Allelic variants confer Arabidopsis adaptation to small regional environmental differences

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, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

[2]Laboratory of Genetics, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

[3]Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

[4]Department of Plant Physiology, Green Life Science Research Cluster, Swammerdam Institute for Life Sciences, University of Amsterdam, Sciencepark 904, 1098 XH Amsterdam, the Netherlands

[5]Biometris, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

[6]Current affiliation: Institute of Genetics, Faculty of Biology, Ludwig Maximilian University of Munich, Grosshaderner Str. 2-4, D-82152, Martinsried, Germany

[*]These authors contributed equally to this work.

[†]Corresponding authors: Dick de Ridder (dick.deridder@wur.nl), Mark Aarts (mark.aarts@wur.nl)

Running title: Genetics of local Arabidopsis adaption

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, due to mutation of the *GIBBERELLIC ACID REQUIRING 5* (*GA5*) gene, 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 *FE SUPEROXIDE DISMUTASE 3* (*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 *LA RELATED PROTEIN 1C* (*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

49 (Kawecki and Ebert, 2004). Unraveling the genomic and physiological basis of adaptation
50 in plants is a central question in modern plant biology. From an evolutionary
51 perspective, linking plant genotypes to adaptive traits helps to clarify how
52 environmental gradients drive natural selection of both standing and novel variation,
53 which ultimately may result in speciation (Sobel et al., 2010). It also enables us to model
54 whether natural populations are able to adapt to future environmental conditions (Bay
55 et al., 2017; Exposito-Alonso et al., 2019) and provides insights on processes or even
56 genes to focus on when breeding crops for enhanced abiotic stress tolerance (Huang
57 and Han, 2014).

58 The model plant species *Arabidopsis thaliana* provides an excellent system to
59 study natural plant adaptation (Weigel and Nordborg, 2015). Recent studies of this
60 species have shifted from studying the global HapMap population (Li et al., 2010),
61 traditionally used for genome-wide association studies (GWAS), to populations sampled
62 from regions containing a range of contrasting climates, including Sweden (Long et al.,
63 2013), the Iberian Peninsula (Tabas-Madrid et al., 2018), and the south-west of France
64 (Frachon et al., 2018), as these local populations are expected to minimize genetic
65 heterogeneity and thus provide more statistical power for detecting quantitative trait
66 loci (QTLs) (Brachi et al., 2013; Korte and Farlow, 2013). While studies on such regional
67 populations have successfully identified several adaptive loci (Frachon et al., 2018;
68 Tabas-Madrid et al., 2018; Arteaga et al., 2022), we know comparatively little about the
69 loci driving adaptation in regions having a milder diversity in environmental conditions.

70 The Netherlands, a region which has remained largely under-sampled so far,
71 provides an excellent opportunity to address this issue. This region covers an area
72 comparable in size to the south-west of France, but much smaller than Sweden and the
73 Iberian Peninsula. Compared to the other three regions, the Netherlands has only mild

74 climatic clines (https://www.knmi.nl/klimaat-viewer), but past work indicated that local
75 adaptation of *A. thaliana* may still be expected (Barboza et al., 2013). The Netherlands
76 contains a moderate frequency of accessions with short inflorescences, mediated by the
77 same loss-of-function allele of the *GA5* gene that managed to spread over more than
78 100 kilometres in the western part of the country (Barboza et al., 2013). There is reason
79 to believe that more such signatures of local adaptation can be found, for instance
80 based on the small differences in average temperature and annual precipitation, or the
81 soil type on which *A. thaliana* grows in the Netherlands (De Vries et al., 2003). While it
82 can be found on the clay and peat regions in the west and north of the country, it
83 mainly grows on sandy soils, which largely occur in the east and south, and in the
84 coastal dunes. It is often found in roadsides and gardens in which substantial movement
85 of soils and seeds from other regions has occurred in the past.
86 To comprehensively address the contribution of genetic variation to plant
87 adaptation in the Netherlands, we generated the Dutch *Arabidopsis thaliana* Map
88 (DartMap) panel, a collection of 192 *A. thaliana* accessions sampled from different sites
89 in the country. We show that this panel contains a surprisingly high level of genetic
90 diversity and provide evidence of local adaptation to mild climatic clines. Moreover, we
91 present strong evidence that variation at two loci is involved in mediating adaptation to
92 wind tolerance and iron deficiency. Our work demonstrates that plant populations
93 sampled in small geographic regions with a low diversity in environmental conditions
94 can contain a level of genetic variation that is sufficiently rich to facilitate local
95 adaptation.

96 # Results

97 ## The degree of *A. thaliana* genetic diversity in the Netherlands is typical of
98 that of most European regions

99 The global *A. thaliana* collection of the 1,001 Genomes (1001G) Project (The 1001
100 Genomes Consortium, 2016) includes only 11 lines from the Netherlands, which likely
101 represents only a small portion of the region's genetic variation due to limited dispersal
102 of minor alleles (Tabas-Madrid et al., 2018). Therefore, we assembled and analysed the
103 DartMap panel and used it together with the Dutch 1001G accessions to characterize
104 the overall degree of genetic diversity of *A. thaliana* in the Netherlands. In total, we
105 identified 2,712,612 SNPs and 353,974 indels in the DartMap panel and the Dutch
106 1001G accessions (collectively referred to as DartMap + 1001G panel for brevity)
107 relative to the *A. thaliana* Col-0 nuclear genome reference sequence. Moreover, we
108 detected 487 SNPs and 209 indels in the chloroplast sequence, and 137 SNPs and 15
109 indels in the mitochondrial sequence.

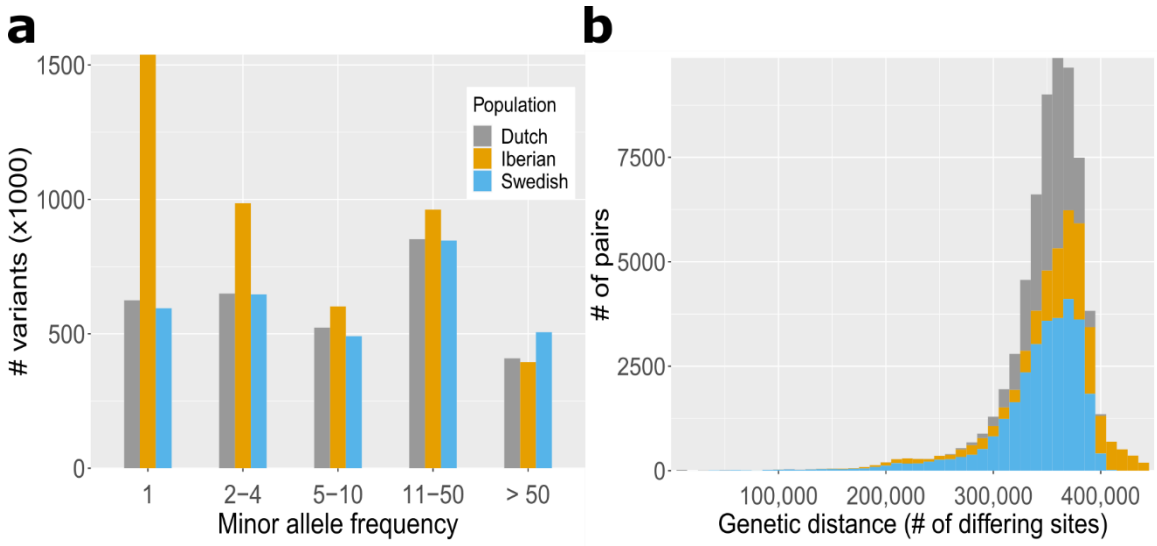
110 Furthermore, we examined copy number variation in the DartMap panel,
111 excluding the Dutch 1001G accessions because they were sequenced on a different
112 sequencing platform. We detected 29,155 copy number variants (CNVs) relative to the
113 Col-0 reference genome. Most of these are deletions,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

120 To contextualize the degree of genetic diversity of the DartMap + 1001G panel, we
121 compared it to similarly sized collections from Sweden (Long et al., 2013) and the
122 Iberian peninsula (Tabas-Madrid et al., 2018), using the same SNP and indel calling
123 approach. Considering only SNPs and indels that are polymorphic in the DartMap +
124 1001G panel (i.e. at least one accession contains a reference allele), the Swedish
125 collection contains a similar number of variants, while the Iberian one is more diverse
126 (Table 1). This disparity can be attributed to admixture of "relict" accessions in the
127 Iberian collection, which originated from *A. thaliana* populations in Europe before the
128 last ice age, and the genetically distinct "non-relict" group that repopulated Europe
129 thereafter (The 1001 Genomes Consortium, 2016).
130
131 Table 1: **The number of polymorphic bi-allelic SNPs and indels ("Variants") found in**
132 **regional *A. thaliana* populations.**

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

133
134 Most SNPs and indels detected in the DartMap + 1001G panel are shared among
135 multiple accessions and the distribution of their minor allele frequencies is similar to
136 that of the Swedish collection (Figure 1a). In contrast, the Iberian collection contains a
137 relatively larger fraction of rare variants (Figure 1a), possibly because the mountain
138 systems in this region strongly limit the spread of minor alleles (Tabas-Madrid et al.,
139 2018).

140 To assess the genetic diversity of the DartMap + 1001G panel, we calculated
141 pairwise genetic distances between individual accessions. Most pairs of accessions differ
142 at approximately 320,000-340,000 variant sites, similar to the Swedish collection and
143 most pairs of the Iberian collection (Figure 1b). However, the Iberian collection shows a
144 right-skewed distribution of pairwise genetic distances, indicating the presence of
145 accessions that are genetically more distant from each other compared to the majority
146 (Figure 1b). This tail in the distribution aligns with previous findings of genetic distance
147 between Iberian accessions and attributed to the presence of the relict accessions in
148 this region (The 1001 Genomes Consortium, 2016). Taken together, our analyses
149 indicate that genomic variation of *A. thaliana* in the Netherlands predominantly
150 originated from the post-glacial expansion event from which most *A. thaliana*
151 accessions in Europe descended (Lee et al., 2017).
152



153

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

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

159 We explored the demographic and evolutionary history of the DartMap + 1001G panel
160 using population structure analyses based on the genetic distances between individual
161 accessions, which have proven useful for this purpose in previous studies of *A. thaliana*
162 (The 1001 Genomes Consortium, 2016; Tabas-Madrid et al., 2018). We computed
163 population structure of the DartMap + 1001G panel through hierarchical clustering based
164 on pairwise organellar (Figure 2a, Table S2) and nuclear (Figure 2b, Table S2) genome-
165 wide similarity, including all accessions of the 1001G collection sampled in Belgium,
166 France, Germany, and the United Kingdom to evaluate relatedness between Dutch *A.*
167 *thaliana* accessions and those of nearby countries. The dendrograms based on
168 chloroplast and mitochondrial variation consist of three nearly equidistant groups (Figure
169 S3) with identical spatial distributions across the Netherlands (Figure 2a), as expected for
170 organellar sequences that are both predominantly maternally inherited. A previous
171 study showed that European accessions contain several highly diverged chloroplast
172 haplogroups of which the estimated time of divergence considerably predates the post-
173 glacial expansion event (Hsu et al., 2019). A principal component analysis, based on
174 chloroplast variation between individual accessions of the DartMap + 1001G panel and of
175 a global collection of *A. thaliana*, separates Dutch accessions in three different groups
176 (Figure 2c) that correspond to different subsets of chloroplast haplogroups identified in
177 the same study (Figure 2d). This implies that the distinct organellar groups of the
178 DartMap + 1001G panel represent ancient variation, while the nuclear groups reflect
179 more recent genomic differentiation, explaining the limited overlap in spatial distribution
180 between the two (Figure 2a and b).

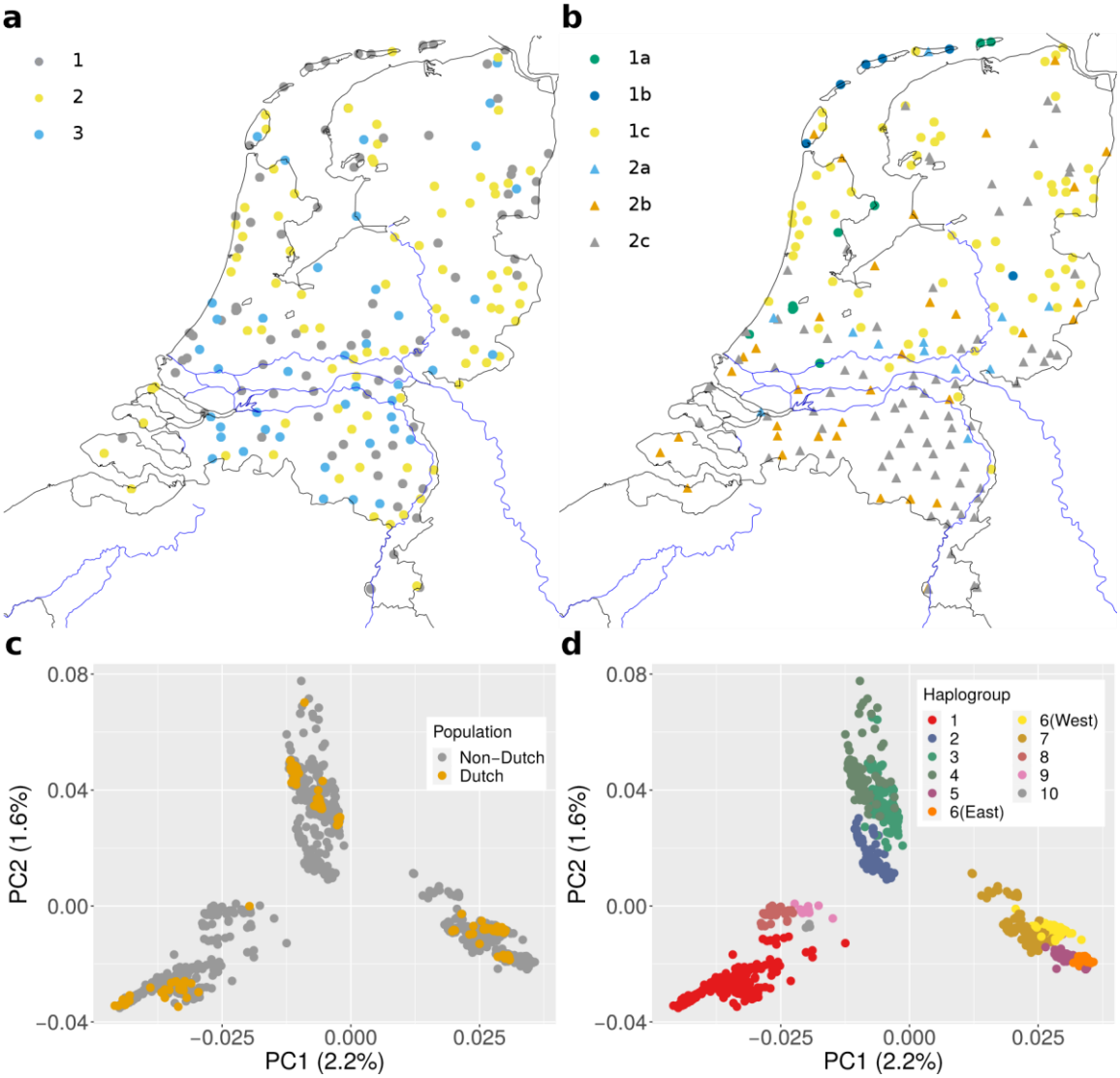
181 The spatial distribution of the nuclear groups provides insights into the factors
182 that shaped genomic variation in the DartMap + 1001G panel following the postglacial
183 expansion event. Based on nuclear variation, the panel can be divided into two main
184 groups, predominantly separated by the Rhine and Meuse river branches (group 1 to the
185 north and group 2 to the south) (Figure 2b). The rivers likely acted as natural dispersion
186 barriers (Pico et al., 2008), restricting genetic flow between the two groups. It is worth
187 noting that *A. thaliana* is generally not found on clay soil accompanying the rivers, except
188 on sandy river dunes and residential areas where it has been introduced through human-
189 induced transport of soil from other regions. Group 1 can be further divided into three
190 subgroups (Figure 2b), of which two contain accessions that are closely related to each
191 other (subgroups 1a and 1b) and less related to the rest of the population (Figure S4).
192 Subgroup 1a contains eight accessions of the Dartmap panel with short inflorescences,
193 referred to as "semidwarfs", and one accession (Tha-1) of the 1001G panel, previously
194 described as a semidwarf (Barboza et al., 2013). The accessions from subgroup 1a are
195 mainly found in the western part of the Netherlands, close to the North Sea coast (Figure
196 2b). Subgroup 1b is predominantly found on the Dutch islands in the north, implying that
197 these islands were colonized by a limited number of founder accessions, mostly from
198 group 1. Group 2 consists of accessions that are genetically similar to accessions from
199 Germany (subgroup 2a), respectively France and the United Kingdom (subgroup 2b)
200 (Figure S5), suggesting that these are part of a large genetic group spread over
201 Northwestern Europe.

202 Besides large clusters, we identified 27 pairs of accessions that are near duplicates
203 in terms of nuclear similarity (Table S3), with differences of fewer than 50,000 variant
204 sites between them. This level of similarity is significantly lower than the median pairwise

205 similarity of 354,388 variant sites (Figure 1a). Strikingly, they include 6 out of 8 of the

206 semidwarf accessions of subgroup 1a, the furthest of which are separated by more than

207 100 km (Table S3) (Flood et al., 2016a).



208

209 Figure 2: **Population structure of *A. thaliana* in the Netherlands.** (a-b) Geographical

210 location of Dutch accessions clustered based on organellar (a) and nuclear (b) genome-

211 wide similarity. Colours are used to distinguish accessions of different clusters. The main

212 rivers of the Netherlands are depicted by blue lines. (c) Principal component analysis of

213 chloroplast variants in Dutch accessions and the global collection of *A. thaliana* analysed

214 in a previous study (Hsu et al., 2019). (d) Samples in (c) coloured according to the

215 chloroplast haplogroups reported in a previous study (Hsu et al., 2019).

216

217 Semidwarf *ga5* mutants display tolerance to windy conditions

218 The widespread geographical distribution of highly identical genotypes that all have the
219 same semidwarf phenotype suggests there may be a selective advantage to this trait,
220 similar to the near-identical and widespread atrazine-resistant genotypes distributed
221 along the United Kingdom railway system, which were identified previously (Flood et al.,
222 2016a).

223 An earlier study already indicated a higher frequency of semidwarfs in the
224 Netherlands compared to the global estimated frequency of at least 1% (Barboza et al.,
225 2013). Next to the eight semidwarfs found in subgroup 1a, we identified a ninth
226 semidwarf accession, Snh-1, in subgroup 2b. Almost all examined semidwarf accessions
227 found in the field are loss-of-function mutants at the *GA5* locus, which encodes the
228 *GIBBERELLIN 20-OXIDASE 1* (*GA20OX1*) gene (Barboza et al., 2013). Further investigation
229 of the genomic sequences of *GA5* revealed that the eight semidwarf accessions in
230 subgroup 1a, that are genotypically nearly identical, share a previously described *ga5*
231 splice site mutation specific to Dutch semidwarfs (Barboza et al., 2013), while Snh-1
232 contains a 52-bp deletion in the last exon of *GA5*.

233 The appearance of a relatively high frequency of natural *ga5* mutants suggests
234 that semidwarfism provides an adaptive advantage. Dwarfism in plants is a common
235 adaptation to high altitude (Billings and Mooney, 1968; Grace, 1988; Luo et al., 2015),
236 although it remains unclear which aspect of high altitude causes this. The majority of
237 the Dutch semidwarfs are collected within 15 km of the coast, often only a few meters
238 above sea level, so there is no association with high altitude. However, the locations
239 where the semidwarfs were collected share an important characteristic providing a
240 potential selective force, namely the presence of high wind speeds (Figure 3a). High

241 wind speeds are also a characteristic of high-altitude environments. We therefore
242 investigated whether the presence of high wind speeds could create a selective
243 environment favouring semidwarf *A. thaliana* over normal-sized plants.
244 We grew six Dutch semidwarf accessions and five unrelated tall accessions (Table
245 S4), in a climate-controlled growth chamber equipped to provide controlled wind
246 speeds typically experienced in coastal areas of the Netherlands in March/April.
247 Moreover, we included Col-0, L*er* and the *ga5* loss-of-function mutants in Col-0 and L*er*
248 backgrounds for a direct comparison of near isogenic lines differing only at the *GA5*
249 locus. Despite our preselection for similar flowering time (Table S4), based on previous
250 data collected on greenhouse-grown plants, the semidwarfs flowered significantly later
251 in the controlled growth room conditions than the taller accessions (Figure 3b). This
252 likely had little effect on our final results though, as all accessions were exposed to
253 windy conditions for the same time and all plants started to flower before the end of
254 the experiment. The tall accessions were strongly affected by the wind, as evidenced by
255 a significant reduction in stem length (approximately 30% on average) and fresh weight
256 (approximately 36% on average) (Figure 3c and 3d). In contrast, the semidwarfs also
257 showed a significant reduction in stem length, but the average reduction was only
258 around 10%. No significant decrease in fresh weight was observed (Figure 3d). Similar
259 effects were observed for Col-0 and L*er*, where the wind treatment significantly
260 decreased their stem length and fresh weight, whereas the respective *ga5* loss-of-
261 function mutants remained unaffected (Figure 3c and d). Collectively, these results
262 strongly suggest that the semidwarf phenotype caused by the *ga5* mutation provides a
263 strong advantage with respect to plant biomass production, when growing in conditions
264 of high wind speeds.
265

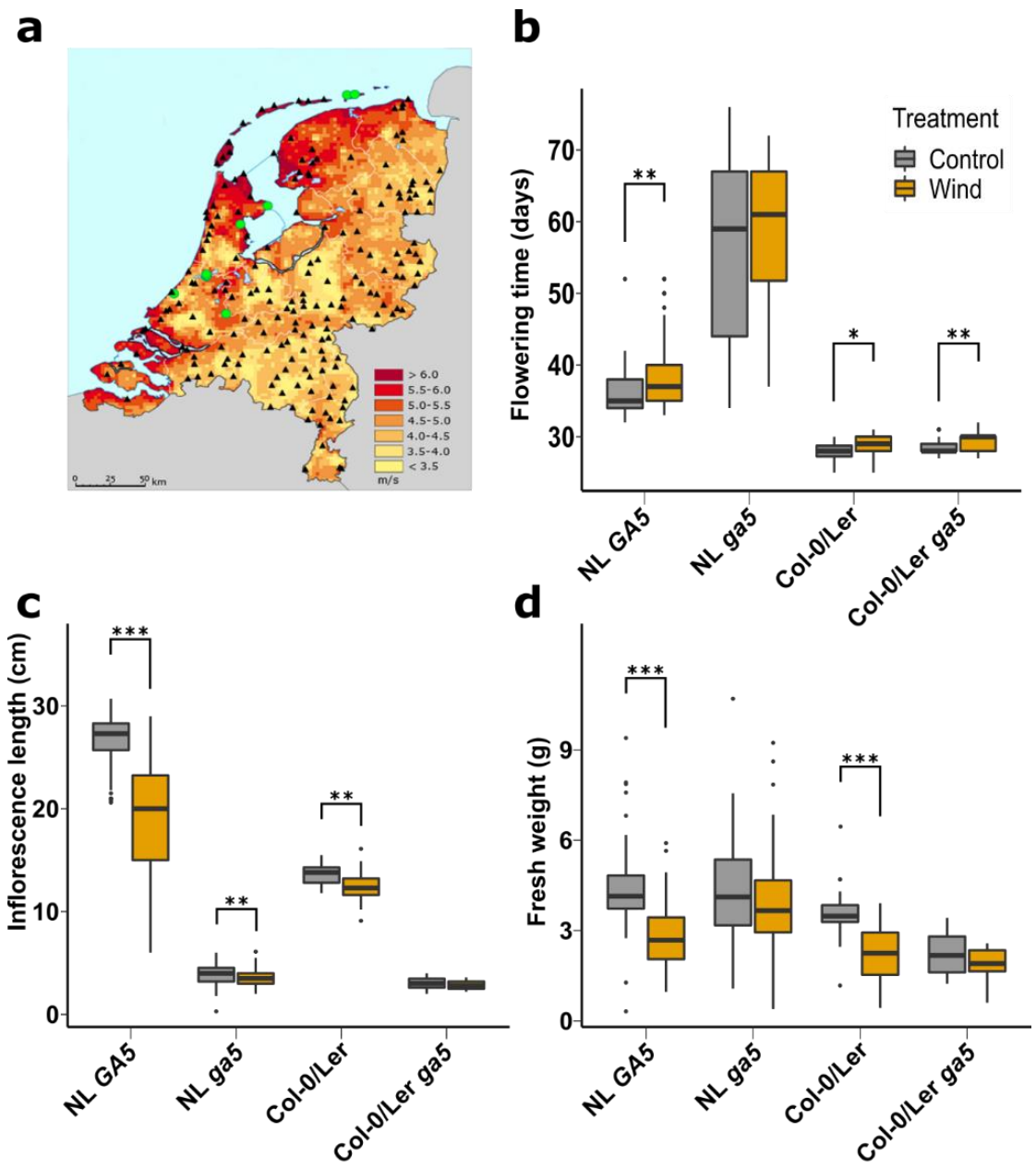


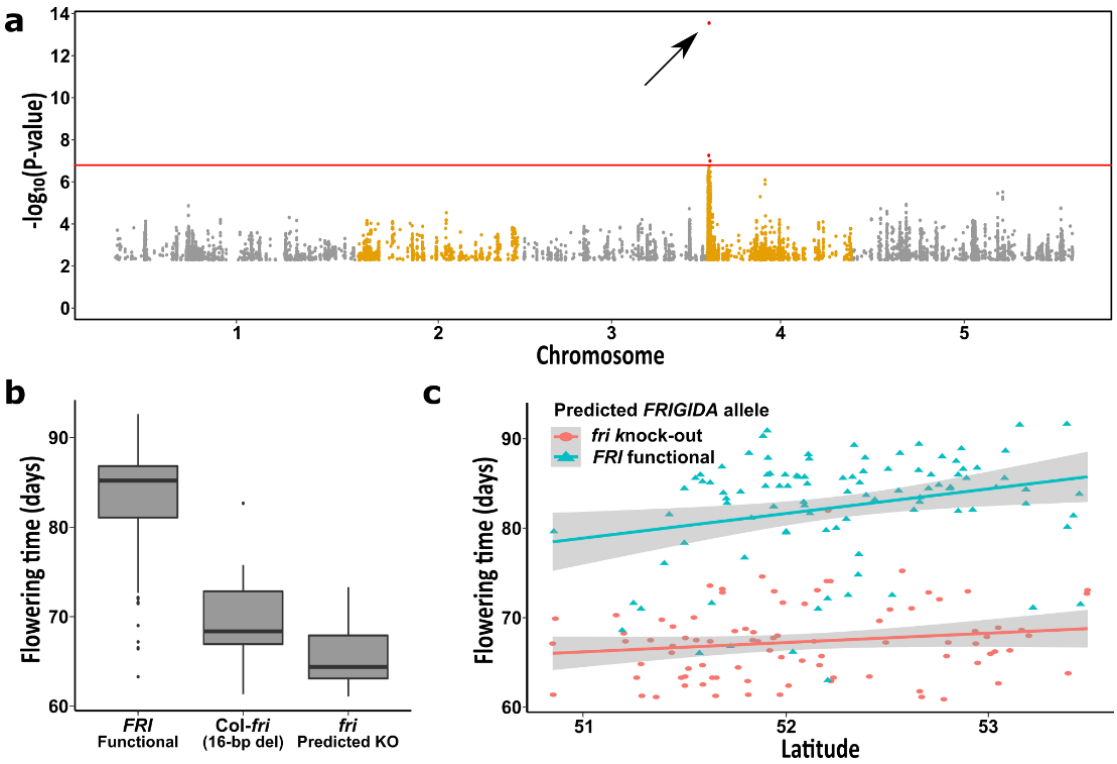
Figure 3: **Phenotypic response to windy conditions.** (a) Average annual wind speed (between 1981-2010) in the Netherlands as measured at ground level. Green circles indicate semidwarf accessions that share the same splice site mutation, black triangles indicate all other DartMap accessions. Figure 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 $*P < 0.05$;

275 $^{**}P < 0.01$; $^{***}P < 0.001$, two-tailed Student's *t*-test. NL – Dutch accessions; *GA5* – wild-
276 type allele, *ga5* – mutant allele.
277

278 Flowering time variants display a latitudinal cline

279 The transition from vegetative to reproductive stage is a key life history trait in *A.*
280 *thaliana*, influenced by seasonal cues like day length and temperature (Andrés and
281 Coupland, 2012). Previous GWAS of flowering time yielded few significant associations,
282 mainly due to allelic heterogeneity (Atwell et al., 2010; Tabas-Madrid et al., 2018; Zhang
283 and Jiménez-Gómez, 2020). We expect this should be less of a problem in a regional
284 population with more closely related individuals. Therefore, we performed GWA
285 analysis of flowering time, with and without vernalization, using the DartMap panel. We
286 discovered a highly significant association between flowering time and a 16-bp indel
287 near the end of the first exon of the *FRIGIDA* (*FRI*) gene (position 269960 on
288 chromosome 4 of Col-0), irrespective of vernalization treatment (Figure 4a). This indel is
289 known to cause a premature stop codon in the *FRI* coding sequence, leading to a non-
290 functional *FRI* allele (Koornneef et al., 1994; Johanson et al., 2000; Andrés and Coupland,
291 2012). *FRI* is a key regulator of flowering time in *A. thaliana* and loss-of-function alleles
292 are widespread in nature (Zhang and Jiménez-Gómez, 2020). As the variant we picked
293 up by GWA analysis is unlikely to be the only loss-of-function mutation to occur in the
294 Netherlands, we examined all polymorphisms in the *FRI* coding sequences (CDS) in the
295 DartMap and found a total of 41 *FRI* alleles. When compared to the *FRI-H51* allele, that
296 encodes a fully functional protein and is regarded as the ancestral *FRI* allele based on
297 sequence comparisons with *A. lyrata* (Le Corre et al., 2002; Zhang and Jiménez-Gómez,
298 2020), another 11 loss-of-function *fri* alleles are identified across 88 accessions. The
299 accessions with a loss-of-function allele are on average early flowering (Figure 4b).

300 Although the Netherlands has a small latitudinal range, we examined if there is a
301 latitudinal cline in the diversity panel regarding flowering time, consistent with previous
302 findings in Northern European and Mediterranean accessions (Stinchcombe et al., 2004).
303 While we did not observe such a cline when considering the entire population, when
304 excluding the accessions carrying the *fri* loss-of-function alleles, we indeed found such a
305 latitudinal cline (Figure 4c). This means other loci, next to *FRI*, are involved in controlling
306 flowering time in the DartMap collection. These analyses collectively demonstrate the
307 suitability of the DartMap panel for conducting GWA analyses on phenotypic variation
308 relevant to local adaptation.
309



310

311 Figure 4: **GWA analysis of flowering time in the DartMap panel upon vernalization.** (a)
312 Manhattan plot for days until flowering after a six-week vernalization treatment,
313 identifying one highly associated polymorphism located in the *FRIGIDA* gene (arrow).
314 The red line corresponds to a permutation-based threshold level of significance at $P <$
315 0.05. (b) Flowering time distribution of three *FRIGIDA* (*FRI*) variants in the DartMap

316 panel. Accessions with a functional *FRI* allele (*FRI F*unctional), accessions with the Col-0
317 variant (Col-*fri*) that corresponds to a 16-bp deletion causing a premature stop codon in
318 the *FRIGIDA* coding sequence and accessions with a predicted functional knock-out
319 allele of *FRI* (*fri* Predicted KO). (c) Average days until flowering for DartMap accessions
320 relative to their latitudinal site of origin with linear fit for accessions with a predicted
321 loss-of-function *fri* allele (red, $R^2 = 0.016$, $P = 0.128$) or a functional *FRI* allele (blue, $R^2 =$
322 0.082, $P = 0.002$).
323

324 Local adaptation to the Dutch climate

325 The Dutch climate is characterized as a temperate maritime climate, which is relatively
326 uniform across the Netherlands. However, there are mild and gradual clines in
327 environmental variables, with slightly higher temperatures and less precipitation in
328 inland areas compared to coastal areas. One advantage of having a densely and
329 uniformly sampled population is the potential to detect whether local adaptation to
330 these gradual environmental clines has influenced genetic diversity. To test this
331 hypothesis, we conducted a genome-environment association (GEA) analysis on 19
332 quantitative climatic variables related to temperature and precipitation (Table S5 and
333 Figure 5a as example). We identified a few QTLs for these variables, but the most
334 notable QTL is located on Chr. 4 (Figure 5b). We named it *qDPT* (QTL for the Dutch
335 Precipitation and Temperature climatic variables). This QTL is associated with seven
336 climatic variables reflecting aspects of temperature and precipitation (Table S5).
337 Further examination of *qDPT* revealed that it appears to consist of two separate
338 QTLs, approximately 230 kilobase pairs apart (Figure S6a). The SNPs that are found most
339 frequently with the highest association in each QTL are SNP $4^{16770900}$ and SNP $4^{17000635}$
340 with minor allele frequencies (MAF) of 14.2% and 17.5%, respectively. To determine

341 whether the -log(*P*) peaks represent distinct QTLs or a single one, we calculated pairwise
342 r^2 values between SNP $4^{16770900}$ and all other biallelic SNPs in the selected genomic
343 region (16,700,000–17,200,000 bp). This revealed only weak evidence of linkage
344 disequilibrium (LD) (pairwise $r^2 = 0.52$) between the two SNPs most strongly associated
345 with the trait (Figure S6b). Hence our conclusion these are two separate QTLs. For
346 convenience, we will refer to them as '*qDPT-A*' and '*qDPT-B*', with their corresponding
347 SNPs labelled as A/a for SNP $4^{16770900}$ and B/b for SNP $4^{17000635}$.
348 The GEA analyses provide insights into whether specific alleles are correlated
349 with particular climatic variables. A strong association at a specific locus can indicate
350 that allelic variation at that locus is important for local adaptation (Ferrero-Serrano and
351 Assmann, 2019). However, such a strong correlation could also arise by accident, if the
352 climatic variable happens to align with the population structure. We confirm that this is
353 not a case of such a potential statistical artifact. First of all, kinship correction is used to
354 perform the GEA analyses, and this should drastically reduce the chance of finding loci
355 that are strongly associated due to the population structure. Secondly, we observe that
356 the geographical distribution of the accessions contributing to the population structure
357 appear to be mostly affected by the river systems (Figure 2b), which can be regarded as
358 a South – North cline. The climatic variables that map at *qDPT* follow a mostly East –
359 West cline. Moreover, the minor allele of *qDPT-A* is found in five out of six nuclear
360 subclusters (the common allele is found in all subclusters), and for *qDPT-B* both alleles
361 are found all subclusters. In conclusion, identification of these QTLs is not an artifact
362 due to the general population structure and thus likely to represent a locus associated
363 with plant adaptation to the environment.
364 It is not immediately apparent to which specific climatic variable(s) this locus
365 contributes. The climate in the Netherlands is primarily influenced by the North Sea,

366 which affects various environmental factors such as temperature, wind speed and
367 precipitation patterns and variability. Therefore, it is not surprising that the same peak
368 is consistently observed for several climatic variables, given the high correlation among
369 the variables associated with *qDPT*. Consequently, unveiling the particular climatic
370 variable or combination of variables to which the adaptation is associated with this
371 specific locus may be challenging. However, we noted that the climatic variables
372 associated to *qDPT* are all connected to seasonality, while annual means in precipitation
373 and temperature are not associated. This suggests that the adaptation is related to
374 conditions that prevail during certain parts of the year, such as one or two seasons, or
375 to adaptations that enable the plant to effectively tolerate greater variability. To explore
376 which climatic factor may be involved in this regard, we tested if allelic variation at *qDPT*
377 resulted in a differential response to a gradual decrease in water availability, which is
378 likely to be seasonal.

379 As *qDPT-A* and *qDPT-B* are in close proximity, we selected the best possible set of
380 accessions distinguishing best the combinations of alleles for these QTLs aiming to
381 disentangle the individual contributions of the QTLs. Given the available genotypes of
382 accessions in the DartMap population, this results in 13 accessions with 'AB', 7 with 'Ab',
383 1 with 'aB' and 7 with 'ab' (Table S6). Initially, all accessions were grown under well-
384 watered conditions. Subsequently, after two weeks, half of the replicates for each
385 accession were no longer watered for eleven days. We refer to this as the drought
386 treatment. We regularly measured the projected leaf area (PLA) during this period in the
387 control and the drought treatment. Due to an uneven distribution of accessions across
388 the genotypic combinations for these QTLs, we compared accessions with 'Ab'
389 genotypes against those with 'ab' genotypes (7 accessions per group) to assess the
390 effect of *qDPT-A*. For *qDPT-B*, all 28 accessions (14 accessions per *qDPT-B/b* allele) were

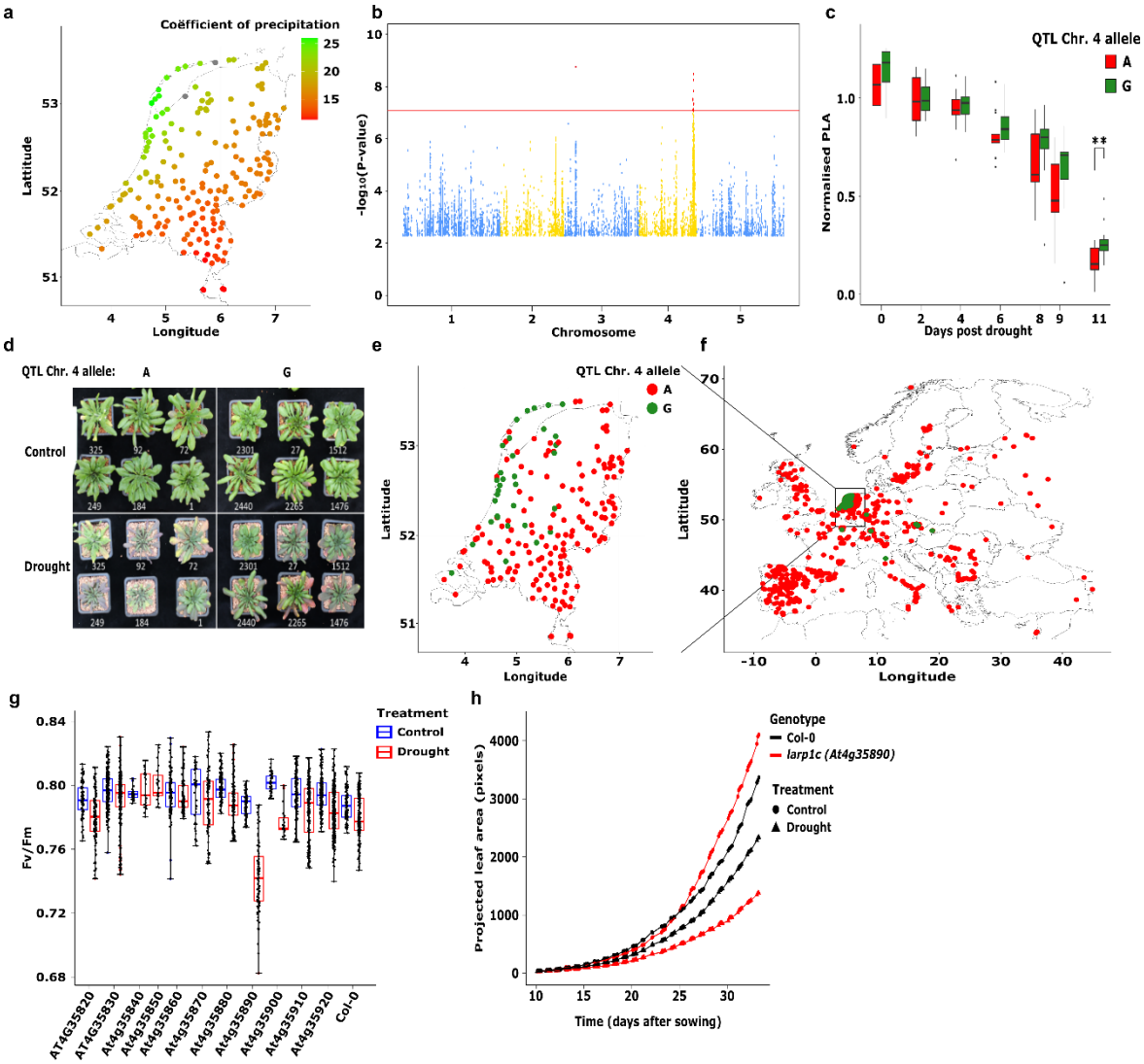
391 included in the analysis. Our results indicated a significant effect of *qDPT-B* on the PLA
392 response to the drought treatment (two-way ANOVA with *qDPT-B/b* allele and days
393 after last watering as variables; $P = 0.003$) (Figure 5c and 5d). However, we did not
394 observe a significant effect for *qDPT-A*. Subsequently, we examined the distribution of
395 *qDPT-B* alleles across the Netherlands and the 1001G panel. We found that SNP $4^{17000635}$
396 (*qDPT-B*) has a much higher MAF in the Netherlands (MAF = 17.5%) (Figure 5e)
397 compared to the 1001G panel (MAF = 3.1%) (Figure 5f).
398 We propose that the region encompassing the genomic locus with a LOD > 6,
399 which is consistently observed across different mappings, is the most likely location for
400 the gene(s), allelic variation of which causes the QTL. This region comprises 13 genes
401 spanning from At4g35820 to At4g35940 (Table S7). Within this region, there are a few
402 genetic variants that are in linkage disequilibrium ($r^2 > 0.4$) with SNP $4^{17000635}$ and that
403 potentially impact the encoded protein sequence of the gene they reside in.
404 Noteworthy are two non-synonymous variants in *OSCA4.1* (At4g35870), also known as
405 *EARLY-RESPONSIVE TO DEHYDRATION STRESS 4* (*ERD4*). The first variant constitutes a
406 change from methionine (neutral-polar) to isoleucine (hydrophobic) at the 10th position,
407 while the second variant changes an isoleucine to leucine (both hydrophobic) at the
408 641st position. These substitutions constitute relatively minor sequence differences that
409 are unlikely to affect the functionality of the protein encoded by the *ERD4* gene. We
410 tested T-DNA insertion lines for 11 of the 13 candidate genes in response to control and
411 drought treatments and measured F_v/F_m as a proxy of plant stress (Maxwell and
412 Johnson, 2000) (Figure 5g). For the remaining two genes, At4g35930 and At4g35940, no
413 T-DNA insertion mutant lines were available. The mutant of *LARP1c* (At4g35890) has
414 significantly lower F_v/F_m values in the drought treatment compared to Col-0, while it
415 does not differ from Col-0 in the control treatment (Figure 5g). Furthermore, the mutant

416 PLA is slightly higher, though not significantly different from Col-0 PLA under the control

417 treatment, but it is significantly impaired under the drought treatment (Figure 5h).

418 These results pinpoint *LARP1c* as the most prominent candidate gene to underlie the

419 difference in response to drought.



420

421

422 Figure 5: **Genome-environment association (GEA) analysis on climatic variables related**

423 **to precipitation and temperature.** (a) Geographical distribution of precipitation

424 seasonality (coefficient of variation) expressed in Seasonality Index (SI) units for the

425 DartMap. SI values <0.19 are defined as 'precipitation throughout the year', whereas

426 values between 0.20 and 0.39 are defined as 'precipitation throughout the year, but

427 with a definite wetter season'. (b) Manhattan plot for the GEA related to precipitation

428 seasonality (coefficient of variation). (c) Response to the drought treatment for
429 accessions differing at their *qDPT-B* allele. Boxplots represent the normalised projected
430 leaf area (PLA) calculated per accession by dividing the PLA in the withholding watering
431 treatment over the PLA in the control treatment for a set of 14 randomly chosen
432 accessions per allele (depicted as an adenine (A) or guanine (G) nucleotide at SNP
433 $4^{17000635}$). (d) Pictures of six representative accessions per *qDPT-B* allele in the control
434 and drought treatment, taken 11 days after the start of the treatment. (e) Geographical
435 distribution of the two alleles ('A' and 'G') across the Netherlands and (f) Europe. (g))
436 F_v/F_m as a proxy of plant stress of Col-0 and candidate gene T-DNA insertion lines. (f)
437 The response of PLA of Col-0 and the *larp1c* mutant in the control and drought
438 treatment.
439

440 Natural allelic variation for *FSD3* affects iron deficiency tolerance

441 Although the Netherlands is relatively uniform in climate, the soil type composition
442 differs substantially across the country, with clay, peat and sand forming the major
443 types (Hartemink and Sonneveld, 2013). This could be an important driver for local
444 adaptation in the Netherlands, as nutrient composition and especially nutrient
445 availability (Veer, 2006), differs between soil types. Iron (Fe) in particular is known to
446 vary in availability and shortage of Fe will severely limit plant growth (Hell and Stephan,
447 2003). We determined the light-use efficiency of photosystem II (PSII) electron transport
448 (Φ_{PSII}) of the DartMap diversity panel grown under control and Fe-deficient conditions,
449 through chlorophyll fluorescence measurements. Low Φ_{PSII} can serve as an early
450 indicator of Fe deficiency (Terry, 1980), as Fe is an essential component of iron-sulphur,
451 and heme proteins that play pivotal roles in photosynthesis and respiration (Hänsch and
452 Mendel, 2009).

453 The low Fe supply leads to only mild visible symptoms, such as reduced plant
454 growth and mild leaf chlorosis, but causes extensive variation for Φ_{PSII}, much more than
455 in the control condition (Figure 6a). GWA mapping identified a single associated region,
456 on chromosome 5, with the most significantly associated SNP located at position
457 7,856,132 (Figure 6b). The same locus, and the same SNP, is also found when the
458 average Φ_{PSII} values in the Fe-deficient condition, and the ratio between Φ_{PSII} in Fe
459 deficiency versus control, are used as traits for GWA mapping (Figure S7). To identify
460 candidate genes for the causal allelic variant, we limited the associated region to 22 kb,
461 based on the genome positions of SNPs that are in linkage disequilibrium (LD) (at $r^2 >$
462 0.55) with the most significantly associated SNP. This region is predicted to contain 10
463 genes (Figure 6c).
464 We identified three common haplotypes for this genomic region in the DartMap
465 panel. Accessions that share the haplotype with the Col-0 variant associated with
466 $SNP5^{7856132}$ are referred to as haplotype 1, whereas accessions containing the non-
467 reference variant are further distinguished into two haplotypes that are referred to as
468 haplotypes 2 and 3 (Table S8). Haplotype 1 associates with higher Φ_{PSII} in Fe-deficient
469 conditions relative to the non-reference haplotypes, which do not significantly differ in
470 Φ_{PSII} (Figure 6d). The higher Φ_{PSII} of accessions with haplotype 1 correlates with a
471 significantly higher projected leaf area in Fe-deficient conditions than haplotypes 2 and
472 3 (Figure 6e), while these do not differ significantly for projected leaf area.
473 In total, 40 sequence variants are in LD with $SNP5^{7856132}$ at a cut-off of $r^2 > 0.55$,
474 of which 27 are intergenic, 7 intronic and 6 exonic (Table S9). Only two of the exonic
475 variants encode a non-synonymous amino acid substitution, but in both cases the
476 substitutions are from a leucine to an isoleucine or vice versa. These are very similar
477 amino acids, and neither of these substitutions, one in *RIBF1* (At5g23330), the other in

478 *DUF789* (At5g23380), appear to substantially change the function of these proteins, as
479 predicted by InterPro (Mitchell et al., 2019).
480 We thus hypothesized that non-protein-coding variants that affect either levels
481 or timing of gene expression should be causing the observed QTL. We measured
482 expression of each candidate gene by means of RT-qPCR in both control and Fe-
483 deficient conditions in shoots of six accessions of each haplotype (Table S9). Two known
484 marker genes for Fe deficiency in leaves, *FE SUPEROXIDE DISMUTASE 1* (*FSD1)* and
485 *FERRIC REDUCTION OXIDASE 3* (*FRO3)* (Waters et al., 2012), indeed showed the
486 expected lower (*FSD1*) or higher (*FRO3*) expression under the Fe deficiency treatment
487 compared to control (Figure S8a). Two of the 10 candidate genes, *FE SUPEROXIDE*
488 *DISMUTASE 3* (*FSD3)* and *PRENYLCYSTEINE ALPHA-CARBOXYL METHYLTRANSFERASE*
489 *(STE14A)*, are about 2-fold higher expressed in accessions with haplotype 1 compared to
490 accessions with either of the other two haplotypes (Figure 6f and Figure S8b). The Φ_{PSII}
491 in Fe deficiency correlates well with expression of these genes (Figure 6g). None of the
492 eight other candidate genes showed such response of expression between haplotypes
493 and correlation with the Φ_{PSII} in Fe deficiency (Figure S8b).
494 Since the absolute expression levels of *STE14A* are very low, we focused on *FSD3*
495 as the most likely candidate to underlie the QTL. Two distinct splice variants are
496 described for this gene that influence functionality regarding chloroplast development
497 (Lee et al., 2019). We measured the relative gene expression of each splice variant
498 independently, but found highly similar splice variant ratios between the different
499 haplotypes (Figure S8c), indicating that natural variation affects gene expression of *FSD3*
500 only quantitatively and not through altering differential splicing. Analysis of the non-
501 protein-coding sequences surrounding *FSD3* identified a variable $(TTC)_n/(GAA)_n$
502 microsatellite repeat directly downstream of the *FSD3* 3' UTR, which lacks either two or

503 four of such repeats in haplotypes 2 and 3 relative to haplotype 1. As there are no
504 differences in either Φ_{PSII} or *FSD3* expression between haplotype-2 and-3 accessions we
505 distinguished two alleles for the *FSD3* locus based on their functionality in Fe deficiency:
506 an Fe deficiency tolerant allele, which has the Col-0 reference sequence, and a sensitive
507 allele lacking two or four $(TTC)_n/(GAA)_n$ repeats. We subsequently confirmed the causal
508 nature of the allelic variation at *FSD3* to explain the QTL using quantitative
509 complementation (Weigel, 2012), in which we test if the difference in Φ_{PSII} between F_1
510 progeny derived from crosses between accessions with a tolerant allele crossed with
511 both Col-0 and the *fsd3* mutant, and F_1 progeny derived from similar crosses with
512 accessions carrying the sensitive allele. In Fe-deficient conditions, the F_1 progeny
513 derived from the *fsd3* mutant and accessions with a tolerant *FSD3* allele significantly
514 outperform the F_1 progeny derived from crosses with accessions with a sensitive *FSD3*
515 allele ($P = 0.046$) (Figure 6h). This is not found under control conditions (Figure 6i).

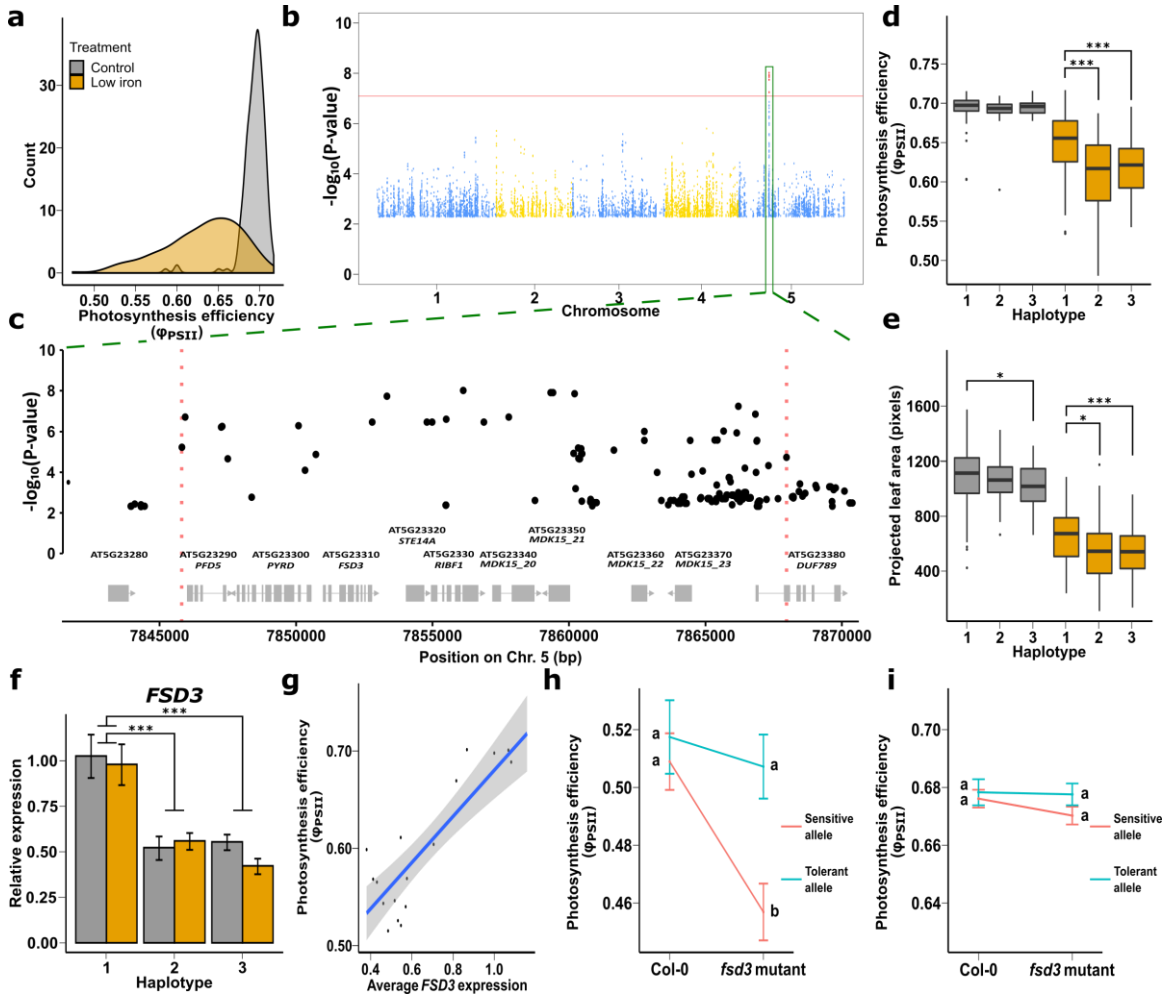


Figure 6: **Genome Wide Association (GWA) analysis of the response of photosynthesis efficiency to iron deficiency.** (a) Distribution of Φ_{PSII}, the light-use efficiency of photosystem II (PSII) electron transport, when grown in control (21 µM Fe^{2+}, grey) and low iron supply (1 µM Fe^{2+}), orange). (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

529 average relative *FSD3* expression level per accession with a linear fit (*FSD3*: $R^2 = 0.74$, P
530 $= 2.94 \times 10^{-6}$). (h-i) Φ_{PSII} of F_1 progeny of crosses between Col-0 (tolerant *FSD3* allele) or
531 the *fsd3* knockout mutant with accessions homozygous for the tolerant allele (N = 5)
532 (blue) and similarly with accessions homozygous for the sensitive allele (N = 9) (red) in
533 iron deficiency (h) and control conditions (i). Error bars represent SE. $*P < 0.05$; $**P <$
534 0.01; $***P < 0.001$, two-tailed Student's *t*-test.
535

536 # Discussion

537 Here we present the DartMap panel, a new biodiversity panel consisting of 192
538 Dutch *A. thaliana* accessions, and we demonstrate its potential for detecting genetic
539 signs of local adaptation. The genetic variation observed in this panel largely falls within
540 the reported diversity of the 1001G collection, which is expected for accessions
541 collected in Northwestern Europe, outside the European relict regions (The 1001
542 Genomes Consortium, 2016). The genetic diversity is unexpectedly high considering the
543 relatively small geographical area without remarkable environmental clines. It illustrates
544 the high level of standing genetic variation that exists among local populations of this
545 predominantly self-pollinating species (Bomblies et al., 2010; The 1001 Genomes
546 Consortium, 2016). Despite, or perhaps due to, the absence of strong environmental
547 clines, the DartMap panel is very suitable for GWAS. The genetic architecture underlying
548 natural adaptation is strongly dependent on the local environment (Brachi et al., 2013;
549 Fournier-Level et al., 2013). Consequently, adaptive loci may exhibit genetic or allelic
550 heterogeneity in regions with contrasting climates, which hampers their detection
551 through GWAS. A milder diversity in environmental conditions can help to alleviate this
552 limitation, as exemplified by the successful identification of the well-known *FRI* gene,
553 which controls flowering time variation. This gene was not previously detected by GWAS

554 in other collections such as the Iberian collection (Tabas-Madrid et al., 2018) or the large
555 1001G panel (The 1001 Genomes Consortium, 2016), primarily due to allelic
556 heterogeneity. As such, the DartMap should be regarded as a valuable complement to
557 collections sampled from regions with contrasting climates, and not a replacement.
558 With the availability of several regional diversity panels, it will be easier to identify such
559 adaptive loci when taking into account that each region may favour its own adaptive
560 variation.

561 Our finding that genes involved in disease resistance are overrepresented among
562 those impacted by CNV is consistent with previous work: a similar phenomenon was
563 found for defence-response genes in a global set of 1,301 natural *A. thaliana* accessions
564 (Göktay et al., 2021) and this same category of genes was found to be enriched in
565 regions highly rearranged between seven diverse *A. thaliana* genomes (Jiao and
566 Schneeberger, 2020). These observations suggest that CNVs may act as a genomic
567 response to biotic stress. The next step would be to investigate whether such structural
568 variants correlate to variation in the prevalence and abundance of known pathogens of
569 *A. thaliana* (Bartoli et al., 2018).

570 The prevalence of plants with short inflorescences and the predominance of a
571 near-isogenic genotype at various dispersed locations, strongly indicate an adaptive
572 advantage for *ga5* mutants in the coastal regions of the Netherlands. Allelic
573 heterogeneity and strong selection of one successful, near-isogenic genotype
574 represented multiple times in the DartMap panel was the reason we could not identify
575 the *GA5* locus through GWAS, but only based on prior knowledge on possible causal loci
576 determining plant semidwarfism (Barboza et al., 2013). Signatures of selection were
577 previously identified in *GA5* among *A. thaliana* accessions collected globally (Barboza et
578 al., 2013), but the environmental cue driving such selection remained elusive. Later (Luo

579 et al., 2015) identified several *ga5* loss-of-function plants in two Alpine populations of *A.*
580 *thaliana* and noted a decrease in plant height with increasing altitude although the
581 adaptive advantage to high altitude environments specifically was not conclusively
582 demonstrated. Our findings demonstrate that high-wind treatment significantly impacts
583 the growth vigour of tall Dutch *A. thaliana* accessions, whereas it has no significant
584 effect on the vigour of predominantly coastally located semidwarf plants. Arguably, we
585 determined this for plants that are growing under optimal water and nutrient supply.
586 Under natural circumstances, with additional growth-limiting factors and probably
587 periods with much stronger winds, this reduction in vigour in the tall accessions is likely
588 to be more detrimental, and also likely to affect their fitness (Younginger et al., 2017) or
589 fecundity, which we could not establish under our artificial windy conditions. The
590 observed phenotypic responses in terms of biomass and stem length to windy
591 conditions are similar between the natural Dutch accessions and the Col-0 and Ler
592 reference strains, as well as their corresponding *ga5* loss-of-function lines. Considering
593 that wind speed generally increases with altitude in mountainous areas
594 (https://map.neweuropeanwindatlas.eu), the selective advantage of the Alpine
595 semidwarfs is also likely to be due to tolerance to wind speed. The selective pressure
596 seems high, given the two independent *ga5* alleles we found in the Netherlands, and
597 others found elsewhere in the world (Barboza et al., 2013; Luo et al., 2015). It is
598 remarkable that the near-isogenic genotype collected at several Dutch sites spread over
599 such large distances, without any obvious means of transport. This is in contrast to a
600 previous case of whole-genome hitchhiking, due to selection on herbicide tolerance,
601 where the spread was clearly linked to transportation along railway tracks (Flood et al.,
602 2016a). However, given the high population density of the Netherlands, we assume that

603 human-mediated movement of soil and seeds would also be sufficient to achieve
604 dispersal of the near-isogenenic *ga5* semidwarfs.

605 The presence of a North-South cline in flowering time and the association
606 between windy conditions and inflorescence height highlight the importance of
607 relatively mild environmental clines for local adaptation. These patterns can be
608 detected in a uniformly and densely sampled population like the DartMap. This is
609 further exemplified by the strong association on chromosome 4 related to the mild cline
610 in seasonality, where the alternative allele of *qDPT-B* is only found at a high local
611 abundance in the West of the Netherlands. However, it should be noted that such
612 associations do not necessarily provide evidence of causality. Demographic processes or
613 genetic drift can produce similar patterns to those created by selection (Rellstab et al.,
614 2015). To further investigate whether this association represents a case of local
615 adaptation, common garden experiments or reciprocal transplanting experiments can
616 be conducted (Hancock et al., 2011; Rellstab et al., 2015; de Villemereuil et al., 2016;
617 Terés et al., 2019). Nonetheless, the high local relative abundance observed in the West
618 of the Netherlands compared to rest of the world, along with the differential sensitivity
619 to drought, suggests the presence of an adaptive locus. However, it seems unlikely that
620 this adaptation is solely attributed to the drought response measured in our
621 experiments, as similar climate characteristics in other regions would also be expected
622 to exhibit a higher relative abundance. Therefore, it is more plausible that a
623 combination of different climatic factors and/or other unconsidered factors such as
624 urbanization contribute to this adaptation.

625 Identifying the causal gene in genotype-environment association analysis can be
626 challenging because allelic variation may affect tolerance to an experimentally
627 measured trait that may not necessarily reflect the trait under natural selection.

628 However, in our study we found that the mutant of *La-RELATED PROTEIN 1c* (*LARP1c*) is
629 more sensitive to drought compared to other tested mutants and Col-0. *LARP1c*
630 encodes a protein with a conserved La-motif domain, which is associated with RNA
631 recognition and is widely found among eukaryotes (Bayfield et al., 2010). Moreover, the
632 sequence variant with the highest statistical association in *qDPT-B* (SNP $4^{17000635}$) is
633 located in the 3' UTR of *LARP1c*. Although neither of these observations provide
634 conclusive evidence on whether this gene is underlying the *qDPT-B*, we consider *LARP1c*
635 as the most plausible candidate gene. In *A. thaliana, LARP1c* is known to be involved in
636 seed germination and the regulation of leaf senescence, which can be induced by
637 various environmental stresses (Zhang et al., 2012; Yan et al., 2023). Additionally, each
638 of the three LARP1 (1a, 1b, 1c) proteins in *A. thaliana* localise at the processing bodies
639 (Zhang et al., 2012; Yan et al., 2023) and are proposed to play a role in RNA processes in
640 response to environmental perturbations. For instance, *LARP1a* is required for
641 thermotolerance of *A. thaliana* to long exposure of moderately high temperatures,
642 where it is involved in regulating the RNA decay machinery together with the
643 cytoplasmic exoribonuclease XRN4 (Merret et al., 2013). However, the specific functions
644 of *LARP1c* in plants are still relatively understudied, and there may be other roles yet to
645 be explored (Yan et al., 2023). *LARP1c* may thus also be involved in the response to
646 drought and possibly also other environmental factors.
647 In addition to the mild and gradual environmental clines, we also examined the
648 panel for sensitivity to low iron availability, as we expected to find variation for this trait
649 given the distinct differences in soil type across the Netherlands. This identified non-
650 coding variation at *FSD3* to cause a major QTL affecting sensitivity to low iron supply.
651 *FSD3* is long known to encode an essential chloroplast-localized iron superoxide
652 dismutase. Its function is to protect developing chloroplasts against excessive damage

653 by superoxide (Myouga et al., 2008). However, natural genetic variation in *FSD3* has not
654 been previously observed. The QTL affecting sensitivity to low iron supply is influenced
655 by subtle allelic variation in *FSD3*, specifically related to the number of repeats in a
656 microsatellite region directly downstream of the transcribed region. This variation
657 results in a two-fold difference in gene expression between contrasting alleles, which
658 strongly correlates with photosynthesis efficiency under iron-deficient conditions. When
659 plants are supplied with sufficient iron, there is no phenotypic difference between
660 contrasting haplotypes. FSD3 is one of two chloroplast-localized iron superoxide
661 dismutases, the other being FSD2 (Gallie and Chen, 2019). While both can form
662 heterocomplexes and *FSD3* overexpression can partly complement the *fsd2* mutant
663 phenotype (Gallie and Chen, 2019), they are not genetically redundant. The *fsd2* loss-of-
664 function mutant displays a compact rosette phenotype, with pale green leaves, and the
665 *fsd3* loss-of-function mutant is barely viable, with a very small and pale rosette (Myouga
666 et al., 2008). Since iron is an essential co-factor for iron superoxide dismutases, a
667 decrease in iron supply is likely to affect their function. Under the mild iron deficiency
668 conditions we tested, this decrease in function was apparently just enough to affect
669 photosynthesis efficiency significantly more in the genotypes with lower *FSD3*
670 expression.
671 We found no reason to suggest strong selection for either of the two alleles.
672 Since the FSD3 function is essential for proper chloroplast development and tolerance
673 to reactive oxygen species generated by photosynthesis, lower *FSD3* expression is likely
674 disadvantageous. The impact of lower *FSD3* expression may even extend beyond
675 chloroplast development and photosynthesis efficiency. For instance, hydrogen
676 peroxide, which is produced when superoxides are converted by superoxide dismutases,
677 can function as a signaling molecule in the systemic response to iron deficiency (Le et

678 al., 2016). Therefore, plants with lower *FSD3* expression may have a delayed response in
679 detecting iron deficiency. Nevertheless, a large ionome survey of 1135 global *A. thaliana*
680 accessions did not reveal any clear indications of local adaptation to iron deficiency
681 (Campos et al., 2021). This suggests that establishing a direct link between allelic
682 variation and local adaptation to iron deficiency may be challenging. Factors such
683 seasonal variation in iron availability depending on precipitation or limited distribution
684 and heterogeneity of the allelic variation could contribute to the difficulty in establishing
685 such a link.

686 In conclusion, our work demonstrates that regions with relatively mild
687 environmental clines can still harbor a wide range of adaptive genetic variants. This
688 highlights the value of establishing regional collections as excellent tools for studying
689 local adaptation of *A. thaliana*. Such a collection does not need to be limited to regions
690 with strong environmental clines as were put forward in previous studies (Brachi et al.,
691 2013; Frachon et al., 2018). Even a diversity panel from a relatively small geographic
692 region with mild environmental clines, such as the Netherlands, provides enough
693 variation to pick up small differences in flowering time and sensitivity to iron deficiency.
694 Given the decreasing costs of whole genome sequencing, it becomes increasingly easier
695 to characterize the genetic variation in such populations in great detail. Our study may
696 serve as a blueprint of this approach, supporting the view that collections of wild plant
697 species can be powerful systems to determine the genomic basis of natural adaptation
698 (Weigel and Nordborg, 2015; Sollars et al., 2017; Monnahan et al., 2019).

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 Dutch *Arabidopsis thaliana* Map (DartMap) panel, suitable for GWAS (Table S2). 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 $^{\circ}$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 $^{\circ}$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 $^{\circ}$C.

722 ## Genomic DNA sequence data

723 Genomic DNA of the DartMap panel was sequenced at GenomeScan B.V., Leiden.
724 Libraries were prepared using the NEBNext Ultra DNA Library Prep kit for Illumina,
725 according to described procedures. 500-700 bp insert size fragments were sequenced
726 using a Illumina Hiseq 4000 device, aiming for at least 30x coverage of paired-end reads
727 (2x151 bp) of each accession, and used to generate the alignment files necessary for
728 calling genomic variants (see Note S1 for details). We additionally generated alignments
729 for the DartMap panel combined with 11 Dutch accessions included in the 1,001
730 Genomes (1001G) Project (The 1001 Genomes Consortium, 2016) (Sequence Read
731 Archive ID: SRP056687). This combined set of samples is referred to as the "DartMap +
732 1001G panel", to clearly distinguish it from the DartMap panel containing accessions
733 collected in this study only.
734

735 ## Single nucleotide polymorphism (SNP) and indel calling

736 Genetic variants were called using a workflow based on the Genome Analysis Toolkit
737 (GATK) Best Practices (Van der Auwera et al., 2013). Base quality scores of aligned reads
738 were recalibrated using GATK (version 4.0.2.1) BaseRecalibrator with default
739 parameters, using a set of variants called in a world-wide panel of 1135 *A. thaliana*
740 accessions as known sites (The 1001 Genomes Consortium, 2016) (obtained from
741 ftp.ensemblgenomes.org/pub/plants/release-37/vcf/arabidopsis_thaliana). SNPs and
742 short insertions/deletions (indels) were called in each sample using GATK
743 HaplotypeCaller, allowing a maximum of three alternate alleles at each site. Samples
744 were jointly genotyped using GATK GenomicsDBImport and GATK GenotypeGVCFs with
745 default parameters. This last step generated three different VCF files containing calls for

746 the nuclear genome, the mitochondrial genome, and the chloroplast genome
747 respectively. We filtered these sets to remove putative false positive calls (see Note S1
748 for details).
749

750 Copy number variation (CNV) calling

751 We called CNVs in the DartMap panel, defined here as deletions, duplications, and
752 insertions of at least 50 bp, by applying the Hecaton workflow (Wijfjes et al., 2019) to
753 the alignment files of each accession. Accessions of the 1001G were excluded from this
754 analysis, as they were sequenced using a different sequencing platform, which can
755 strongly affect CNV calling: a linear model of genetic distance between accessions based
756 on CNVs as a function of differences in coverage, sequencing technology, insert size, and
757 read length (see Note S1 for details) explained 30% of the variation in the number of
758 CNVs found in each sample (R^2 = 0.3). CNVs called in each accession were genotyped
759 and merged to generate a single VCF file as output (see Note S1 for details).
760

761 Characterizing the impact of deletions

762 We detected deletions overlapping protein-coding genes (by at least 1 bp) using
763 bedtools intersect. Genes were obtained from the TAIR10 genomic annotation of
764 Ensembl Plants (release 40) and from the annotation of the improved *A. thaliana*
765 mitochondrial assembly (BK010421.1). Gene function was considered disrupted if the
766 associated gene overlapped with a deletion and at least 76 bp (more than half the
767 length of a read) of the coding sequence lacked read coverage. Deletions were defined
768 as having a moderate frequency if they were present in at least 48 samples (25%) of the
769 DartMap panel. We defined disease resistance genes by first collecting all Uniprot

770 entries linked to all *A. thaliana* reference genes using the Retrieve/ID mapping tool
771 (https://www.uniprot.org/uploadlists/), excluding a small number of genes (53) for
772 which no corresponding entries could be found. Genes were labeled as involved in
773 disease resistance if the name of their protein product contained the terms "disease
774 resistance" or "Disease resistance". Enrichment tests were performed using Fisher's
775 exact test as implemented in the Python SciPy library (Virtanen et al., 2020) (version
776 1.3.1).
777

778 Population structure analysis

779 We clustered samples based on bi-allelic sites missing calls in fewer than 10% of the
780 samples and having a non-reference allele in at least 10 of them, to ensure that
781 accessions are grouped on variation at the population level, rather than rare variants
782 found in few individuals. For clusters based on mitochondrial and chloroplast variants,
783 this criterium was relaxed to a minimum of two samples, as a relatively large number of
784 sites would have been discarded otherwise. To generate a distance metric, we
785 represented samples by binary vectors with 1s for homozygous variants and 0 for all
786 other cases and computed the Hamming distance between them, i.e. the number of
787 variant sites differing between samples, using the dist.gene function of the R package
788 ape (Paradis and Schliep, 2019) (v5.3). Samples were clustered with the hclust function
789 in R (R Core Team, 2013), using complete linkage as the agglomeration method, similar
790 to analyses on the Swedish and Iberian collection (Long et al., 2013; Tabas-Madrid et al.,
791 2018) (Figure S9). We cut trees into different numbers of groups using the R function
792 cutree, to explore the geographical distribution of clusters formed at various
793 hierarchical levels. Geographical distance between accessions was computed using the
794 distance function of the geopy library (https://github.com/geopy/geopy) (v2.0.0) with

795 default parameters and their geographical locations were plotted using the basemap
796 library (https://matplotlib.org/basemap/) (v1.2.2).
797

798 Principal component analysis

799 Principal component analysis of chloroplast variants called in the Dutch and global
800 collection was performed using the SeqArray (Zheng et al., 2017) (v1.28.0) and
801 SNPrelate (Zheng et al., 2012) (v1.22.0) libraries in R. To remove putative false positive
802 calls, variants were excluded if they had a quality-by-depth (QD) score lower than 25,
803 leaving a total of 4095 sites. Furthermore, we removed 10 accessions in which at least
804 50% of the variants were called as heterozygous, suggesting DNA contamination of the
805 samples used for sequencing. Results were visualized using ggplot2 (Wickham, 2016)
806 (v3.3.2). Three outlier accessions were excluded from the final figures to improve the
807 visualization of Dutch accessions with respect to the global collection. Accessions were
808 coloured based on their chloroplast haplogroup assigned in a previous study (Hsu et al.,
809 2019). Haplogroups of accessions that were not included in that paper (i.e. the DartMap
810 panel) were predicted with a 1-nearest neighbour classifier on the accessions that were,
811 using the first two principal components as features.
812

813 Phenotyping flowering time and stem length

814 For all experiments, seeds were sown on wet filter paper and stratified for 3 days at 4
815 °C before they were germinated. Seedlings were grown hydroponically on rockwool
816 blocks (Grodan Rockwool Group, 40 x 40 mm in size) pre-soaked in a nutrient solution
817 designed for *A. thaliana* (1.7mm NH_4^+, 4.5mm K^+, 0.4mm Na^+, 2.3mm Ca^{2+}, 1.5mm
818 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

819 SiO_3^{2-}, 21 µM Fe^{2+} (chelated with 3% diethylene triaminopentaacetic acid), 3.4 µM
820 Mn^{2+}, 4.7 µM Zn^{2+}, 14 µM BO_3^{3-} and 6.9 µM Cu^{2+} at pH 7, EC 1.4 mS cm^{-1}). Plants were
821 grown in a greenhouse until 97 days after sowing. For each treatment, plants were
822 equally divided over 20 plastic trays, with 76-78 plants per tray. Eight replicates per
823 genotype per treatment were used, randomly assigned over four separate trays with
824 two replicates per tray per genotype. Each tray consisted of a unique, randomly
825 generated combination of genotypes. Trays were placed in a random order in the
826 greenhouse, and were moved once every two weeks to a new position to minimize
827 possible spatial effects in the greenhouse. A vernalisation treatment of six weeks was
828 performed in a climate-controlled growth chamber set at 4 °C, with 12h/12h light/dark
829 period, 70 µM m^{-2} s^{-1} irradiance and 70% humidity starting at day 20 after sowing. After
830 vernalisation, plants were returned to the greenhouse. Flowering time and height up to
831 the first silique were scored in the greenhouse, either with or without a vernalisation
832 period of six weeks. Flowering time was determined as the day when the first flower
833 had opened. Stem length up to the first silique was measured one week after the first
834 flower had opened, as this trait remains stable thereafter (Barboza et al., 2013). To
835 obtain the set of phenotypes for further analysis, we calculated the average phenotypic
836 values per accession per treatment.
837

838 ## Wind simulation treatment

839 In a separate experiment, plants of six semidwarf accessions and five tall accessions of
840 the DartMap panel, with similar flowering time, as well as the accessions Col-0 and Ler,
841 and their respective loss-of-function *ga5* mutants (Table S4) were grown for the first 14
842 days after sowing in a climate-controlled growth chamber, set at 20/18 °C (day/night)
843 with 10 h light (from 08:00 - 18:00) and 70% humidity at an irradiance of 200 µM m^{-2} s^{-1}

844 at ambient CO_2 levels. The growth chamber contained two treatment basins. The
845 treatment basins have the same design as described in a previous study (Flood et al.,
846 2016b), with the distinction that the two treatment basins are physically separated and
847 are on either side of the climate chamber. Each treatment basin was divided into five
848 equally sized blocks to hold plants. Accessions were grown in a randomized complete
849 block design with six replicates per block. Wind was simulated with two fans that were
850 positioned at either end of one of the treatment basins. The fans were set to provide a
851 wind speed that varied around 5.6 m/s for the majority of the treatment basin. The
852 wind speed varied depending on the distance of the plant to the fan, with plants closest
853 to the fan experiencing wind speeds up to 7 m/s and plants furthest away of the fan and
854 at the edge of the treatment basin experiencing 3.5 m/s wind. The wind speed for the
855 control treatment basin varied between 0.5 and 1.0 m/s. Wind speeds were measured
856 with a TSI Velocalc (model: 8347-M-GB) air velocity monitor. Wind treatment was
857 simulated by alternating patterns of six hours with wind by six hours without wind and
858 was terminated when the first plants in the wind treatment basin had fully ripened
859 siliques.
860

861 ## Extraction of climatic variables

862 Climatic variables were obtained from google earth engine using R and RGEE (Aybar et
863 al., 2020). All bands from Worldclim BIO data (Hijmans et al., 2005) selected from 250m-
864 by-250m squares using the coordinates from the collection site for each accession as
865 input.
866

867 ## Drought treatment

868 To test which of the two nearby QTLs at *qDPT* affected drought tolerance, plants were
869 grown in a greenhouse on ceramic-based granular soil. For the first two weeks, all plants
870 were watered three times a week with the nutrient solution described previously. This
871 watering regime was continued for the control treatment while for the drought
872 treatment excess nutrient solution was removed and watering was ceased. Accessions
873 (Table S6) were grown in a randomised complete block design with five replicates per
874 treatment. Photographs were taken with a Nikon D3000 with the same settings, zoom
875 and distance to the plants and were subsequently analysed to obtain the projected leaf
876 area (PLA) with ImageJ (settings Hue 26-114, Saturation 58-218 and Brightness 121-
877 255). The PLA was then normalised per accession by dividing the PLA in the drought
878 treatment by the PLA in the control treatment to correct for innate differences in size of
879 different accessions. For candidate gene validation T-DNA insertion lines were used
880 (Table S7). Plants were grown in our phenotyping growth chamber, the Phenovator
881 (Flood et al., 2016b), with the same conditions in the growth chamber as described for
882 the wind simulation treatment. The drought treatment was induced in a similar way, by
883 using ceramic-based granular soil. In contrast to the experiment in the greenhouse, the
884 growth substrate was covered with black plastic cover to prevent algal growth that
885 might interfere with high throughput phenotyping. Moreover, we watered plants in the
886 drought treatment only once at the start, as there was less evaporation in this setup
887 compared to the greenhouse. Starting at 10 days after sowing, we measured projected
888 leaf area based on near infrared reflectance (NIR) imaging as described by (Flood et al.,
889 2016b) four times per day, and F_v/F_m using chlorophyll fluorescence once per day.
890

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 1 µM Fe^{2+} instead of 21 µM 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., 2016b), 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., 2016b), 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., 2015). 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

916 the entire population. We calculated the residuals of each accession relative to this
917 linear regression to be used as the phenotypic value for GWA analysis. Additionally, we
918 used the average Φ_{PSII} values in the iron deficient condition and the calculated ratio
919 between Φ_{PSII} in iron deficiency versus control as alternative traits for GWAS.
920

921 ## Plant material for expression analysis

922 Plants that were used for gene expression analysis were grown in the greenhouse on
923 rockwool blocks that were watered with either the control nutrient solution (21 µM
924 Fe^{2+}) or the adjusted low iron solution (1 µM Fe^{2+}). For both treatments we used five
925 plastic trays to contain the nutrient solution. Two replicate plants per accession were
926 distributed over trays in a complete randomised block design. Whole shoots of two
927 replicate plants per accession originating from the same tray were pooled and
928 harvested as one sample for RNA isolation 21 days after germination, while five such
929 samples per accession per treatment were harvested in total. The material was
930 snapfrozen in liquid nitrogen immediately after harvesting and stored at -80 °C until
931 RNA isolation. Accessions were selected for gene expression analysis (Table S8) based
932 on their haplotype for the region with the highest association in the GWAS. Haplotypes
933 were determined by looking at pairwise correlations between all variants with a minor
934 allele frequency > 10% in a window of 20 kb surrounding the SNP with the highest
935 association.
936

937 ## Quantitative complementation

938 Quantitative complementation was used to test if the difference in Φ_{PSII} between the
939 F_1's of Columbia-0 (Col-0) and the knockout mutant *fsd3* in Col-0 background crossed

940 with accessions homozygous for the tolerant allele was significantly different from the
941 response in Φ_{PSII} for F_1's of Col-0 and *fsd3* crossed with accessions homozygous for the
942 sensitive allele. Five accessions with a tolerant, and nine accessions with a sensitive
943 allele for *FSD3* were crossed to both Col-0 and *fsd3* (Table S8). Col-0 and *fsd3* were used
944 as pollen donor in all crosses. The *fsd3* T-DNA insertion knockout mutant, SALK_103228,
945 was ordered from the Arabidopsis Biological Resource Center and insertion of the T-
946 DNA was validated by PCR as earlier described (Myouga et al., 2008). This T-DNA
947 insertion line has no *FSD3* transcript (Myouga et al., 2008), and we could only grow it
948 beyond the seedling stage on a 0.3% (w/v) gel-rite medium supplemented with 3% (w/v)
949 sucrose.

950 Plants (N = 10 per cross per treatment) were grown as described in the section
951 "Plant material for expression analysis". Plants were phenotyped for Φ_{PSII} from 21 DAS
952 onwards as described by (Baker and Oxborough, 2004) with the PlantScreenTM SC
953 System (Photosystems Instruments). Phenotyping was done for twenty plants per run
954 and lasted for three consecutive days overall. We phenotyped crosses with the tolerant
955 allele and the sensitive allele equally spread out over the three days to prevent possible
956 phenotypic differences caused by the delay between measurements. Moreover, in each
957 measurement of twenty plants we included F_1's from both the cross between a given
958 accession with Col-0 and with *fsd3* in a single measurement. We included one reference
959 tray that was measured each day in the morning, at the start of the afternoon and at
960 the end of the afternoon to test for possible differences in Φ_{PSII} during the day, but no
961 significant differences were found. We corrected the phenotypic data for possible block
962 effects by fitting a linear model, with random terms underlined:

963

964 $\underline{Y} = (FSD3\ allele\ group \times Pollen\ donor) + \underline{Block} + \underline{\varepsilon}$

965

966 The model was analysed using the restricted maximum likelihood procedure with the
967 lme4 package in R (Bates et al., 2015). We tested interaction between the *FSD3* allele
968 group (natural tolerant or sensitive allele) and the pollen donor (Col-0 or the *fsd3*
969 mutant) with a two-way ANOVA, for which a Kenward-Roger approximation for the
970 degrees of freedom was used. Tukey's post-hoc multiple comparison test ($\alpha = 0.05$) was
971 used to assess significance.

972

973 RNA isolation and gene expression analysis

974 Total RNA was extracted with the Direct-zol RNA mini-prep kit (Zymo Research)
975 according to the company's instructions and then subjected to a DNAse (Promega)
976 treatment at 37 °C for one hour. Total removal of DNA was validated by means of a no-
977 reverse-transcriptase PCR reaction on a few mixtures of samples. The RNA quality was
978 assessed for purity (A260/A280) with a spectrophotometer (Thermo scientific Nanodrop
979 2000) and for possible RNA degradation by means of a visual inspection of the RNA on a
980 1% (w/v) agarose gel. cDNA was then synthesized from 1 µg total RNA (measured by
981 spectrophotometer) with the iScript™ cDNA synthesis kit (Bio-RAD) according to the
982 company's instructions. Reverse transcriptase quantitative PCR (RT-qPCR) reactions
983 were performed with the SensiFast SYBR Green Mastermix (Bioline) on a Bio-RAD cfx96
984 machine. RT-qPCR primer sets were designed in parts of the coding region that did not
985 contain SNPs between the used set of accessions. For each RT-qPCR primer set used, we
986 first validated its efficiency by means of a standard curve and only used primer sets with
987 efficiencies ranging between 90% and 110%. Gene expression was normalized, using the
988 delta cycling threshold (ΔCt) method (Livak and Schmittgen, 2001), to three reference
989 genes, *SAND*¸ *YLS8*, and *TIP41-like* as earlier described (Han et al., 2013). Differential

990 expression was assessed by two-sample t-tests on the ΔCt values. All primer sequences
991 can be found in Table S10. We included *FSD1* and *FRO3* in the expression analysis as
992 marker genes for iron deficiency in shoots (Waters et al., 2012).
993

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

996 GWA and GEA mapping were performed in GEMMA (Zhou and Stephens, 2012) (version
997 0.98) with a univariate linear mixed model and the minor allele frequency cut-off set at
998 0.05. A kinship matrix was constructed in GEMMA to correct for population structure.
999 CNVs were included in the analysis in a similar fashion as SNPs and indels by converting
1000 them into a reference variant, an alternative variant or heterozygous. To estimate a
1001 suitable genome-wide p-value threshold, we performed 1,000 permutations of the
1002 phenotype values. For each permutation we performed association analysis and
1003 extracted the highest association score. The distribution of these 1,000 highest
1004 association scores was then used to determine the empirical threshold value of $p <$
1005 0.05.
1006

1007 Data availability

1008 Raw sequencing data of the DartMap panel can be found on the repository of the
1009 National Center for Biotechnology Information (NCBI) (BioProject ID: PRJNA727738).

1010 Funding

1011 This work was supported by the project grant ALWGR.2015.9 of the Netherlands
1012 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.

Authors' contributions

SS, DDR and MGMA designed the project with RB and RYW. RB composed the panel. 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 contributed especially 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 contributed to the experimental design, interpretation of results, and preparation of the manuscript. All authors read and approved the final manuscript.

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.

Competing interests

The authors declare that they have no competing interests.

References

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

Arteaga, N., Méndez-Vigo, B., Fuster-Pons, A., Savic, M., Murillo-Sánchez, A., Picó, F. X., and Alonso-Blanco, C. (2022). Differential environmental and genomic architectures shape the natural diversity for trichome patterning and morphology in different Arabidopsis organs. *Plant, Cell & Environment* **45**:3018–3035.

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

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

Baker, N. R., and Oxborough, K. (2004). Chlorophyll Fluorescence as a Probe of Photosynthetic Productivity. In *Chlorophyll a Fluorescence: A Signature of Photosynthesis* (ed. Papageorgiou, G. C.) and Govindjee, pp. 65–82. Dordrecht: Springer Netherlands.

Barboza, L., Effgen, S., Alonso-Blanco, C., Kooke, R., Keurentjes, J. J. B., Koornneef, M., and 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**:15818–15823.

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

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* **67**:1–48.

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

Bayfield, M. A., Yang, R., and 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**:365–378.

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

Bomblies, K., Yant, L., Laitinen, R. A., Kim, S.-T., Hollister, J. D., Warthmann, N., Fitz, J., and Weigel, D. (2010). Local-scale patterns of genetic variability, outcrossing, and

1077 spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics*
1078 **6**:e1000890.
1079 **Brachi, B., Villoutreix, R., Faure, N., Hautekèete, N., Piquot, Y., Pauwels, M., Roby, D.,**
1080 **Cuguen, J., Bergelson, J., and Roux, F.** (2013). Investigation of the geographical
1081 scale of adaptive phenological variation and its underlying genetics in *Arabidopsis*
1082 *thaliana*. *Molecular Ecology* **22**:4222–4240.
1083 **Cameron, D. L., Di Stefano, L., and Papenfuss, A. T.** (2019). Comprehensive evaluation
1084 and characterisation of short read general-purpose structural variant calling
1085 software. *Nature Communications* **10**:3240.
1086 **Campos, A. C. A. L., van Dijk, W. F. A., Ramakrishna, P., Giles, T., Korte, P., Douglas, A.,**
1087 **Smith, P., and Salt, D. E.** (2021). 1,135 ionomes reveal the global pattern of leaf
1088 and seed mineral nutrient and trace element diversity in *Arabidopsis thaliana*. *The*
1089 *Plant Journal* **106**:536–554.
1090 **de Villemereuil, P., Gaggiotti, O. E., Mouterde, M., and Till-Bottraud, I.** (2016). Common
1091 garden experiments in the genomic era: new perspectives and opportunities.
1092 *Heredity* **116**:249–254.
1093 **De Vries, F., De Groot, W. J. M., Hoogland, T., and Denneboom, J.** (2003). *De*
1094 *Bodemkaart van Nederland digitaal; toelichting bij inhoud, actualiteit en*
1095 *methodiek en korte beschrijving van additionele informatie*. Wageningen: Alterra.
1096 **Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C.,**
1097 **Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., et al.** (2017).
1098 African genomes illuminate the early history and transition to selfing in
1099 *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **114**:5213–
1100 5218.
1101 **Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R., and Weigel, D.** (2019).
1102 Natural selection on the *Arabidopsis thaliana* genome in present and future
1103 climates. *Nature* **573**:126–129.
1104 **Ferrero-Serrano, Á., and Assmann, S. M.** (2019). Phenotypic and genome-wide
1105 association with the local environment of *Arabidopsis*. *Nature Ecology & Evolution*
1106 **3**:274–285.
1107 **Flood, P. J., Van Heerwaarden, J., Becker, F., De Snoo, C. B., Harbinson, J., and Aarts, M.**
1108 **G. M.** (2016a). Whole-genome hitchhiking on an organelle mutation. *Current*
1109 *Biology* **26**:1306–1311.
1110 **Flood, P. J., Kruijer, W., Schnabel, S. K., van der Schoor, R., Jalink, H., Snel, J. F. H.,**
1111 **Harbinson, J., and Aarts, M. G. M.** (2016b). Phenomics for photosynthesis, growth
1112 and reflectance in *Arabidopsis thaliana* reveals circadian and long-term
1113 fluctuations in heritability. *Plant Methods* **12**:1–14.
1114 **Fournier-Level, A., Wilczek, A. M., Cooper, M. D., Roe, J. L., Anderson, J., Eaton, D.,**
1115 **Moyers, B. T., Petipas, R. H., Schaeffer, R. N., and Pieper, B.** (2013). Paths to
1116 selection on life history loci in different natural environments across the native
1117 range of *Arabidopsis thaliana*. *Molecular Ecology* **22**:3552–3566.
1118 **Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., Roby, D., and**
1119 **Roux, F.** (2018). A genomic map of climate adaptation in *Arabidopsis thaliana* at a
1120 micro-geographic scale. *Frontiers in Plant Science* **9**:967.
1121 **Fulgione, A., Koornneef, M., Roux, F., Hermisson, J., and Hancock, A. M.** (2018).
1122 Madeiran *Arabidopsis thaliana* Reveals Ancient Long-Range Colonization and
1123 Clarifies Demography in Eurasia. *Molecular Biology and Evolution* **35**:564–574.
1124 **Gallie, D. R., and Chen, Z.** (2019). Chloroplast-localized iron superoxide dismutases FSD2
1125 and FSD3 are functionally distinct in *Arabidopsis*. *PLoS One* **14**:e0220078.
1126 **Göktay, M., Fulgione, A., and Hancock, A. M.** (2021). A new catalog of structural variants

1127 in 1,301 *A. thaliana* lines from Africa, Eurasia, and North America reveals a
1128 signature of balancing selection at defense response genes. *Molecular Biology and*
1129 *Evolution* **38**:1498–1511.
1130 **Goslee, S. C., and Urban, D. L.** (2007). The ecodist Package for Dissimilarity-based
1131 Analysis of Ecological Data. *J. Stat. Soft.* **22**:1–19.
1132 **Grace, J.** (1988). 3. Plant response to wind. *Agriculture, Ecosystems & Environment*
1133 **22**:71–88.
1134 **Han, B., Yang, Z., Samma, M. K., Wang, R., and Shen, W.** (2013). Systematic validation of
1135 candidate reference genes for qRT-PCR normalization under iron deficiency in
1136 Arabidopsis. *Biometals* **26**:403–413.
1137 **Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G.,**
1138 **Toomajian, C., Roux, F., and Bergelson, J.** (2011). Adaptation to climate across
1139 the *Arabidopsis thaliana* genome. *Science* **334**:83–86.
1140 **Hänsch, R., and Mendel, R. R.** (2009). Physiological functions of mineral micronutrients
1141 (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). *Current Opinion in Plant Biology* **12**:259–266.
1142 **Hartemink, A. E., and Sonneveld, M. P. W.** (2013). Soil maps of the Netherlands.
1143 *Geoderma* **204**:1–9.
1144 **Hell, R., and Stephan, U. W.** (2003). Iron uptake, trafficking and homeostasis in plants.
1145 *Planta* **216**:541–551.
1146 **Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.** (2005). Very high
1147 resolution interpolated climate surfaces for global land areas. *International*
1148 *Journal of Climatology* **25**:1965–1978.
1149 **Hsu, C., Lo, C., and Lee, C.** (2019). On the postglacial spread of human commensal
1150 *Arabidopsis thaliana*: journey to the East. *New Phytologist* **222**:1447–1457.
1151 **Huang, X., and Han, B.** (2014). Natural variations and genome-wide association studies in
1152 crop plants. *Annual Review of Plant Biology* **65**:531–551.
1153 **Jiao, W.-B., and Schneeberger, K.** (2020). Chromosome-level assemblies of multiple
1154 *Arabidopsis* genomes reveal hotspots of rearrangements with altered
1155 evolutionary dynamics. *Nature Communications* **11**:989.
1156 **Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C.** (2000).
1157 Molecular analysis of FRIGIDA, a major determinant of natural variation in
1158 *Arabidopsis* flowering time. *Science* **290**:344–347.
1159 **Kawecki, T. J., and Ebert, D.** (2004). Conceptual issues in local adaptation. *Ecology Letters*
1160 **7**:1225–1241.
1161 **Koornneef, M., Blankestijn-de Vries, H., Hanhart, C., Soppe, W., and Peeters, T.** (1994).
1162 The phenotype of some late-flowering mutants is enhanced by a locus on
1163 chromosome 5 that is not effective in the Landsberg *erecta* wild-type. *The Plant*
1164 *Journal* **6**:911–919.
1165 **Korte, A., and Farlow, A.** (2013). The advantages and limitations of trait analysis with
1166 GWAS: a review. *Plant Methods* **9**:1–9.
1167 **Le, C. T. T., Brumbarova, T., Ivanov, R., Stoof, C., Weber, E., Mohrbacher, J., Fink-**
1168 **Straube, C., and Bauer, P.** (2016). ZINC FINGER OF ARABIDOPSIS THALIANA12
1169 (ZAT12) interacts with FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION
1170 FACTOR (FIT) linking iron deficiency and oxidative stress responses. *Plant*
1171 *Physiology* **170**:540–557.
1172 **Le Corre, V., Roux, F., and Reboud, X.** (2002). DNA polymorphism at the *FRIGIDA* gene in
1173 *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local
1174 selection for flowering time. *Molecular Biology and Evolution* **19**:1261–1271.
1175 **Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-**
1176 **Blanco, C., Weigel, D., and Nordborg, M.** (2017). On the post-glacial spread of

1177 human commensal *Arabidopsis thaliana*. *Nature Communications* **8**:14458.
1178 **Lee, S., Joung, Y. H., Kim, J.-K., Do Choi, Y., and Jang, G.** (2019). An isoform of the plastid
1179 RNA polymerase-associated protein FSD3 negatively regulates chloroplast
1180 development. *BMC Plant Biology* **19**:1–12.
1181 **Li, H.** (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-
1182 MEM. *arXiV* Advance Access published 2013.
1183 **Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,**
1184 **G., Durbin, R., and 1000 Genome Project Data Processing Subgroup** (2009). The
1185 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079.
1186 **Li, Y., Huang, Y., Bergelson, J., Nordborg, M., and Borevitz, J. O.** (2010). Association
1187 mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*.
1188 *Proceedings of the National Academy of Sciences* **107**:21199–21204.
1189 **Livak, K. J., and Schmittgen, T. D.** (2001). Analysis of relative gene expression data using
1190 real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**:402–408.
1191 **Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., Zhang, Q.,**
1192 **Vilhjálmsson, B. J., Korte, A., and Nizhynska, V.** (2013). Massive genomic
1193 variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature*
1194 *Genetics* **45**:884–890.
1195 **Luo, Y., Dong, X., Yu, T., Shi, X., Li, Z., Yang, W., Widmer, A., and Karrenberg, S.** (2015). A
1196 single nucleotide deletion in *gibberellin20-oxidase1* causes alpine dwarfism in
1197 Arabidopsis. *Plant Physiology* **168**:930–937.
1198 **Martin, M.** (2011). Cutadapt removes adapter sequences from high-throughput
1199 sequencing reads. *EMBnet.journal* **17**:10–12.
1200 **Maxwell, K., and Johnson, G. N.** (2000). Chlorophyll fluorescence—a practical guide.
1201 *Journal of Experimental Botany* **51**:659–668.
1202 **McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A.,**
1203 **Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al.** (2010). The Genome
1204 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
1205 sequencing data. *Genome Research* **20**:1297–1303.
1206 **Merret, R., Descombin, J., Juan, Y., Favory, J.-J., Carpentier, M.-C., Chaparro, C., Charng,**
1207 **Y., Deragon, J.-M., and Bousquet-Antonelli, C.** (2013). XRN4 and LARP1 are
1208 required for a heat-triggered mRNA decay pathway involved in plant acclimation
1209 and survival during thermal stress. *Cell Reports* **5**:1279–1293.
1210 **Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D.,**
1211 **Chang, H.-Y., El-Gebali, S., and Fraser, M. I.** (2019). InterPro in 2019: improving
1212 coverage, classification and access to protein sequence annotations. *Nucleic Acids*
1213 *Research* **47**:D351–D360.
1214 **Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., Laenen, B., Schmickl,**
1215 **R., Paajanen, P., and Šrámková, G.** (2019). Pervasive population genomic
1216 consequences of genome duplication in *Arabidopsis arenosa*. *Nature Ecology &*
1217 *Evolution* **3**:457–468.
1218 **Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y.,**
1219 **Nagata, N., Ikeuchi, M., and Shinozaki, K.** (2008). A heterocomplex of iron
1220 superoxide dismutases defends chloroplast nucleoids against oxidative stress and
1221 is essential for chloroplast development in Arabidopsis. *The Plant Cell* **20**:3148–
1222 3162.
1223 **Paradis, E., and Schliep, K.** (2019). ape 5.0: an environment for modern phylogenetics
1224 and evolutionary analyses in R. *Bioinformatics* **35**:526–528.
1225 **Pedersen, B. S., and Quinlan, A. R.** (2019). Duphold: scalable, depth-based annotation
1226 and curation of high-confidence structural variant calls. *GigaScience* **8**:giz040.

1227 **Pico, F. X., Méndez-Vigo, B., Martínez-Zapater, J. M., and Alonso-Blanco, C.** (2008).
1228 Natural genetic variation of *Arabidopsis thaliana* is geographically structured in
1229 the Iberian Peninsula. *Genetics* **180**:1009–1021.
1230 **R Core Team, R.** (2013). R: A language and environment for statistical computing Advance
1231 Access published 2013.
1232 **Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., and Holderegger, R.** (2015). A
1233 practical guide to environmental association analysis in landscape genomics.
1234 *Molecular Ecology* **24**:4348–4370.
1235 **Schneeberger, K., Hagmann, J., Ossowski, S., Warthmann, N., Gesing, S., Kohlbacher, O.,**
1236 **and Weigel, D.** (2009). Simultaneous alignment of short reads against multiple
1237 genomes. *Genome Biology* **10**:R98.
1238 **Sloan, D. B., Wu, Z., and Sharbrough, J.** (2018). Correction of Persistent Errors in
1239 Arabidopsis Reference Mitochondrial Genomes. *The Plant Cell* **30**:525–527.
1240 **Sobel, J. M., Chen, G. F., Watt, L. R., and Schemske, D. W.** (2010). The biology of
1241 speciation. *Evolution* **64**:295–315.
1242 **Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H.,**
1243 **Swarbreck, D., Kaithakottil, G., Cooper, E. D., Uauy, C., and Havlickova, L.** (2017).
1244 Genome sequence and genetic diversity of European ash trees. *Nature* **541**:212–
1245 216.
1246 **Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S.,**
1247 **Purugganan, M. D., and Schmitt, J.** (2004). A latitudinal cline in flowering time in
1248 *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings*
1249 *of the National Academy of Sciences* **101**:4712–4717.
1250 **Tabas-Madrid, D., Méndez-Vigo, B., Arteaga, N., Marcer, A., Pascual-Montano, A.,**
1251 **Weigel, D., Xavier Pico, F., and Alonso-Blanco, C.** (2018). Genome-wide
1252 signatures of flowering adaptation to climate temperature: Regional analyses in a
1253 highly diverse native range of *Arabidopsis thaliana*. *Plant, Cell & Environment*
1254 **41**:1806–1820.
1255 **Terés, J., Busoms, S., Perez Martín, L., Luís-Villarroya, A., Flis, P., Álvarez-Fernández, A.,**
1256 **Tolrà, R., Salt, D. E., and Poschenrieder, C.** (2019). Soil carbonate drives local
1257 adaptation in *Arabidopsis thaliana*. *Plant, Cell & Environment* **42**:2384–2398.
1258 **Terry, N.** (1980). Limiting factors in photosynthesis: I. Use of iron stress to control
1259 photochemical capacity in vivo. *Plant Physiology* **65**:114–120.
1260 **The 1001 Genomes Consortium** (2016). 1,135 genomes reveal the global pattern of
1261 polymorphism in *Arabidopsis thaliana*. *Cell* **166**:481–491.
1262 **Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-**
1263 **Moonshine, A., Jordan, T., Shakir, K., Roazen, D., and Thibault, J.** (2013). From
1264 FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best
1265 practices pipeline. *Current Protocols in Bioinformatics* **43**:10–11.
1266 **Veer, G. van der** (2006). *Geochemical soil survey of the Netherlands. Atlas of major and*
1267 *trace elements in topsoil and parent material; assessment of natural and*
1268 *anthropegenic enrichment factors*. Utrecht University.
1269 **Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D.,**
1270 **Burovski, E., Peterson, P., Weckesser, W., and Bright, J.** (2020). SciPy 1.0:
1271 fundamental algorithms for scientific computing in Python. *Nature Methods*
1272 **17**:261–272.
1273 **Waters, B. M., McInturf, S. A., and Stein, R. J.** (2012). Rosette iron deficiency transcript
1274 and microRNA profiling reveals links between copper and iron homeostasis in
1275 *Arabidopsis thaliana*. *Journal of Experimental Botany* **63**:5903–5918.
1276 **Weigel, D.** (2012). Natural variation in Arabidopsis: from molecular genetics to ecological

genomics. *Plant Physiology* **158**:2–22.

Weigel, D., and 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.

Wijfjes, R. Y., Smit, S., and 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., and Han, Y. (2023). The Arabidopsis LARP1s are Involved in Regulation of Seed Germination. *Journal of Plant Growth Regulation* **42**:1775–1788.

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

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

Zhang, B., Jia, J., Yang, M., Yan, C., and 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.

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

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

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: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., et al. (2017). Adaptation of *Arabidopsis thaliana* to the Yangtze River basin. *Genome Biology* **18**:239.