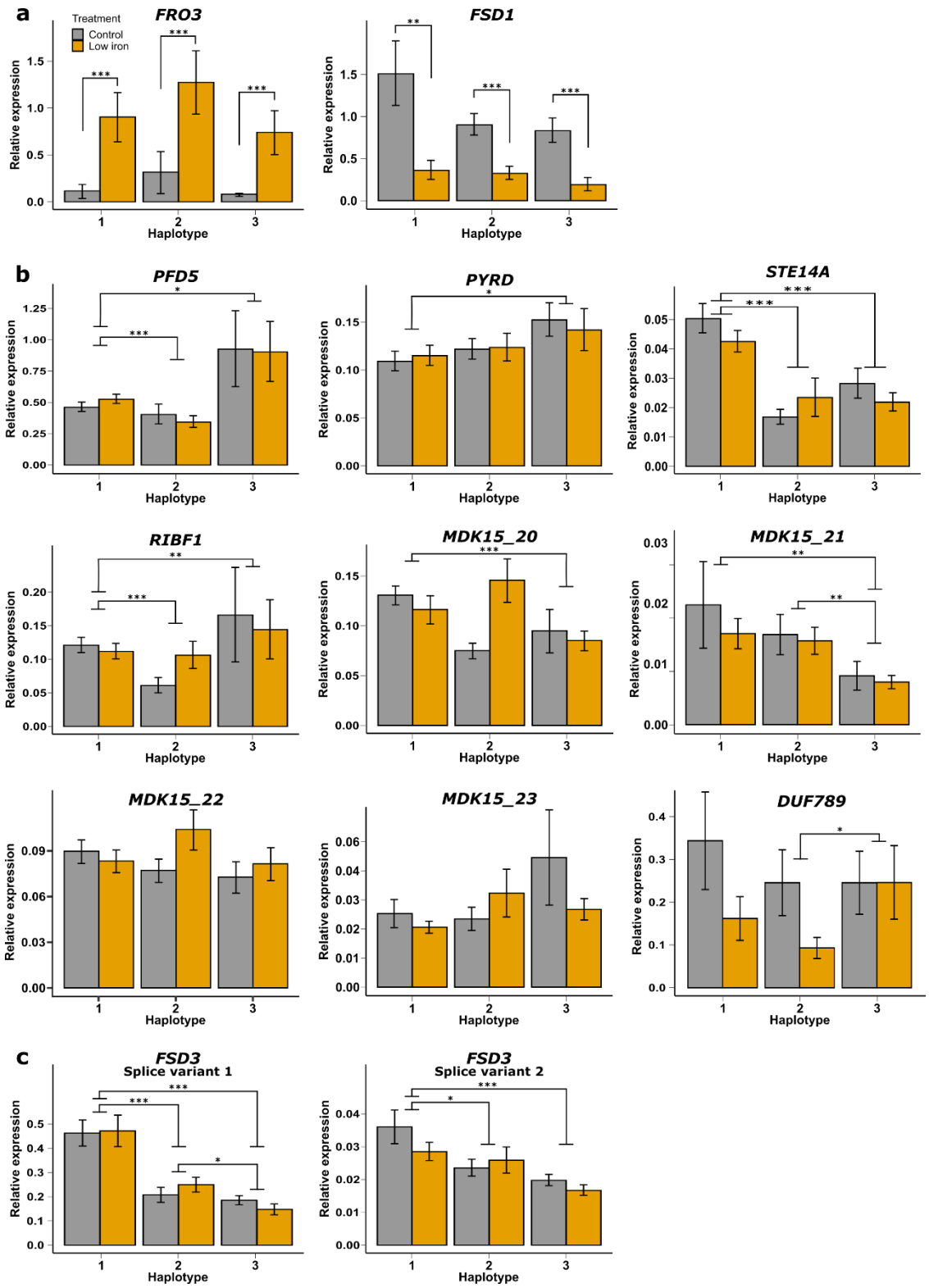

Figure S8: **Gene expression levels per haplotype group as measured by RTqPCR.** The average gene expression levels at (21µm Fe^{2+} (Control, depicted in dark grey) and (1µm Fe^{2+}) (Low iron, depicted in orange) per haplotype group for two iron deficiency marker genes (a), nine candidate genes from the GWAS (b) and two splice variants of *FSD3* (c). Average gene expression levels were taken from six natural accessions (with 5 replicates per accession per treatment) and normalized relative to three reference genes (*SAND*¸ *YLS8*, and *TIP41-like*). Significance of differences was assessed using two-sample t-tests on all Ct values from both treatments per haplotype. Error bars represent SE.*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

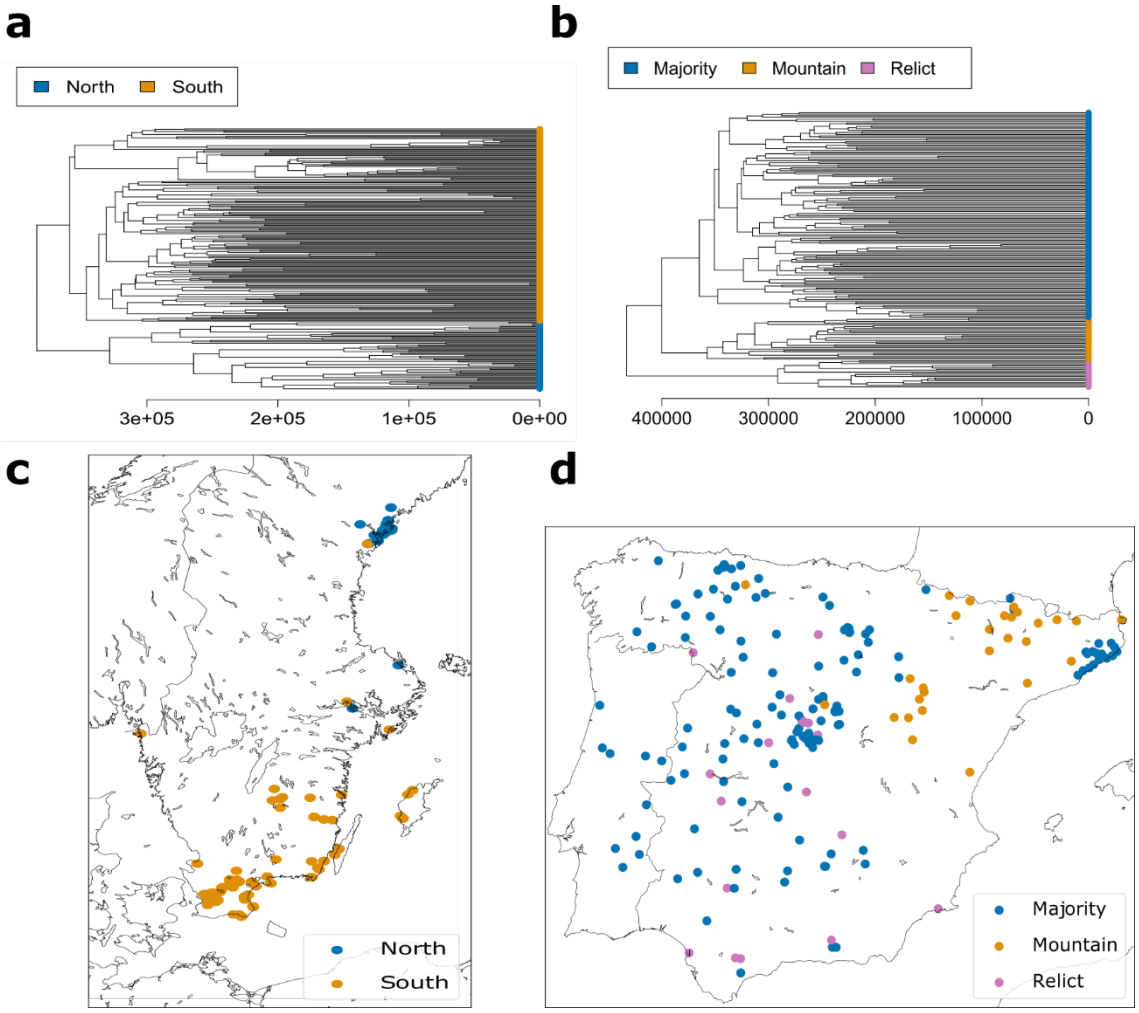

Figure S9: **Population structure of *A. thaliana* in Sweden and the Iberian peninsula. (a) Dendrogram of Swedish accessions.** (b) Dendrogram of Iberian accessions. (c) Geographical location of groups of Swedish accessions after cutting the dendrogram into two clusters. Groups correspond to accessions sampled in the north and south, as previously reported (Long et al., 2013). (d) Geographical location of groups of Iberian accessions after cutting the dendrogram into three clusters. The two outgroups of that are distinct from the majority correspond to relicts and accessions sampled in mountains, as previously reported (Tabas-Madrid et al., 2018).

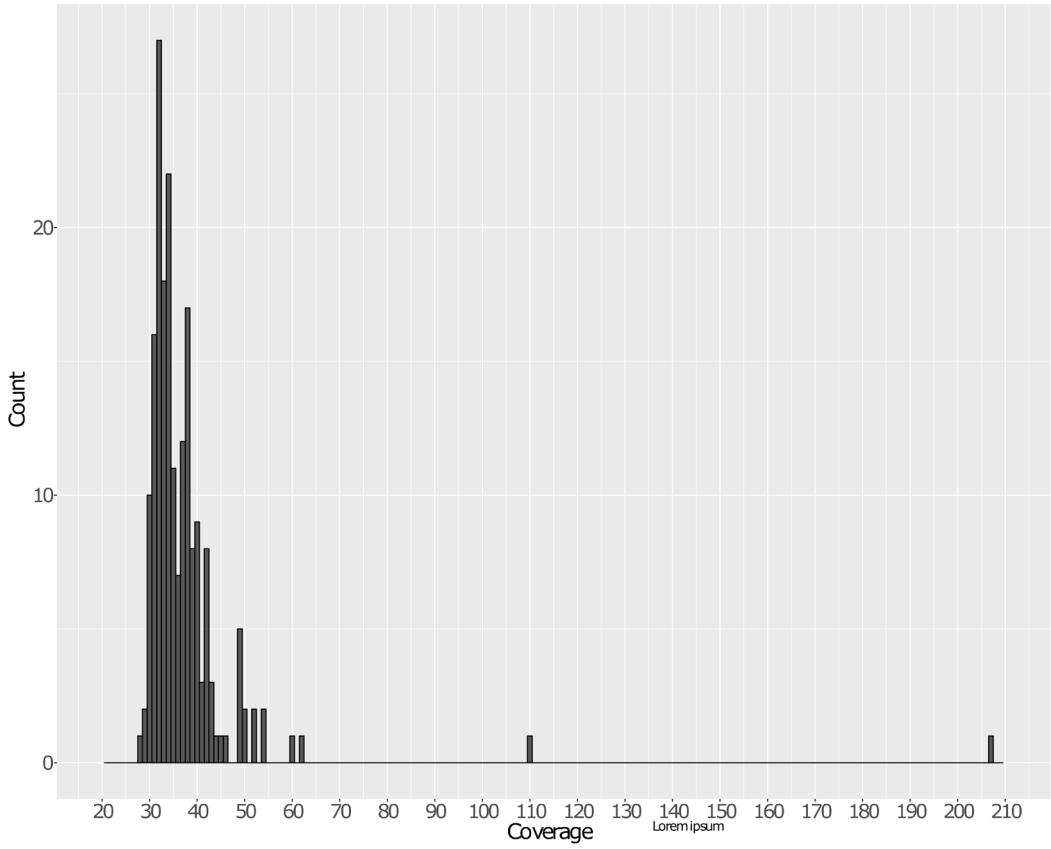

Figure S10: **Genomic coverage of each DartMap accession after pre-processing.** All samples have at least 28x coverage.

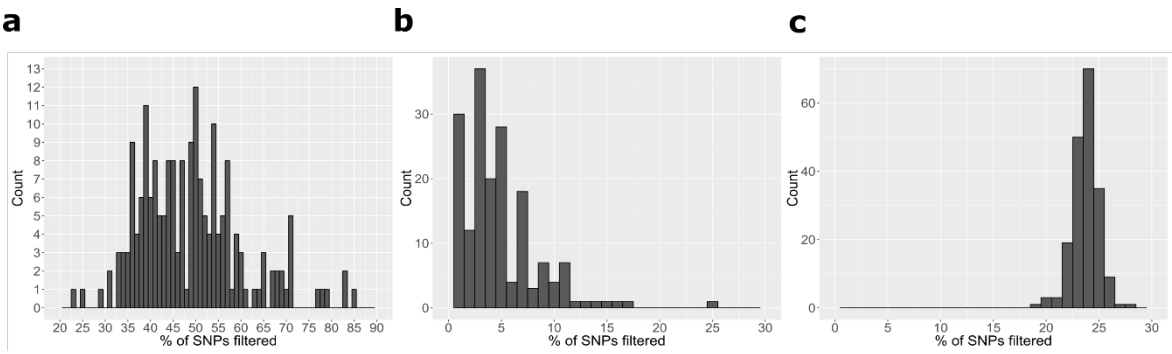

Figure S11: **Effect of filtering steps on number of retained SNPs for each sample.** The percentage of filtered SNPs is shown for the mitochondrial (a), chloroplast (b), and nuclear (c) callsets. The y-axes show the number of samples in which a particular percentage of SNPs (x-axes) were filtered.

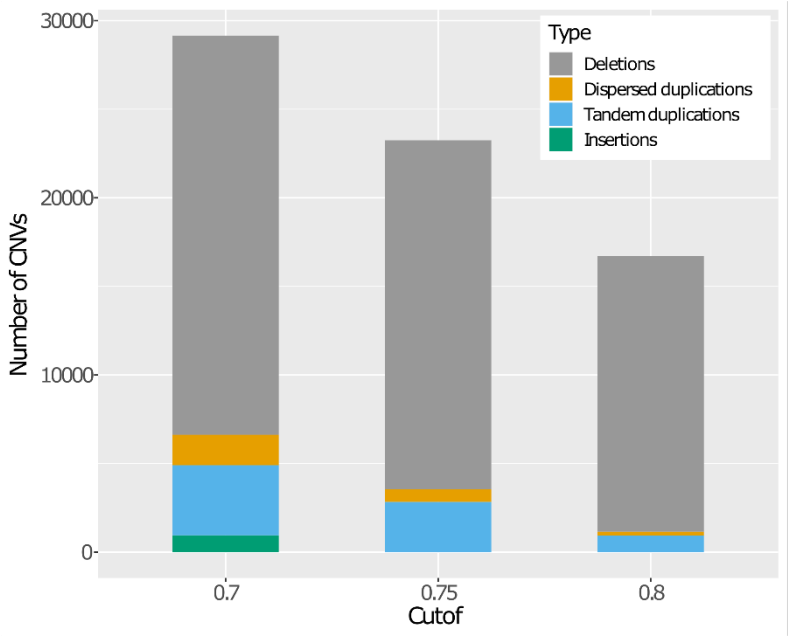

Figure S12: **Number and types of CNV detected at different cut-offs of Hecaton.** Insertions were not reported at high cut-offs, as Hecaton generally considers them to be inaccurate.