Title

Indirect genetic effects are shaped by demographic history and ecology in *Arabidopsis thaliana*

Author list

Germain Montazeaud[1,†], Quentin Helleu[1,2], Samuel E. Wuest[3], Laurent Keller[1,*, †]

Author's affiliations

[1]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

[2] Structure et Instabilité des Génomes, Muséum National d'Histoire Naturelle, CNRS UMR 7196, INSERM U1154, 43 rue Cuvier, F-75005 Paris, France

[3]Group Breeding Research, Division Plant Breeding, Agroscope, Wädenswil, Switzerland

[*]Present address: Social Evolution Unit, Cornuit 8, BP 855, Chesières, Switzerland.

[†]corresponding authors: g.montazeaud@gmail.com; laurentkeller01@gmail.com

Abstract

The phenotype of an individual can be affected by the genes of its conspecifics through Indirect Genetic Effects (IGEs). IGEs have been studied across different organisms including wild and domesticated animals and plants, but little is known about their genetic architecture. Here, we show in a large-scale intraspecific interaction experiment that the contribution of IGEs to the biomass variation of *Arabidopsis thaliana* is comparable to values classically reported in animals. Moreover, we identify eleven loci explaining 85.1% of the variability in IGEs. We find that positive IGE alleles (i.e., with positive effects on neighbour biomass) occur both in Relict accessions from southern Eurasia and in post-glacial colonizers from northern Scandinavia, and that they most likely have two divergent origins: for nine loci, they evolved in the post-glacial colonizers independently from the

Relicts, while the two others were introgressed in the post-glacial colonizer from the Relicts. Finally, we find that variation in IGEs most likely reflects divergent adaptations to the contrasting environments of the edges and the centre of the native range of the species. These findings reveal a surprisingly tractable genetic basis of IGEs in *A. thaliana* that is shaped by the ecology and the demographic history of the species.

Introduction

Interactions between biological organisms shape the diversity and functioning of ecosystems[1,2] and influence evolutionary processes[3]. Despite their importance for the evolutionary dynamics of populations, biological interactions do not feature explicitly in the classical paradigm of evolutionary genetics because they are hidden in the environmental component of the phenotype, and thus considered as unaffected by natural selection[4]. Griffing challenged this view by extending the classical formalism of quantitative genetics to account not only for the effect that genes have on the phenotype of their bearer (Direct Genetic Effects, DGEs), but also for the effect that genes exert on other individuals (Indirect Genetic Effects, IGEs)[5]. His work and later theoretical extensions have shown that IGEs can significantly alter both the magnitude and the direction of response to individual selection[4–8]. For example, when IGEs are negatively correlated with DGEs, they can reverse the response to selection in the opposite direction to the direction of individual selection[8].

Griffing's work was motivated by a practical problem faced by plant breeders: plants selected for high individual yield usually produce low yields when cultivated at high density in monoculture due to intense competition between individuals[5,9–11]. This problem was solved by the development of selection methods such as pedigree selection where crop varieties are selected based on stand (group) performance instead of individual performance[12,13]. Similar problems led to similar solutions in animal breeding where selection at the group level was shown to decrease aggressiveness between individuals and increase productivity and animal well-being[14,15].

While stand selection has remained the main approach to improve productivity in plants, animal breeding has built from Griffing's original models and developed quantitative genetics approaches

49 that allow the quantification of individual's IGEs, which can then be used as a direct selection
50 criterion[8,16]. These quantitative genetics approaches allow to quantify the relative contribution of
51 DGEs and IGEs on phenotypic variation, and to predict the response to selection for traits that are
52 affected by intraspecific interactions[16–19]. More recently, these methods have been applied to non-
53 farmed animals to better understand the genetic basis of inter-individual interactions. These studies
54 showed that IGEs affect many behavioural, morphological, and physiological traits[20,21], including in
55 humans[22], and even identified genes involved in interactions such as stress coping strategies and
56 wound healing in mice[20].

57 Plants engage in many types of interactions, including competition[23], facilitation[24,25] and allelopathy[26]
58 which can be modulated by the degree of relatedness between individuals[27–31]. However, contrary to
59 animals, very few studies have analysed intraspecific plant-plant interactions with the quantitative
60 genetic framework of IGEs[32–35]. Because IGE models partition the phenotypic variance into the direct
61 and the indirect effects of the genes, they provide a relevant framework to assess the evolutionary
62 consequences of intraspecific interactions. Moreover, they could help dissecting the genetic basis of
63 plant-plant interactions, which currently remain largely unknown[36,37]. Finally, as shown in animals,
64 the IGE framework could open new perspectives for plant breeding and help to identify and select
65 genes that have positive effects on stand-level performance (e.g., biomass or grain yield).

66 *Arabidopsis thaliana* has long been the model species for plant genetics and molecular biology.
67 Recently, it has also attracted the interest of ecologists and evolutionary biologists[38] following a
68 massive international effort to collect, characterize, and sequence large collections of natural
69 accessions[39,40]. These large-scale genomic data helped to reconstruct the recent demographic history
70 of the species, showing that the current distribution of *A. thaliana* in Eurasia results from two
71 consecutive waves of colonization that followed the last glacial maximum (~ 10 kya)[40,41]. First, plants
72 from several glacial refugia in southern Eurasia spread northwards as the glaciers retreated. After this
73 first wave of expansion, a second colonization happened from a population near the northern Balkan
74 and eastern Europe. This post-glacial colonizer, referred to as "non-Relict", expanded mainly along
75 the east-west axis and replaced most populations in the middle part of the range, leaving some of the

first colonizers referred to as "Relicts" only at the northern and southern edges. While Relicts are only found at the southern end of the range in present days, segments of Relict genomes are still present in northern and southern non-Relicts, potentially carrying adaptations to local abiotic environments[41,42]. Non-Relicts and Relicts differ not only in their evolutionary histories, but also in their preferred habitats: while Relicts are associated with relatively stable natural plant communities, non-Relicts exhibit a classical ruderal plant strategy adapted to disturbed habitats and anthropized environments[40,43]. The rapid and widespread colonization of Europe by a single population and the current wide geographic distribution of *A. thaliana* have motivated many studies to investigate the ecological and genomic basis of climate adaptation[44–48]. While most of these studies focused on interactions between plants and their abiotic environments, plant-plant interactions and their underlying genetic determinisms remain understudied[36].

In the present study, we combine large-scale genomic data and IGE models to identify the genetic basis of intraspecific plant-plant interactions and investigate how they have been shaped by evolution in *Arabidopsis thaliana*. We first use data from a published experiment[49] based on 98 natural accessions from the RegMap[39] panel to estimate the contribution of IGEs to phenotypic variation in three important traits (flowering time, rosette diameter, and aboveground biomass). We then use a genome-wide association study (GWAS) to identify the genomic regions associated with IGEs. Finally, we use the most up-to-date genomic data[40] to map the distribution of IGE alleles in the natural range of the species and to investigate the evolutionary forces that may have shaped this distribution.

Results

Plant biomass is affected by IGEs

In the experiment, 98 accessions from the RegMap[39] panel were each grown with 10 different neighbours, referred to as testers (Fig. 1a and b, see also ref[49]). Testers were chosen among the 98 accessions. They had different plant sizes which allowed to maximize the range of competitive environments faced by the 98 accessions (see Methods). Each accession was also grown as a monoculture (i.e., with a neighbour of the same genotype). Paired plants were grown in the same pot,

102 allowing to share both light and soil resources (Fig. 1a and b). Plants were harvested eight weeks after
103 sowing, and three traits were recorded: flowering time, rosette diameter, and aboveground biomass.
104 We fitted three models for each trait: Model 1 only accounted for direct genetic effects, Model 2
105 accounted for both direct and indirect genetic effects, and Model 3 accounted for both direct and
106 indirect genetic effects as well as their covariance. All three traits showed significant direct genetic
107 variance (Fig. 1c). Model 1 was the best model for both flowering time and rosette diameter
108 (Supplementary Table 1), indicating that these two traits were not significantly affected by IGEs in
109 the experiment. For these two traits, direct genetic variance accounted for 58.2% and 25.5% of the
110 total phenotypic variance, respectively (Fig. 1c). Model 3 was the best model for aboveground
111 biomass (Supplementary Table 1), indicating significant IGEs and significant DGE-IGE covariance
112 on this trait. Direct and indirect genetic variances accounted for 22.0% and 2.7% of the total
113 phenotypic variance of plant biomass, respectively (Fig. 1c). The genetic correlation between DGEs
114 and IGEs on biomass was strongly negative ($r = -0.88 \pm 0.03$, $p < 0.001$, Fig. 1d, Supplementary
115 Table 2), indicating that the alleles that increased the biomass of their bearer decreased the biomass of
116 their neighbours.

117 ***Eleven loci are associated with IGEs***

118 To investigate the genetic basis of IGEs, we conducted a Genome-Wide Association Study (GWAS)
119 on plant biomass using 206,416 bi-allelic SNPs distributed along the *A. thaliana* genome (all
120 accessions were homozygous at all SNPs). For each SNP, we fitted a linear model with biomass as the
121 response variable and two fixed effects: the allele carried by the focal individual, which was the DGE
122 of the SNP, and the allele carried by the neighbour individual, which was the IGE of the SNP (see
123 Methods). The GWAS detected eleven genomic regions with significant IGEs (Fig. 2a and
124 Supplementary Fig. 1). All eleven loci had DGEs and IGEs of opposite sign on plant biomass (i.e., the
125 allele associated with increased biomass in the neighbour (positive IGE) was systematically
126 associated with decreased biomass in the focal (negative DGE)) (Supplementary Fig. 2 and 3).
127 However, these DGEs were not significant when correcting for multiple testing at the genome-wide
128 level, and thus not detected by the GWAS (Supplementary Fig. 4). For all eleven IGE loci, the

129 positive IGE allele (i.e., the allele with a positive effect on neighbour biomass) was the rarest allele
130 among the 98 accessions (Supplementary Fig. 2). Individual IGE loci explained between 22.1 and
131 52.9% of the IGE variance (Fig. 2b), and as much as 85.1% jointly which accounts for 2.3% of the
132 total phenotypic variance (Fig. 2c). There was a strong positive linear relationship between the
133 biomass of a focal plant and the number of IGE loci bearing a positive IGE allele in its neighbour
134 (Fig. 2d, $F_{1,4170} = 157.27$, $p < 0.001$), showing that the IGEs of the different loci are cumulative.

135 ***Positive IGE alleles are enriched at extremal latitudes***

136 Naïvely, we would expect alleles that increase the biomass of neighbours but decrease the biomass of
137 their bearers to be outcompeted and lost from natural populations. To investigate where and why these
138 alleles are maintained, we first localized IGE variants in the Eurasian range of *A. thaliana* using
139 genomic data from 972 accessions from the 1001 Genomes project[40]. We considered nine non-Relict
140 ancestry groups (incl. one admixed) that diverged from each other after the last-glaciation, and one
141 group of Relicts that diverged from the non-Relicts before the last-glaciation[40,41] (Fig. 3a). Among
142 these accessions, linkage disequilibrium was moderate between IGE loci ($0 < r^2 < 0.29$,
143 Supplementary Fig. 5). Across all eleven loci, positive IGE alleles were much more frequent in the
144 non-Relicts from North Sweden than in all other non-Relicts (Fig. 3b, average frequency in North
145 Sweden = 0.76 vs 0.11 in all other non-Relicts). Positive IGE alleles were also present at intermediate
146 frequency in the Relicts ($f = 0.33$). This pattern was consistent for all IGE loci, except for three loci
147 where positive IGE alleles were absent from the Relicts (Supplementary Fig. 6).

148 ***Two IGE alleles were introgressed from Relicts***

149 Genomic analyses indicate that non-Relicts admixed with resident Relicts when they colonized the
150 northern and southern edges of Europe, leading to the introgression of Relict alleles adapted to the
151 local environments[40,41]. The presence of positive IGE alleles in North Sweden could therefore be the
152 result of introgressions from past Scandinavian Relicts into Scandinavian non-Relicts. This hypothesis
153 was supported by the presence of similar positive IGE alleles, sometimes at high frequency, in present
154 day Relicts (Fig. 3b and Supplementary Fig. 6). Near two IGE loci (chr1:6301080 and chr5:2838468),

155 the non-Relicts from North Sweden and the Relicts also shared similar haplotypes that were absent or
156 rare in all the other ancestry groups (Fig. 4a and Supplementary Fig. 7). We thus tested if these two
157 IGEs loci were introgressed from the Relicts into the non-Relicts from North Sweden using ABBA-
158 BABA statistics (Patterson's *D* statistic[50], Fig. 4b and d; and f_d statistic[51], Fig. 4c and e) computed
159 with the following topology: (((Western Europe, North Sweden), Relicts), *A. lyrata*) (see Methods).
160 As expected under the hypothesis that these two loci were introgressed from the Relicts into the non-
161 Relicts from North Sweden, both the *D* and f_d statistics were significantly higher in genomic windows
162 surrounding IGE loci than in background windows (chr1:6301080: $D = 0.40$ vs 0.05, $t_8 = -3.30$, $p =$
163 0.011, Fig. 4b; $f_d = 0.23$ vs 0.09, $t_8 = -2.42$, $p = 0.042$, Fig. 4c) and chr5:2838468 ($D = 0.31$ vs 0.05,
164 $t_{11} = -2.34$, $p = 0.040$, Fig. 4d; $f_d = 0.20$ vs 0.09, $t_{11} = -1.81$, $p = 0.098$, Fig. 4e). For all other IGE loci,
165 the *D* and f_d values were not significantly higher in the genomic windows surrounding IGE loci
166 compared to background windows, providing no evidence for genomic introgression (Supplementary
167 Fig. 8 and 9).

168 ***Two IGE loci are associated with local soil properties***

169 To investigate whether the maintenance of positive IGE alleles could be explained by adaptations to
170 the local environment, we ran Genome-Environment Association (GEA) studies using 196 climatic
171 variables measured at the collection site of each accession[52]. We considered that an IGE SNP was
172 associated with an environmental variable whenever a GEA SNP (i.e., a SNP significantly associated
173 with an environmental variable) was close to and in high linkage with that IGE SNP (see Methods and
174 Supplementary Fig. 10). These analyses revealed significant associations between IGE loci
175 chr2:19614933 and chr5:2838468 and local soil properties (Fig. 5). 18 SNPs close (~ 700 bp to 80 kb
176 to chr2:19614933, and ~70 kb to 400 kb to chr5:2838468) to and in high linkage ($0.51 \leq r^2 \leq 0.66$
177 with chr2:19614933, $0.52 \leq r^2 \leq 0.74$ with chr5:2838468) with these two loci were significantly
178 associated with soil organic carbon content (Fig. 5a-c). At these two loci, the positive IGE allele was
179 most common in environments with high levels of organic carbon in the soil (Fig. 5d: organic carbon
180 content = 2.9% vs 1.02%, $F_{1,904} = 65.90$, $p < 0.001$; Fig. 5e: organic carbon content 3.5% vs 1.1%,
181 $F_{1,904} = 68.09$, $p < 0.001$). 1 SNP ~25 kb away from and in high linkage with chr5:2838468 ($r^2 = 0.64$)

182 was also significantly associated with the amount of water extractable by plants in the soil (Fig. 5f-g).
183 At this locus, the positive IGE allele mostly occurred in environments with low amounts of water
184 available (Fig. 5h: plant extractable water = 10.12 cm/cm vs 13.01 cm/cm, $F_{1,853} = 15.33$, $p < 0.001$).

185 ***IGE loci are linked to plant growth strategies***

186 Overall, positive IGEs alleles mostly occurred at the northern and southern ends of the Eurasian range
187 of *A. thaliana* (Fig. 3b and Supplementary Fig. 6). These habitats are characterized by limited human
188 activity and more stable plant communities compared to the central part of the range. Previous studies
189 have indeed shown that the southern Relict accessions tend to occur in old oak and pine forests[40,43]
190 while the non-Relicts from North Sweden, too, live in tree-covered environments (Supplementary Fig.
191 11). This raises the possibility that IGE alleles could be favoured in habitats with contrasting
192 disturbance regimes and plant community types. This hypothesis is supported by the genetic
193 association between some IGE loci and soil organic carbon content (Fig. 5a-e), a variable expected to
194 be strongly affected by vegetation cover and perturbation regimes. Ecological theory predicts that
195 habitats with stable plant communities and high vegetation cover should select for a slower growth
196 rate and a more conservative resource-use strategy than disturbed environments[53–55]. We thus tested
197 this hypothesis by comparing the effect of IGE alleles on biomass production in the *absence* of
198 intraspecific interactions. To do so, we grew a single plant per pot of a subset of 83 accessions from
199 the initial set of 98. As predicted, positive IGE alleles were associated with lower biomass production
200 for nine of the eleven IGE loci, and this was also the case for the two other IGE loci, although the
201 effect was not significant (Supplementary Fig. 12). Moreover, biomass production was negatively
202 correlated with the number of positive IGE alleles across the eleven IGE loci (Fig. 6a). These findings
203 corroborate the negative correlation between IGEs and DGEs (Fig. 1d, Supplementary Fig. 2 and 3)
204 and support an association between IGEs and growth rate.

205 We also searched for genes close to IGE loci that had at least one non-synonymous, nonsense, or
206 frameshift mutation in high linkage ($r^2 \geq 0.5$) with these loci (see Methods). We obtained 128 genes,
207 among which 107 had functional annotations available (Supplementary Data 1). Among these 107
208 genes, there was a significant enrichment of biological functions related to the metabolism of salicylic

209 acid (SA), a key hormone involved in plant growth regulation[56,57] (Fold Enrichment = 34.03, p =
210 0.0277). SA has notably been shown to be a major regulator of the shade-avoidance syndrome
211 (SAS)[58,59], a set of competitive responses (e.g., stem elongation or reduction of branching) expressed
212 by plants when grown in shaded or crowded environments[60]. Eleven genes related to light responses
213 and shade-avoidance were close to and in high linkage with five IGEs loci including chr5:2838468,
214 the locus with the strongest IGE (Fig. 6b-f).

215 To further test the hypothesis that these eleven genes could be subject to divergent selection in
216 different ecological habitats, we compared their fixation index (F_{ST}) values between the non-Relicts
217 from North Sweden (northern edge, low disturbance, high vegetation cover), the Relicts from
218 southern Eurasia (southern edge, low disturbance, high vegetation cover), and all other non-Relict
219 groups from central Eurasia (centre, high disturbance, low vegetation cover). Apart for two genes
220 which had low F_{ST} values between the non-Relicts from North of Sweden and Germany, the F_{ST}
221 values between the non-Relicts from North Sweden and all other non-Relicts were greater for the
222 eleven light-related genes than the median F_{ST} values for all the other annotated genes (n = 27,005
223 annotated genes, Fig. 6g). Moreover, the F_{ST} values for light-related genes between the non-Relicts
224 from North Sweden and the Relicts were on average lower than the F_{ST} values between the non-
225 Relicts from North Sweden and the other non-Relicts (Fig. 6h), showing that the North Swedish non-
226 Relicts are more similar to the Relicts than to the other non-Relicts at these genes. Altogether, these
227 results further support the view that natural variation at IGE loci is associated with differences in
228 growth rates and light responses, two life-history traits expected to be under differential selection in
229 different ecological habitats.

230 Discussion

231 Using a representative sample of the natural variation of *A. thaliana*, our study showed that plant
232 biomass production was significantly influenced by intraspecific interactions mediated through IGEs.
233 We identified eleven loci explaining 85.1% of the IGE variation corresponding to 2.3% of phenotypic
234 variation. At all loci, the rarest alleles had a positive effect on neighbour biomass. We found that these

235 positive IGE alleles mostly occurred in non-Relicts from Northern Sweden and in Iberian Relicts. Our
236 analyses showed that most of these alleles (i.e., nine out of eleven) seem to have evolved
237 independently in these two groups, while the two others were most likely introgressed from
238 Scandinavian Relicts into the non-Relicts. Allelic variation at two IGE loci, including one
239 introgressed from the Relicts, was associated with soil organic carbon content and soil water
240 availability, suggesting that positive IGE alleles could provide adaptations to the environments of the
241 northern and southern ends of the species 'range. Consistent with this, we found genetic associations
242 between IGE loci and both growth rates and growth responses to light stimulus, two traits expected to
243 be under divergent selection between the ecological habitats found at the margins (high environmental
244 stress, low disturbance, stable plant communities) and at the centre of the range (low environmental
245 stress, higher disturbance linked to human activities, and less stable plant community).

246 In our experiment, the additive contribution of neighbour's genes to the biomass of the focal plant
247 accounted for 2.7% of the phenotypic variance. Although this proportion appears small, it is
248 comparable to social effects classically observed in animals. For example, the contribution of IGEs to
249 aggressive behaviours in crickets (*Gryllus bimaculatus*) or mink (*Neovison vison*) has been shown to
250 account for about 3% of the total phenotypic variance[18,21]. In larger-scale studies, IGEs have been
251 shown to account on average for 2.9% of the phenotypic variance in 170 behavioural, physiological,
252 and morphological traits in laboratory mice (*Mus musculus*) and 1.5% of the phenotypic variance in
253 51 dietary, mental health, and disease related traits in humans (*Homo sapiens*)[20,22]. This does not mean
254 that some traits may be more strongly affected by intraspecific interactions. For example, IGEs have
255 been shown to account for 22% and 25% of the phenotypic variance in LDL cholesterol levels in mice
256 and shoot number in the clonal herb *Sedum album*, respectively[20,35]. However, the IGE estimate of the
257 *Sedum* study is difficult to compare with our own because it combines the additive and non-additive
258 components of IGEs, whereas here we have quantified only the additive component (i.e., the only
259 component that directly affects the response to selection).

260 The strong negative correlation between DGEs and IGEs on biomass suggests that IGEs contribute
261 mainly to competitive plant-plant interactions. Competition can explain both positive and negative

262 interactions: competitive individuals increase their biomass at the expense of their neighbour (positive
263 DGEs, negative IGEs), while weaker competitors leave a greater proportion of local resources
264 available for their neighbour (negative DGEs, positive IGEs). This interpretation is consistent with the
265 fact that IGE loci are associated with growth rates and growth responses such as shade-avoidance. A
266 previous IGE study in *A. thaliana* found positive DGE-IGE correlations for most traits[34]. However,
267 this study relied on Recombinant Inbred Lines derived from only two accessions and therefore may
268 have examined only a small subset of the natural variation in plant-plant interactions. In addition, the
269 authors did not correct for non-genetic sources of correlation between paired plants in their analysis.
270 Because paired plants share a similar pot, positive DGE-IGE correlations might arise from similar
271 environmental effects on the phenotype of the two interaction partners. Consistent with our results,
272 most IGE studies conducted on trees (*Eucalyptus globulus* and *Pinus taeda*) also found strong
273 negative DGE-IGE correlations for growth-related traits[32,33,61], which has also been attributed to the
274 effect of competition between adjacent trees.

275 Eleven loci in low-to-moderate linkage disequilibrium with each other accounted for 85.1% of the
276 IGEs variance. There was a strong linear relationship between the number of positive IGE alleles in a
277 focal plant and the biomass of its neighbour, meaning that the effects of the eleven loci are cumulative
278 Previous studies that investigated the genetic basis of IGEs in *A. thaliana* generally found a lower
279 number of QTLs than in our study. For example, Mutic and Wolf only found two QTLs with major
280 effects on neighbour biomass[34], and Baron *et al*. found a single GWAS peak associated with the effect
281 of *A. thaliana* on the biomass of another competitor species (*Trifolium arvense*)[62]. A possible
282 explanation is that these two studies used a smaller number of genotypes obtained from laboratory
283 crosses or a single regional population whereas we combined many genotypes from different
284 environments. Despite the high number of IGE QTLs identified in our study, we did not identify a
285 locus that was previously found to be associated with cooperative plant traits using the same
286 experimental data[49]. This can be explained by a difference in the analytical approaches used in the two
287 studies: while the previous study corrected for size differences between genotypes, our IGE analyse
288 did not, and because the allele associated with cooperative traits happened to be found -most likely by

289 sampling effect- in genotypes that were slightly larger on average, the positive effect of cooperation is
290 masked by the negative effect of competition (caused by size differences) in our analysis.

291 Our results suggest that the positive IGE alleles have two distinct evolutionary origins. For nine of the
292 eleven loci, the positive IGE alleles most likely evolved once in the Relicts from Iberia that were
293 present in southern edge of the range before the last glaciation, and another time in the non-Relicts
294 that colonized northern Europe after the last glaciation. For the two remaining IGE loci, positive
295 alleles were most likely introgressed from a local Relict population into the post-glacially spreading
296 lineage during the colonization of Scandinavia. Previous studies have already shown that non-Relict
297 genomes at the northern and southern edges of the range harbour many haplotypes that were
298 introgressed from the local Relicts during colonization, and it has been suggested that these
299 haplotypes could have provided the post-glacial colonizer with locally adapted alleles[41]. Our results
300 partly support this hypothesis, as one of the introgressed IGE locus is associated with local soil
301 properties.

302 Because positive IGE alleles seem to have a cost for their bearer (i.e., reduced biomass), one would
303 naively expect them to be lost from natural populations. Based on our results, we can advance two
304 hypotheses to explain how they could be maintained despite this cost. First, positive IGE alleles could
305 either provide adaptations to the habitats found at the edges of the range or be genetically linked to
306 variants that have a fitness advantage in these habitats (which is more likely in a highly selfing species
307 such as *A. thaliana* because selfing reduces the effective rate of recombination[63]). These adaptations
308 could in turn affect plant-plant interactions because they involve changes in growth forms and growth
309 responses to light stimulus (i.e., photomorphogenesis). Previous studies have already reported
310 variation in growth strategies[46–48] and light responses[64–67] along latitudinal gradients in *A. thaliana*.
311 Latitudinal variation in light responses has classically been attributed to adaptation to the latitudinal
312 gradient in light intensity (i.e., higher light intensity selects for lower light sensitivity at lower
313 latitudes, and vice versa). Our results suggest that such variation could also be linked to plant-plant
314 interactions. Second, positive IGE alleles could be selected and maintained by kin selection (i.e.,
315 costly traits can be selected if the cost (c) for the bearer of the trait is lower than the benefit (b)

316 received by the other interacting partners, weighted by the relatedness between the bearer of the trait
317 and the other interacting individuals (r): c < rb[68]). According to this hypothesis, we would expect a
318 higher r, or a greater effect of the trait on b, or a weaker effect on c, in populations of *A. thaliana* from
319 North Sweden and Southern Iberia than in populations from the central part of the range.

320 In conclusion, our study reveals significant IGEs in an annual plant. Theoretical work has shown that
321 strong negative DGE-IGE correlations, as observed in our study, may gradually reduce heritable
322 variation and eventually suppress evolutionary potential[33,69,70]. Biomass evolution could then be or
323 could have been constrained by plant-plant interactions in *A. thaliana*. The fact that genetic variation
324 exists for plant-plant interactions also suggests that these interactions should be considered in plant
325 breeding programs, as has been done in animals. This would allow the selection for less competitive
326 and more cooperative crop varieties.

327 METHODS

328 *Plant material*

329 The experiment is already described in ref[49] and was conducted with a subset of 98 natural accessions
330 from the RegMap panel. The RegMap panel consists of 1,310 worldwide accessions, including
331 several regional panels, that have been genotyped with a 250K SNP (Single Nucleotid
332 Polymorphisms) chip and for which high-quality geographic coordinates have been collected[39]. It
333 constitutes one of the largest genetic resource available for non-human species, and it has already
334 been used to investigate the genetic basis of many ecologically important traits in *A. thaliana*[71–74]. The
335 98 accessions used in the experiment are a subset of the panel for which a comprehensive set of
336 phenotypic traits has been collected[75]. The geographic distribution of these 98 accessions covers most
337 of the Eurasian range of the species (Supplementary Fig. 13).

338 *Experimental design*

339 Pairs of individual plants were grown in 6 x 6 x 5.5 cm pots. A pair was composed of one accession
340 and one tester genotype. Each accession was grown with ten different testers (Bay-0, C24, Col-0, Cvi-

341 0, Ler-1, Sav-0, Sf-2, Shahdara, St-0, Uk-1) following a full factorial design. Each accession was also
342 grown as a monoculture with two plants of the same genotype in the same pot (Fig. 1a and b). Testers
343 were a subset of the 98 lines. They constitute the parents of different publicly available recombinant
344 inbred line populations, which facilitate more advanced genetic studies (e.g., see ref[49]). They also had
345 different plant sizes which allowed to maximize the range competitive environments faced by the 98
346 accessions. A PCA (Principal Component Analysis) based on 206,416 SNPs (Single Nucleotide
347 Polymorphisms) confirmed that the ten testers were representative of the overall genetic variation
348 present in the 98 accessions (Supplementary Fig. 14, PERMANOVA comparing testers vs other
349 accessions: $F_{1,96} = 1.13$, $p = 0.1196$). The overall design was replicated in two blocks. The initial set
350 of genotypes comprised 97 accessions, but the lack of seeds for one accession led to its replacement in
351 the second block with another accession, resulting in a total of 98 accessions each grown with ten
352 testers. This gave a total of 2134 pots containing two plants each. A second experiment was
353 conducted to measure plant biomass in the absence of intraspecific interactions. This experiment was
354 conducted with a subset of 83 accessions with a single plant per accession per pot (7 x 7 x 8 cm), and
355 no replication, as described previously[49].

356 ***Plant growth conditions***

357 Seeds of all accessions were sown directly in soil (four parts Einheitserde ED73, Gebrüder Patzer,
358 Germany; one part quartz sand) in February 2016. Pots of a given block were randomly placed into
359 trays covered with plastic lids for germination. To ensure the growth of two plants per pot, multiple
360 seeds were sown (approx. 5-20 seeds) per position in a pot, and the two genotypes (and all
361 monocultures) were sown approximately 3-4 cm apart. Once seeds had germinated, surplus seedlings
362 were removed, such that only one (two for monocultures) healthy seedling remained per genotype per
363 pot. Block 1 was sown on February 17th and block 2 on February 18th 2016, and pots were placed in
364 trays in a greenhouse compartment. Additional light was provided if necessary to achieve a
365 photoperiod of 14 hours. Day-time and night-time temperatures were maintained around 20–25 °C
366 and 16–20 °C, respectively. Trays were randomly re-arranged within the greenhouse every 3-5 days.
367 After 5-5.5 weeks, pots were transferred from trays onto three tables with automated watering and

368 randomly re-arranged weekly. Flowering shoots of individual plants were tied to wooden sticks as
369 they grew taller than approx. 10 cm. Plants received no vernalization treatment, so accessions with
370 strict vernalization requirement never flowered. For this reason, aboveground plant biomass (instead
371 of fecundity) was measured destructively as a performance trait, and at a time when most accessions
372 were still in their active growth phase. Plants were harvested on April 14th (Block 1) and April 15th
373 (Block 2) 2016, approx. eight weeks after sowing. For the experiment with single individuals per pot,
374 plants were grown for 43 days on a mixture of one part ED73 and four parts quartz sand.

375 ***Trait measurements***

376 At harvest, each plant was cut below the rosette and individually dried at 65°C for 4-5 days and then
377 stored at room temperature until weighing to measure aboveground dry biomass. Flowering time was
378 measured over the course of the experiment by scoring all individuals that had a flowering bolt of >
379 0.5 cm every 2-3 days. Rosette diameter was measured on day 34 and 35 after sowing with a precision
380 of 0.5 cm.

381 ***Statistical analysis***

382 Unless otherwise stated, all statistical analysis were conducted with R v. 4.2.3[76].

383 Quantification of Indirect Genetic Effects: To estimate the contribution of IGEs to the total
384 phenotypic variance, we followed the quantitative genetic models described in ref[8]. For each of the
385 three traits (flowering time, rosette diameter, and aboveground biomass), we fitted three linear mixed
386 models. Model 1 only accounted for direct genetic effects:

387 (1) $$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{Z}_b\boldsymbol{b} + \boldsymbol{Z}_D\boldsymbol{a}_D + \boldsymbol{\varepsilon},$$

388 Where $\boldsymbol{y}$ is the vector of phenotypic observations, $\boldsymbol{Z}_b$ and $\boldsymbol{Z}_D$ are known incidence matrices, μ is the
389 mean phenotype, $\boldsymbol{b} \sim N(0, \boldsymbol{I}_b\sigma_b^2)$ ("~" meaning "distributed as") is the random block effect,
390 $\boldsymbol{a}_D \sim N(0, \boldsymbol{A}\sigma_{A_D}^2)$ is the random additive direct genetic effect, and $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{R}\sigma_e^2)$ is the random error.
391 σ_b^2, $\sigma_{A_D}^2$, and σ_e^2 are the block variance, additive direct genetic variance, and residual variance,
392 respectively. The vector of phenotypic observations, $\boldsymbol{y}$, concatenates the phenotypes of both the 98

accessions and the ten testers. $\boldsymbol{I}_b$ is an identity matrix, $\boldsymbol{A}$ is a matrix of additive genetic relationships among all accessions, and $\boldsymbol{R}$ is a matrix of correlation between residuals such that $R_{ii} = 1$, $R_{ij} = \rho$ when plants *i* and *j* are in the same pot, and $R_{ij} = 0$ when plants *i* and *j* are in different pots. $\rho \in [-1,1]$ is the correlation between the residuals of two plants that share the same pot. It is included in the model to account for sources of non-genetic covariance between the phenotypes of two interacting plants. Such non-genetic covariance can arise if, for example, pots differ slightly in their soil composition or microclimates, which may create some correlation between the phenotypes of the two interacting plants and bias estimates of genetic variances[8]. Model 2 has an additional random term to account for indirect (sometimes also referred to as "social") genetic effects:

(2) $$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{Z}_b\boldsymbol{b} + \boldsymbol{Z}_D\boldsymbol{a}_D + \boldsymbol{Z}_S\boldsymbol{a}_S + \boldsymbol{\varepsilon},$$

where $\boldsymbol{Z}_S$ is a known incidence matrix, and $\boldsymbol{a}_S \sim N\left(0, \boldsymbol{A}\sigma^2_{A_S}\right)$ is the random additive indirect (or "social") genetic effect with $\sigma^2_{A_S}$ being the additive indirect genetic variance. Model 3 has the same equation as model 2 (equation (2)) but $\boldsymbol{a}_D$ and $\boldsymbol{a}_S$ are assumed to covary such that:

(3) $$\begin{bmatrix} \boldsymbol{a}_D \\ \boldsymbol{a}_S \end{bmatrix} \sim MVN(0, \boldsymbol{C} \otimes \boldsymbol{A}),$$

with

(4) $$\boldsymbol{C} = \begin{bmatrix} \sigma^2{}_{A_D} & \sigma_{A_{DS}} \\ \sigma_{A_{DS}} & \sigma^2{}_{A_S} \end{bmatrix},$$

in which *MVN* is the multivariate normal distribution, $\sigma_{A_{DS}}$ is the covariance between additive direct and indirect genetic effects. $\sigma_{A_{DS}}$ was used to compute the genetic correlation between direct and indirect genetic effects, $r_{A_{DS}}$, with $r_{A_{DS}} = \frac{\sigma_{A_{DS}}}{\sigma_{A_D}\sigma_{A_S}}$ (see Supplementary Table 2 and Fig. 1d).

The three models were fitted using Restricted Maximum Likelihood (REML) and then compared using likelihood ratio tests. Results from these comparisons are reported in Supplementary Table 1. Raw variance estimates from the best models for each trait are reported in Supplementary Table 2. Mixed model fitting and comparison were done with ASReml-R[77]. We computed the direct and

416 indirect heritabilities of the traits by dividing their estimated direct genetic variance ($\widehat{\sigma^2_{A_D}}$) or indirect
417 genetic variance ($\widehat{\sigma^2_{A_S}}$) by the total phenotypic variance.

418 Genome-Wide Association Study: To identify the SNPs associated with IGEs, we conducted a
419 Genome-Wide Association Study (GWAS) using the plant biomass data from the experiment[49] and
420 SNP data at 206,416 sites[39] (sites with minor allele frequency < 5% were preliminary removed). The
421 GWAS was conducted by fitting a linear mixed model recursively for each SNP. We used Model 3 as
422 a baseline model, to which we added the fixed effect of the IGE of the tested SNP (i.e., the effect of
423 the allele carried by the neighbour individual at the tested SNP). This model allowed us to account for
424 population structure, the variance of the random genetic terms being structured with the additive
425 relationship matrix $\boldsymbol{A}$ (see previous section)[78]. Because previous work showed that not accounting for
426 the DGEs of the SNPs when running GWAS on IGEs can increase the number of false positives in
427 settings like ours where both accessions and testers are included in the model and DGEs are strongly
428 correlated to IGEs (Fig. 1d, see Supplementary Note in ref[20]), we included the fixed effect of the DGE
429 of the SNP (i.e., the effect of the allele carried by the focal individual at the tested SNP) before the
430 IGE fixed effect in the model. We obtained two vectors of *p*-values (i.e., one for DGEs, one for IGEs)
431 on which we independently applied multiple testing corrections to detect the SNPs significantly
432 associated with DGEs and IGEs. We used a genome-wide False-Discovery Rate (FDR[79]) of 5% for
433 both effects. We fitted the GWAS models and tested fixed effect significance with ASReml-R[77]. We
434 then computed the indirect additive genetic variance of individual top GWAS SNPs with the
435 following formulae:

436 (5) $$\widehat{\sigma^2_{A_{S_i}}} = \widehat{\beta_{S_i}}^2 \times 2MAF_i(1 - MAF_i)$$

437 Where $\widehat{\beta_{S_i}}^2$ is the squared estimated marginal effect size of SNP *i* of the neighbour on the biomass of
438 the focal plant and is obtained from the single-locus GWAS model, and MAF_i is the minor allele
439 frequency of SNP *i*. The proportion of IGE variance explained by a given SNP was computed as

440 $\frac{\widehat{\sigma^2_{A_{S_i}}}}{\widehat{\sigma^2_{A_S}}} \times 100$. We also computed the joint IGE variance of the *p* top SNPs from different GWAS loci

441 accounting for LD between the *p* variants:

442 (6) $$\widehat{\sigma^2_{A_{S_{joint}}}} = \widehat{\boldsymbol{\beta}_S}^{*T} \boldsymbol{R}^{-1} \widehat{\boldsymbol{\beta}_S}^{*},$$

443 Where $\widehat{\boldsymbol{\beta}_S}^{*}$ is the vector of estimated scaled marginal effect size of the *p* SNPs ($\widehat{\beta_{S_i}}^{*} =$

444 $\widehat{\beta_{S_i}} \times \sqrt{2MAF_i(1 - MAF_i)}$) and $\boldsymbol{R}$ is the LD-matrix (Pearson correlations) of the *p* SNPs. The

445 proportion of IGE variance jointly explained by a set of *p* SNPs was computed as $\frac{\widehat{\sigma^2_{A_{S_{joint}}}}}{\widehat{\sigma^2_{A_S}}} \times 100$. To

446 further test the combined effect of the different IGE loci, we checked the linear relationship between

447 plant biomass and the number of loci with a positive IGE allele across the *p* top SNPs in the

448 neighbour.

449 <u>Localization of IGEs variants:</u> To locate IGE variants within the native range of *A. thaliana*, we used

450 genomic data from the 1001 Genomes project[40]. We kept 972 accessions with latitudes comprised

451 between 25°N and 72°N, and longitudes comprised between 20°W and 75°E (i.e., removing

452 accessions outside Eurasia). We compared the frequency of positive IGE alleles among ancestry

453 groups identified by the 1001 Genomes project[40]. We estimated Linkage Disequilibrium (LD)

454 between the top IGE SNPs in the 972 accessions using the snpStats package[80] (Supplementary Fig. 5).

455 <u>Introgression tests:</u> To test whether IGE loci could have been introgressed from Relicts into

456 Scandinavian accessions, we first visually compared haplotypes around IGE sites between the

457 different ancestry groups using the GenotypePlot package[81]. We then computed ABBA-BABA

458 statistics, again using genomic data of 972 Eurasian accessions from the 1001 Genomes project[40].

459 These statistics are aimed to test for deviation from strict bifurcating evolutionary history and are thus

460 commonly used to detect introgressions. Given a phylogenetic topology with three populations and an

461 outgroup such as (((P1,P2),P3,)O), and calling the ancestral alleles (the ones present in the outgroup)

462 A and the derived alleles B, we expect to observe two discordant genealogies across the genome :

463 ABBA patterns, which group P2 and P3 together, and BABA patterns, which group P1 and P3

464 together. These discordant genealogies are expected to occur in roughly equal proportions due to

465 incomplete lineage sorting. An excess of ABBA relative to BABA patterns, however, is indicative of
466 gene flow between P2 and P3, and *vice versa*. The Patterson's *D* statistic is the difference in the sum
467 of ABBA and BABA patterns across the entire genome or a given genomic region[50]. A positive *D*
468 value is thus indicative of introgression between P2 and P3. *D* was initially developed to detect
469 introgression at the genome-wide level, and it has been shown that *D* values can be inflated when
470 computed over small genomic regions[51]. The f_d statistic was thus developed as an alternative to *D* to
471 detect introgression in particular genomic regions[51]. Here, we computed both *D* and f_d in non-
472 overlapping 20 kb genomic windows using custom Python scripts from
473 https://github.com/simonhmartin/genomics_general. We only included windows with at least 250
474 SNPs. To test for introgression between the Relicts and the non-Relicts from North Sweden, we used
475 the following tree topology: (((Wester Europe, North Sweden), Relicts), *A. lyrata*). We used *A. lyrata*
476 as an outgroup because it is the sequenced species the most closely related to *A. thaliana*. We
477 retrieved genomic data for 10 *A. lyrata* accessions from NCBI BioProject PRJEB30473. We then
478 compared ABBA-BABA statistics between IGE windows (window midpoint located in the ± 115 kb
479 interval around top IGE SNPs) and background genomic windows (window midpoint located outside
480 the ± 115 kb interval around top IGE SNPs) using a two-sided Welch's *t*-tests.

481 Genome-Environment Association study: To test whether IGE loci could be linked to climatic
482 adaptations, we ran a Genome-Environment Association (GEA) study using 196 climatic variables
483 extracted from the CLIMtools database[52] and SNP data for the 1,135 accessions available in the 1001
484 Genomes project[40]. This analysis tests for statistical associations between SNPs and climate at site of
485 origin using a similar approach to GWAS, except that climatic variables are used as response
486 variables instead of phenotypic traits. We ran this analysis at the genome-wide level, i.e., blind to
487 IGEs loci. We first removed all SNPs with minor allele frequencies < 5%, which left us with 441,192
488 SNPs. We then computed the additive relationship matrix between the 1,135 accessions which was
489 used in the GEA model to account for population structure[78]. We computed the additive relationship
490 matrix and conducted the GEA with the gemma program[82]. We used a False-Discovery Rate (FDR[79])
491 of 5% to determine which SNPs were significantly associated with a given climatic variable (hereafter

492 called GEA SNPs). We considered that an IGE SNP was associated with a climatic variable whenever
493 a GEA SNP for that variable was close to and in high linkage with that IGE SNP. More precisely, IGE
494 SNPs had to have at least one significant GEA SNP located within their half LD decay distance and
495 with a LD value higher than 0.5. Pairwise LD between individual SNPs was measured with r^2 across
496 the complete 1001 genomes dataset using plink-ng (v. 1.9.20200712)[83]. To compute the half LD
497 decay distance (i.e., the distance at which LD has halved), we first removed all SNPs with LD values
498 lower than the quadratic mean of the pairwise interchromosomal LD. Because LD values were highly
499 skewed towards zero, we applied a second filter by computing the 0.95 percentile LD (r^2) value in
500 non-overlapping 2500 kb distance bins. With the remaining LD values, we then modelled the LD vs
501 distance relationship with a non-linear regression following equations by Hill and Weir (1988)[84], and
502 we used the fitted values of the regression to compute the half LD distance. We repeated this
503 operation for each IGE SNP, which gave us window sizes adjusted for local LD decay
504 (Supplementary Fig. 10). We also compared land cover types between *A. thaliana* accessions using
505 the Global Land Cover 2000 database[85] accessed through CLIMtools[52].

506 Single-plant biomass analysis: We compared biomass production between IGE alleles in the absence
507 of intraspecific interactions using the second experiment with 83 accessions grown with one plant per
508 pot. We fitted a mixed model with biomass as the response variable, allelic value at IGE locus ("+ +"
509 vs "- -") as a fixed effect, and accession as a random effect. We accounted for covariance between
510 accessions using an additive relationship matrix for the random term. We fitted the model and
511 assessed significance of the fixed effect using the ASReml-R program[77].

512 Candidate gene investigation: To investigate candidate genes underlying IGEs, we used genomic data
513 of the 1,135 accessions from the 1001 Genome project[40]. We first identified all non-synonymous,
514 nonsense, or frameshift polymorphisms that were close to and in high LD with the SNPs associated
515 with IGEs using the R package VariantAnnotation v3.17[86]. We used the same criteria of proximity
516 and LD than for the GEA analysis, i.e., polymorphisms had to be in the half LD decay distance and to
517 have a r^2 value higher than 0.5 with IGE SNPs. Candidate genes were the ones with at least one
518 polymorphism meeting these two criteria. We also computed F_{ST}[87] over all annotated genes using

EggLib v3.1.0[88]. Finally, we checked the biological functions of the candidate genes using annotations from the TAIR10 genome release (https://www.arabidopsis.org/). We used these annotations to perform GO (Gene Ontology)-enrichment analysis with the TAIR online tool for GO Term Enrichments for Plants (powered by PANTHER).

Data Availability

The data analysed in this study are available in Zenodo: https://doi.org/10.5281/zenodo.7944154

Code Availability

All the code used for statistical analysis is available in Zenodo:
https://doi.org/10.5281/zenodo.7944154

Acknowledgements

We thank Christian Fankhauser, Tomas Kay, and Qiaowei Pan, and for helpful comments on the manuscript. We thank Aurélien Estarague, François Vasseur, and Cyrille Violle and for insightful discussions. This work was founded by the University of Lausanne, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement SOCLE (no. 101030712) to G.M. S.E.W. acknowledges funding from the Swiss National Science Foundation, project 310030_192537.

Author contributions

G.M. and S.E.W conceptualized the study based on data previously collected by S.E.W. G.M. conducted the quantification of indirect genetic effects, the genome-wide association study, and the genome-environment associations. G.M. and Q.H. performed introgression tests and investigated candidate genes. G.M. wrote the manuscript with L.K. All the authors commented on the manuscript. L.K. advised and oversaw the project.

541 # Competing interests

542 The authors declare no competing interests.

543 # Tables

544 # Figure captions

545 **Fig. 1: Overview of the experimental design and quantification of Indirect Genetic Effects**
546 **(IGEs). a**: Schematic representation of the experimental design. G1, G2, … G98: focal genotypes,
547 natural *A. thaliana* accessions sampled from the RegMap panel. T1, T2, … T10: one of ten tester
548 genotypes chosen to capture a large portion of the genetic variation present within *A. thaliana* and to
549 represent different plant sizes, M: monocultures. Plant illustrations from BioRender.com. **b**: Picture of
550 the experimental setup (credit: S.E. Wuest). **c**: Bar plots showing the estimated variance components
551 (top of the bar, center of error bars) ± standard errors (error bars). For each trait, we obtained one
552 estimate of each variance components from mixed model solving (see Methods). Variance estimates
553 were divided by the total phenotypic variance and multiplied by 100. DGE: direct genetic variance,
554 IGE: indirect genetic variance, cov(DGE, IGE): covariance between direct and indirect genetic
555 effects. All raw variance estimates are reported in Supplementary Table 2. **d**: Relationship between
556 DGEs and IGEs on aboveground biomass. Points correspond to the direct and indirect breeding values
557 of each accession ($n = 98$, see Methods). Direct and indirect breeding values are expressed as
558 deviation from the population mean. The reported p-value refers to the simple linear regression
559 between IGE and DGE breeding values (two-sided F-test, $F_{1,96} = 942.35$, $p < 2.2e-16$). r: genetic
560 correlation between direct and indirect breeding values (Supplementary Table 2).

561 **Fig. 2: Genome-Wide Association Study (GWAS) of Indirect Genetic Effects (IGEs) on plant**
562 **biomass. a**: Manhattan plot reporting p-values (-log10 transformed) for the association tests between
563 IGEs on biomass and allelic variation at 206,416 SNPs distributed along the genome of *A. thaliana*.
564 The p-values correspond to a per SNP two-sided Wald-test corrected for multiple comparison using a
565 False Discovery Rate (FDR) of 5% (10%), here represented with a solid red (blue) line. The points

566 highlighted in red are the most significant SNPs and the surrounding SNPs at ± 300 kb. **b**: Proportion
567 of IGE variance explained by each of the eleven top IGE SNPs. **c**: Proportion of IGE variance
568 explained by the eleven top IGE SNPs jointly (i.e., accounting for their respective effect sizes and
569 linkage disequilibrium between them). **d:** Violin plots showing the distribution of plant biomass of a
570 focal plant as a function the number of loci bearing a positive IGE allele in the neighbour. Points and
571 error bars represent means ± standard deviation. The reported *p*-value refers to the simple linear
572 regression between biomass and positive IGE allelic count in the neighbor (two-sided *F*-test, $F_{1,4170}$ =
573 157.27, $p < 2.2e-16$). The number of accessions in each category is reported below each violin plot.

574 **Fig. 3: Geographic distribution of IGE variants. a**: Localization of 972 *A. thaliana* accessions from
575 the 1001 Genomes project. Accessions are colored according to ancestry groups. Ancestry groups
576 include the non-Relicts (North Sweden, South Sweden, Germany, Asia, Western Europe, Admixed,
577 Central Europe, Spain, Italy-Balkans-Caucasus), and the Relicts. **b**: Frequency of positive IGE alleles
578 in each ancestry group, averaged across the eleven IGE loci. Points and error bars represent means ±
579 standard deviation (North Sweden, n = 64; South Sweden, n = 156; Germany, n = 61; Asia, n = 48;
580 Western Europe, n = 110; Admixed, n = 128; Central Europe, n = 64; Spain, n = 61; Italy-Balkans-
581 Caucasus, n = 92; Relicts, n = 24.).

582 **Fig. 4: Admixture between non-Relicts from North Sweden and Relicts at IGE loci. a**: Patterns of
583 haplotype sharing between ancestry groups in an 80kb region surrounding chr5:2838468 (red arrow).
584 Each row corresponds to an individual, individuals are grouped by ancestry groups, and ancestry
585 groups are separated by black horizontal lines. Each column corresponds to a SNP, with the reference
586 allele colored in yellow, the alternative allele colored in black (all individuals are homozygous), and
587 missing genotypes colored in white. A ~ 80 kb haplotype block shared between the non-Relicts from
588 North Sweden and the Relicts is highlighted in red. **b-e**: ABBA-BABA statistics (**b** & **d**: Patterson's
589 *D*, **c** & **e**: f_d) compared between background genomic windows (grey, n = 5549 windows) and IGE
590 windows (red) for the two IGE loci chr1:6301080 (n = 9 windows) and chr5:2838468 (n = 12
591 windows). *D* and f_d were both computed in 20 kb sliding windows along the genome with the
592 following population relationships: ((P1,P2),P3),O) with Western Europe as P1, North Sweden as P2,

593 Relicts as P3, and *A. lyrata* as the outgroup (O). Points correspond to mean values across genomic
594 windows and error bars to standard deviations. Background windows and IGE windows were
595 compared with two-sided Welch's t-tests (".": $p < 0.1$, "*": $p < 0.05$; Exact p-values: **b**: $t_8 = -3.30$, $p =$
596 0.011, **c**: $t_8 = -2.42$, $p = 0.042$, **d**: $t_{11} = -2.34$, $p = 0.040$, **e**: $t_{11} = -1.81$, $p = 0.098$).

597 **Fig. 5: Association between IGE loci and environmental variables**. **a** & **f**: Manhattan plots
598 reporting p-values (-log10 transformed) for the association tests between allelic variation at 441,192
599 SNPs and the content of organic carbon in the topsoil (log transformed, **a**), and the amount of water
600 extractable by plants in the soil (**f**). The p-values correspond to a per SNP two-sided Wald-test
601 corrected for multiple comparison using a False Discovery Rate (FDR) of 5%, here represented with a
602 solid red line. The points highlighted in red are SNPs in a window of ± 500 kb around the SNP with
603 the most significant association with the environmental variable. **b-g**: Zoomed Manhattan plots
604 showing local linkage disequilibrium (LD) between environmental SNPs and the closest top IGE SNP
605 (chr2:19304933 in **b**, and chr5:2838468 in **c** and **g**). LD was measured with r^2 and is represented with
606 a color gradient going from blue (low LD) to red (high LD). **d-h**: Distribution of soil variables of the
607 two alternative IGE alleles ("+ +" in red, "- -" in grey) at the top IGE SNPs associated with soil
608 organic carbon content (log transformed, **d** & **e**), and the amount of water extractable by plants in the
609 soil (**h**). The local environments of the non-Relicts from North Sweden and the Relicts are reported
610 below the distributions with blue and red dots, respectively. Environmental variables were compared
611 between the two IGE alleles using a two-sided F-test (**d**: $F_{1,904} = 157{,}37$, $p < 2.2e\text{-}16$; **e**: $F_{1,904} =$
612 $190{,}83$, $p < 2.2e\text{-}16$; **h**: $F_{1,853} = 15{,}33$, $p = 9.754e\text{-}05$).

613 **Fig. 6: Association between IGE loci and plant growth. a**: Relationship between single plant (i.e.,
614 grown without neighbours) biomass and allelic composition at the eleven IGE loci. Allelic
615 composition is characterized by the number of loci bearing a positive IGE allele (going from 0, when
616 all loci bear negative IGE alleles, to 11 when all loci bear positive IGE alleles). All accessions are
617 inbred lines, i.e., fully homozygous, and all SNPs are bi-allelic. Points and error bars represent means
618 ± standard deviation. The reported p-value refers to the simple linear regression between biomass and
619 allelic composition (two-sided F-test, $F_{1,81} = 30.32$, $p = 4.207e\text{-}07$). The number of accessions per

allelic composition is given. **b-f**: Zoomed GWAS peaks (cf Fig. 2a) reporting the *p*-values (-log10 transformed) for the association test between IGE on plant biomass and allelic variation around significant IGE SNPs. SNPs with significant associations using a 5% False-Discovery Rate (FDR) threshold are colored in red. Black boxes below the plots represent the annotated genes in the region (positive strand genes above, negative strand genes below). Genes related to light response, photomorphogenesis, and shade-avoidance are represented with colored points. **g**: Allele frequency differentiation between the non-Relicts from North Sweden and the other groups. Differentiation was measured with F_{ST} computed on gene sequences. Grey distributions represent F_{ST} values on all annotated genes. Vertical black bars represent the median of the F_{ST} distribution over all genes. The upper 10th percentile is colored in red. Values for individual genes related to light responses are plotted with colored points below each distribution. WEU: Western Europe, SSW: South Sweden, SPA: Spain, REC: Relicts, IBC: Italy-Balkans-Caucasus, GER: Germany, CEU: Central Europe, ASI: Asia. **h:** Distribution of F_{ST} values at the eleven light-related genes. The red curve represents F_{ST} values between the non-Relicts from North Sweden and the Relicts, whereas the grey distribution represents F_{ST} values between the non-Relicts from North Sweden and all other non-Relicts. Dashed lines represent the medians of the distributions, and the *p*-value refers to the two-sided Welch *t*-test comparing the means of the two distributions ($t_{12.64} = 2.91$, $p = 0.013$).

References

1. Wisz, M. S. *et al.* The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* **88**, 15–30 (2013).

2. Gaüzère, P. *et al.* The diversity of biotic interactions complements functional and phylogenetic facets of biodiversity. *Current Biology* (2022) doi:10.1016/j.cub.2022.03.009.

3. Whitham, T. G. *et al.* Community and ecosystem genetics: a consequence of the extended phenotype. *Ecology* **84**, 559–573 (2003).

4. Wolf, J. B., Brodie III, E. D., Cheverud, J. M., Moore, A. J. & Wade, M. J. Evolutionary consequences of indirect genetic effects. *Trends in Ecology & Evolution* **13**, 64–69 (1998).

646 5. Griffing, B. Selection in reference to biological groups. I. Individual and group selection applied to
647 populations of unordered groups. *Australian Journal of Biological Sciences* **20**, 127–139 (1967).

648 6. Moore, A. J., Brodie III, E. D. & Wolf, J. B. Interacting phenotypes and the evolutionary process:
649 I. Direct and indirect genetic effects of social interactions. *Evolution* **51**, 1352–1362 (1997).

650 7. Griffing, B. Selection for populations of interacting genotypes. in *Proceedings of the International*
651 *Congress on Quantitative Genetics, August 16–21, 1976* pp 413-434 (Iowa State University Press,
652 1977).

653 8. Bijma, P., Muir, W. M., Ellen, E. D., Wolf, J. B. & Van Arendonk, J. A. M. Multilevel selection 2:
654 Estimating the genetic parameters determining inheritance and response to selection. *Genetics* **175**,
655 289–299 (2007).

656 9. Jennings, P. R. & de Jesus, J. Studies on competition in rice I. Competition in mixtures of varieties.
657 *Evolution* **22**, 119–124 (1968).

658 10. Hamblin, J. Effect of environment, seed size and competitive ability on yield and survival of
659 *Phaseolus vulgaris* (L.) genotypes in mixtures. *Euphytica* **24**, 435–445 (1975).

660 11. Kawano, K. & Thung, M. D. Intergenotypic competition and competition with associated
661 crops in Cassava. *Crop Science* **22**, cropsci1982.0011183X002200010013x (1982).

662 12. Murphy, G. P., Swanton, C. J., Acker, R. C. V. & Dudley, S. A. Kin recognition, multilevel
663 selection and altruism in crop sustainability. *Journal of Ecology* **105**, 930–934 (2017).

664 13. Montazeaud, G. *et al.* Farming plant cooperation in crops. *Proceedings of the Royal Society*
665 *B: Biological Sciences* **287**, 20191290 (2020).

666 14. Muir, W. M. Group selection for adaptation to multiple-hen cages: selection program and
667 direct responses. *Poultry Science* **75**, 447–458 (1996).

668 15. Wade, M. J., Bijma, P., Ellen, E. D. & Muir, W. Group selection and social evolution in
669 domesticated animals. *Evolutionary Applications* **3**, 453–465 (2010).

16. Muir, W. M. Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* **170**, 1247–1259 (2005).

17. Bergsma, R., Kanis, E., Knol, E. F. & Bijma, P. The contribution of social effects to heritable variation in finishing traits of domestic pigs (*Sus scrofa*). *Genetics* **178**, 1559–1570 (2008).

18. Alemu, S. W., Bijma, P., Møller, S. H., Janss, L. & Berg, P. Indirect genetic effects contribute substantially to heritable variation in aggression-related traits in group-housed mink (*Neovison vison*). *Genetics Selection Evolution* **46**, 30 (2014).

19. Ellen, E. D., Visscher, J., van Arendonk, J. a. M. & Bijma, P. Survival of laying hens: Genetic parameters for direct and associative effects in three purebred layer lines. *Poultry Science* **87**, 233–239 (2008).

20. Baud, A. *et al.* Dissecting indirect genetic effects from peers in laboratory mice. *Genome Biology* **22**, 216 (2021).

21. Santostefano, F., Wilson, A. J., Niemelä, P. T. & Dingemanse, N. J. Indirect genetic effects: a key component of the genetic architecture of behaviour. *Scientific Reports* **7**, 10235 (2017).

22. Xia, C., Canela-Xandri, O., Rawlik, K. & Tenesa, A. Evidence of horizontal indirect genetic effects in humans. *Nature Human Behaviour* **5**, 399–406 (2021).

23. Tilman, D. Resource Competition and Community Structure.(MPB-17), Volume 17. in *Resource Competition and Community Structure.(MPB-17), Volume 17* (Princeton university press, 2020).

24. Callaway, R. M. Positive interactions among plants. *Botanical Review* **61**, 306–349 (1995).

25. Bertness, M. D. & Callaway, R. Positive interactions in communities. *Trends in Ecology & Evolution* **9**, 191–193 (1994).

26. Hierro, J. L. & Callaway, R. M. The ecological importance of allelopathy. *Annual Review of Ecology, Evolution, and Systematics* **52**, 25–45 (2021).

694 27. Anten, N. P. R. & Chen, B. J. W. Detect thy family: Mechanisms, ecology and agricultural
695 aspects of kin recognition in plants. *Plant, Cell & Environment* **44**, 1059–1071 (2021).

696 28. Crepy, M. A. & Casal, J. J. Photoreceptor-mediated kin recognition in plants. *New*
697 *Phytologist* **205**, 329–338 (2015).

698 29. Bhatt, M. V., Khandelwal, A. & Dudley, S. A. Kin recognition, not competitive interactions,
699 predicts root allocation in young *Cakile edentula* seedling pairs. *New Phytologist* **189**, 1135–1142
700 (2011).

701 30. Dudley, S. A., Murphy, G. P. & File, A. L. Kin recognition and competition in plants.
702 *Functional Ecology* **27**, 898–906 (2013).

703 31. Dudley, S. A. Plant cooperation. *AoB PLANTS* **7**, plv113 (2015).

704 32. Cappa, E. P. & Cantet, R. J. C. Direct and competition additive effects in tree breeding:
705 Bayesian estimation from an individual tree mixed model. *Silvae Genetica* **57**, 45–56 (2008).

706 33. Costa e Silva, J. & Kerr, R. J. Accounting for competition in genetic analysis, with particular
707 emphasis on forest genetic trials. *Tree Genetics & Genomes* **9**, 1–17 (2013).

708 34. Mutic, J. J. & Wolf, J. B. Indirect genetic effects from ecological interactions in *Arabidopsis*
709 *thaliana*. *Molecular Ecology* **16**, 2371–2381 (2007).

710 35. Andersson, S. Indirect genetic effects from competition in the clonal herb *Sedum album*
711 (Crassulaceae). *PLOS ONE* **9**, e106104 (2014).

712 36. Subrahmaniam, H. J. *et al.* The genetics underlying natural variation of plant-plant
713 interactions, a beloved but forgotten member of the family of biotic interactions. *The Plant Journal*
714 **93**, 747–770 (2018).

715 37. Becker, C. *et al.* The ecologically relevant genetics of plant–plant interactions. *Trends in*
716 *Plant Science* **0**, (2022).

717 38. Takou, M. *et al.* Linking genes with ecological strategies in *Arabidopsis thaliana*. *Journal of*
718 *Experimental Botany* **70**, 1141–1151 (2019).

719 39. Horton, M. W. *et al.* Genome-wide patterns of genetic variation in worldwide *Arabidopsis*
720 *thaliana* accessions from the RegMap panel. *Nature Genetics* **44**, 212–216 (2012).

721 40. 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in
722 *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).

723 41. Lee, C.-R. *et al.* On the post-glacial spread of human commensal *Arabidopsis thaliana*.
724 *Nature Communications* **8**, 14458 (2017).

725 42. Exposito-Alonso, M. *et al.* Genomic basis and evolutionary potential for extreme drought
726 adaptation in *Arabidopsis thaliana*. *Nature Ecology & Evolution* **2**, 352–358 (2018).

727 43. Toledo, B., Marcer, A., Méndez-Vigo, B., Alonso-Blanco, C. & Picó, F. X. An ecological
728 history of the relict genetic lineage of *Arabidopsis thaliana*. *Environmental and Experimental*
729 *Botany* **170**, 103800 (2020).

730 44. Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R. & Weigel, D. Natural
731 selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* **573**, 126–129
732 (2019).

733 45. Hancock, A. M. *et al.* Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*
734 **334**, 83–86 (2011).

735 46. Vasseur, F. *et al.* Climate as a driver of adaptive variations in ecological strategies in
736 *Arabidopsis thaliana*. *Annals of Botany* **122**, 935–945 (2018).

737 47. Sartori, K. *et al.* Leaf economics and slow-fast adaptation across the geographic range of
738 *Arabidopsis thaliana*. *Scientific Reports* **9**, 10758 (2019).

739 48. Estarague, A. *et al.* Into the range: a latitudinal gradient or a center-margins differentiation of
740 ecological strategies in *Arabidopsis thaliana*? *Annals of Botany* **129**, 343–356 (2022).

741 49. Wuest, S. E. *et al.* Increasing plant group productivity through latent genetic variation for
742 cooperation. *PLOS Biology* **20**, e3001842 (2022).

743 50. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

744 51. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA–BABA statistics to
745 locate introgressed loci. *Molecular Biology and Evolution* **32**, 244–257 (2015).

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

748 53. Grime, J. P. Evidence for the existence of three primary strategies in plants and its relevance
749 to ecological and evolutionary theory. *The American Naturalist* **111**, 1169–1194 (1977).

750 54. Grime, J. P. & Mackey, J. The role of plasticity in resource capture by plants. *Evolutionary*
751 *Ecology* **16**, 299–307 (2002).

752 55. Campbell, B. D., Grime, J. P. & Mackey, J. M. L. A trade-off between scale and precision in
753 resource foraging. *Oecologia* **87**, 532–538 (1991).

754 56. Rivas-San Vicente, M. & Plasencia, J. Salicylic acid beyond defence: its role in plant growth
755 and development. *Journal of Experimental Botany* **62**, 3321–3338 (2011).

756 57. Koo, Y. M., Heo, A. Y. & Choi, H. W. Salicylic acid as a safe plant protector and growth
757 regulator. *The Plant Pathology Journal* **36**, 1–10 (2020).

758 58. Nozue, K. *et al.* Network analysis reveals a role for salicylic acid pathway components in
759 shade avoidance. *Plant Physiology* **178**, 1720–1732 (2018).

760 59. Yang, C. & Li, L. Hormonal regulation in shade avoidance. *Frontiers in Plant Science* **8**,
761 (2017).

762 60. Casal, J. J. Photoreceptor signaling networks in plant responses to shade. *Annual Review of*
763 *Plant Biology* **64**, 403–427 (2013).

764 61. Brotherstone, S. *et al.* Competition effects in a young Sitka spruce (*Picea sitchensis*, Bong.
765 Carr) clonal trial. *Silvae Genetica* **60**, 149–155 (2011).

766 62. Baron, E., Richirt, J., Villoutreix, R., Amsellem, L. & Roux, F. The genetics of intra- and
767 interspecific competitive response and effect in a local population of an annual plant species.
768 *Functional Ecology* **29**, 1361–1370 (2015).

769 63. Hartfield, M. & Glémin, S. Hitchhiking of deleterious alleles and the cost of adaptation in
770 partially selfing species. *Genetics* **196**, 281–293 (2014).

771 64. Filiault, D. L. *et al.* Amino acid polymorphisms in *Arabidopsis* phytochrome B cause
772 differential responses to light. *Proceedings of the National Academy of Sciences of the United*
773 *States of America* **105**, 3157–3162 (2008).

774 65. Balasubramanian, S. *et al.* The PHYTOCHROME C photoreceptor gene mediates natural
775 variation in flowering and growth responses of *Arabidopsis thaliana*. *Nature Genetics* **38**, 711–715
776 (2006).

777 66. Maloof, J. N. *et al.* Natural variation in light sensitivity of *Arabidopsis*. *Nature Genetics* **29**,
778 441–446 (2001).

779 67. Stenøien, H. K., Fenster, C. B., Kuittinen, H. & Savolainen, O. Quantifying latitudinal clines
780 to light responses in natural populations of *Arabidopsis thaliana* (Brassicaceae). *American Journal*
781 *of Botany* **89**, 1604–1608 (2002).

782 68. Hamilton, W. D. The genetical evolution of social behaviour. I & II. *Journal of Theoretical*
783 *Biology* **7**, 1–52 (1964).

784 69. Bijma, P. The quantitative genetics of indirect genetic effects: a selective review of modelling
785 issues. *Heredity* **112**, 61–69 (2014).

786 70. Bijma, P. A general definition of the heritable variation that determines the potential of a
787 population to respond to selection. *Genetics* **189**, 1347–1359 (2011).

788 71. Shen, X. *et al.* Natural CMT2 variation is associated with genome-wide methylation changes
789 and temperature seasonality. *PLOS Genetics* **10**, e1004842 (2014).

790 72. Brachi, B. *et al.* Coselected genes determine adaptive variation in herbivore resistance
791 throughout the native range of *Arabidopsis thaliana*. *Proceedings of the National Academy of*
792 *Sciences* **112**, 4032–4037 (2015).

793 73. Brachi, B. *et al.* Investigation of the geographical scale of adaptive phenological variation and
794 its underlying genetics in *Arabidopsis thaliana*. *Molecular Ecology* **22**, 4222–4240 (2013).

795 74. Gloss, A. D. *et al.* Genome-wide association mapping within a local *Arabidopsis thaliana*
796 population more fully reveals the genetic architecture for defensive metabolite diversity.
797 *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**, 20200512 (2022).

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

800 76. R Core Team. R: a language and environment for statistical computing. *R Foundation for*
801 *Statistical Computing* (2019).

802 77. Butler, D., Cullis, B., Gilmour, A., Gogel, B. & Thompson, R. ASReml-R reference manual
803 version 4. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK* (2017).

804 78. Vilhjálmsson, B. J. & Nordborg, M. The nature of confounding in genome-wide association
805 studies. *Nature Reviews Genetics* **14**, 1–2 (2013).

806 79. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: a practical and powerful
807 approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*
808 **57**, 289–300 (1995).

809 80. Clayton, D. snpStats: SnpMatrix and XSnpMatrix classes and methods. *R package version*
810 *1.50.0.*

811 81. Whiting, J. R. JimWhiting91/genotype_plot: Genotype Plot. (2022)
812 doi:10.5281/zenodo.5913504.

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

815 83. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based
816 linkage analyses. *American Journal of Human Genetics* **81**, 559–575 (2007).

817 84. Hill, W. G. & Weir, B. S. Variances and covariances of squared linkage disequilibria in finite
818 populations. *Theoretical Population Biology* **33**, 54–78 (1988).

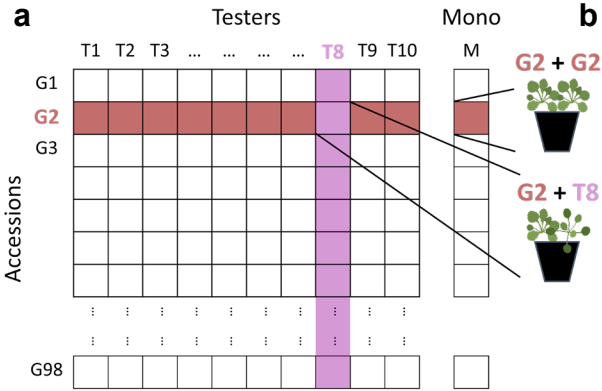
819 85. European Commission. Global Land Cover 2000 database. (2003).

820 86. Obenchain, V. *et al.* VariantAnnotation : a Bioconductor package for exploration and
821 annotation of genetic variants. *Bioinformatics* **30**, 2076–2078 (2014).

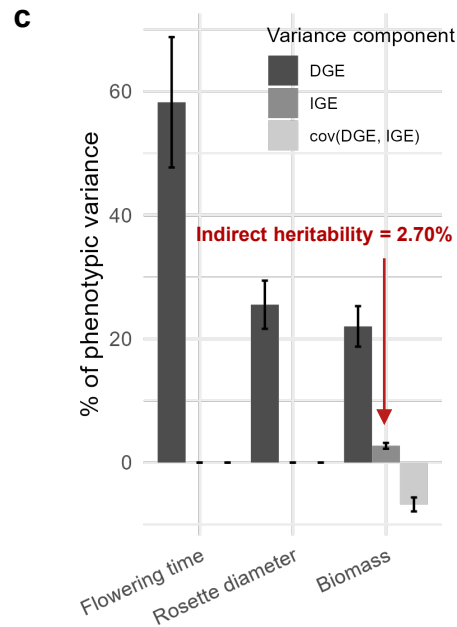
822 87. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the analysis of population
823 structure. *Evolution* **38**, 1358–1370 (1984).

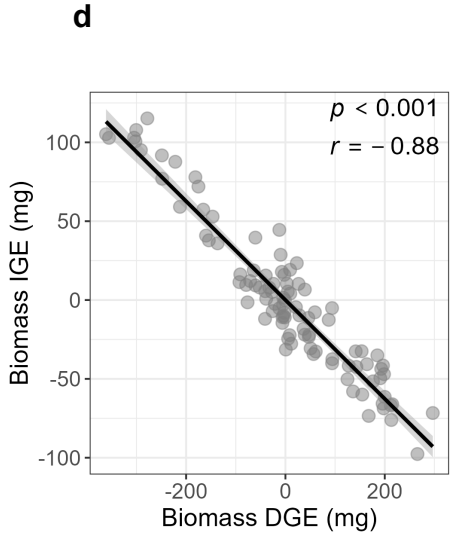
824 88. De Mita, S. & Siol, M. EggLib: Processing, analysis and simulation tools for population
825 genetics and genomics. *BMC Genetics* **13**, 27 (2012).

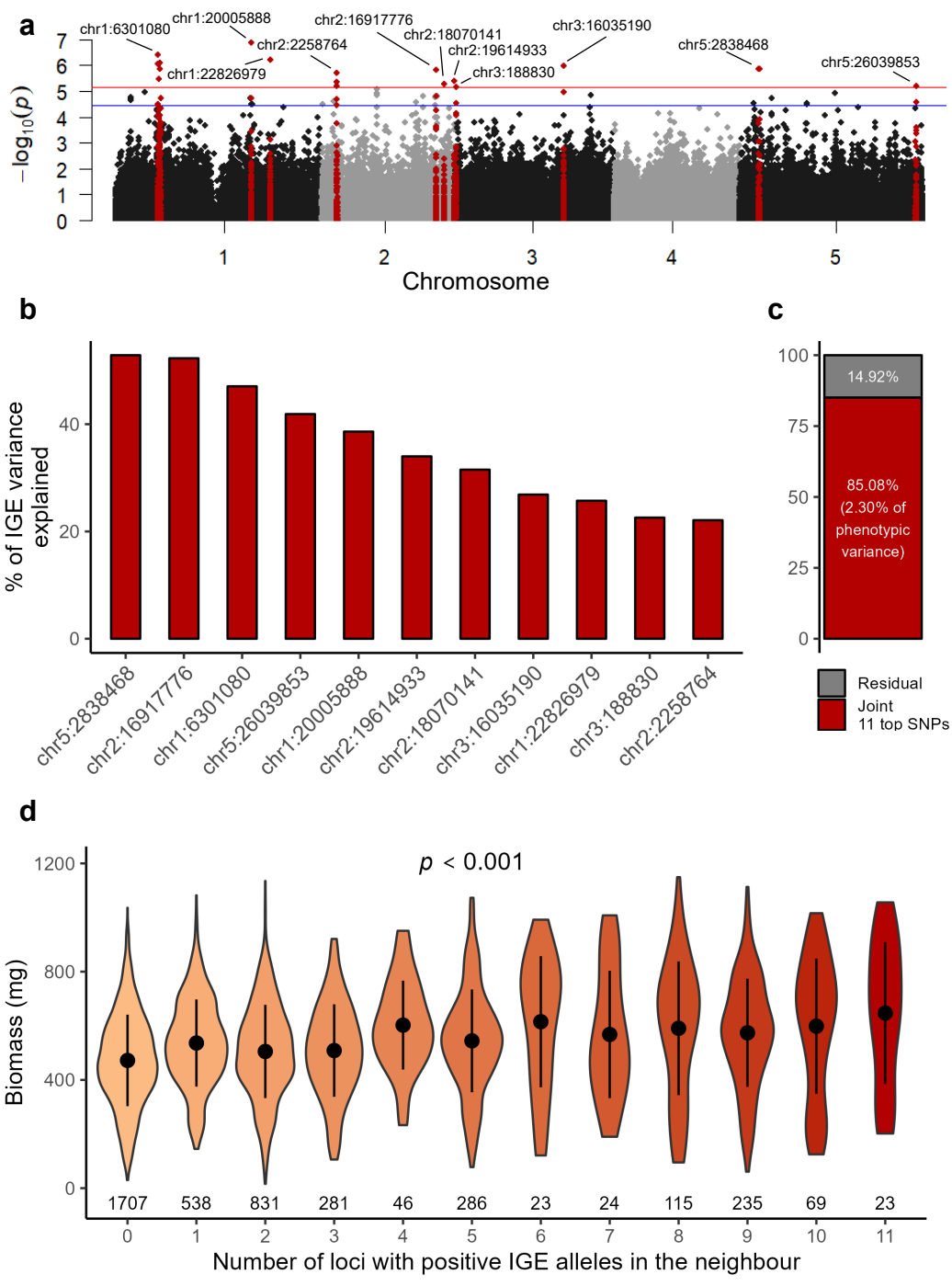
MAIN FIGURES

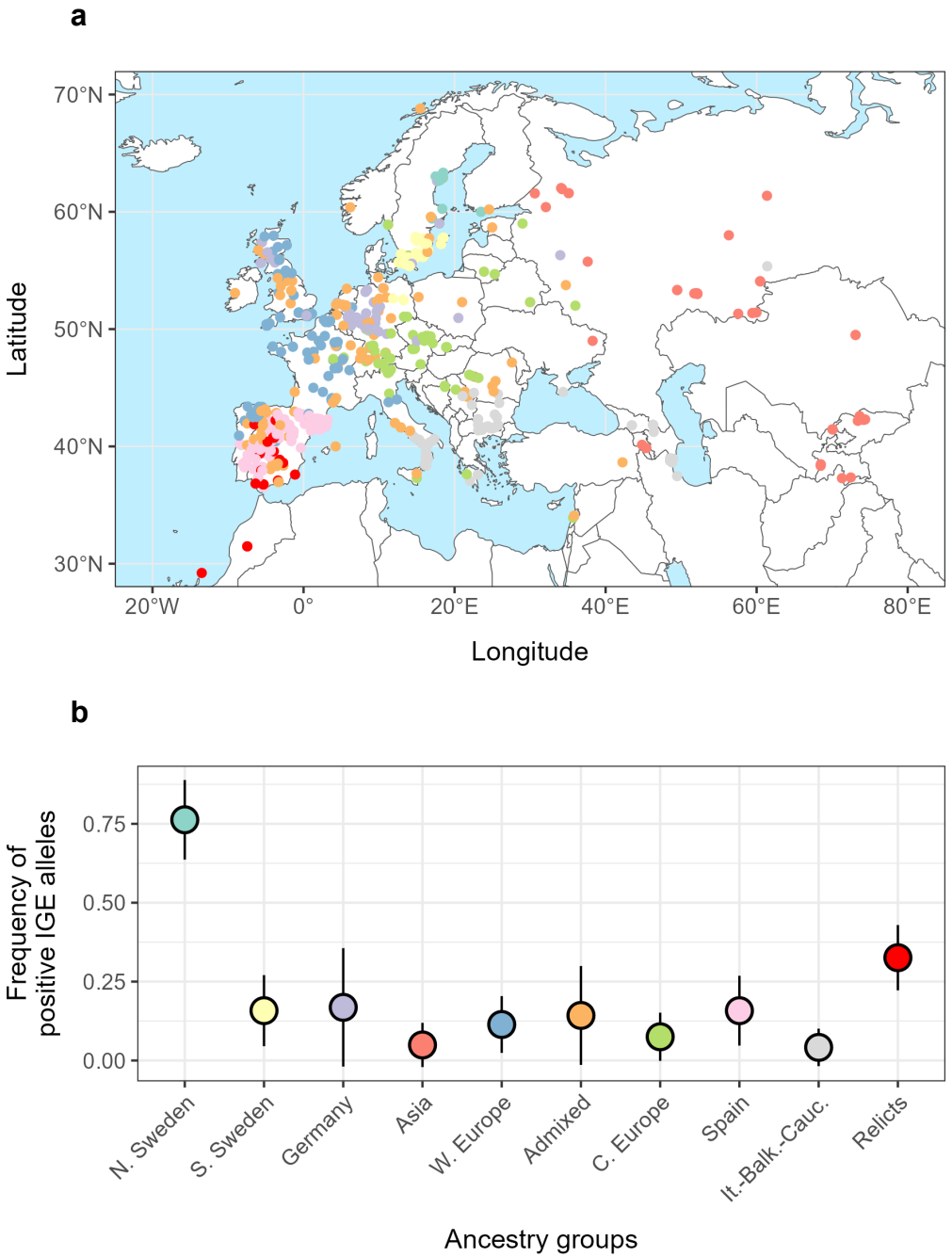

a
Testers
Mono
T1
T2
T3
...
...
...
...
T8
T9
T10
M
G1
G2
G3
G98
Accessions
G2 + G2
G2 + T8

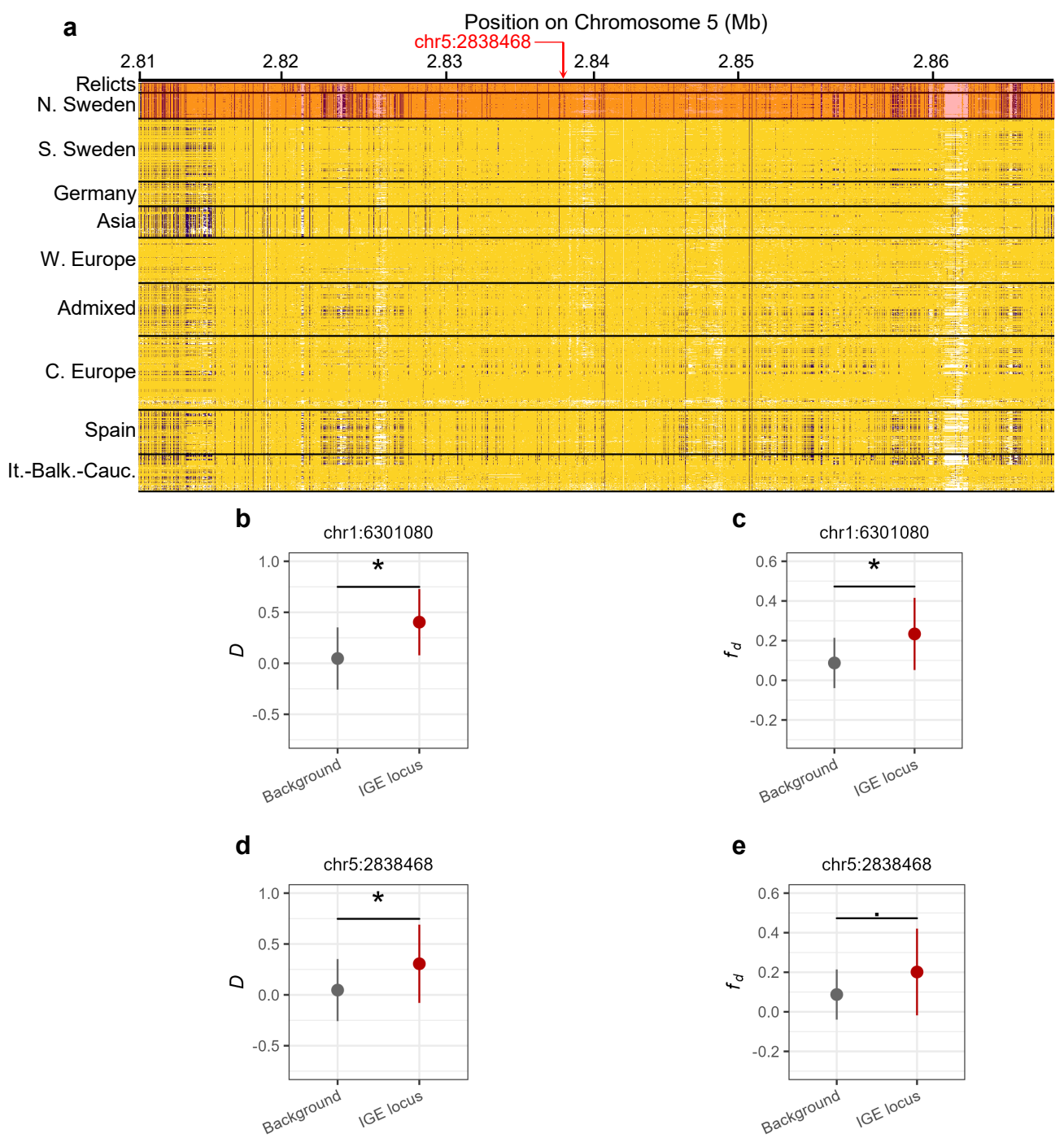

b

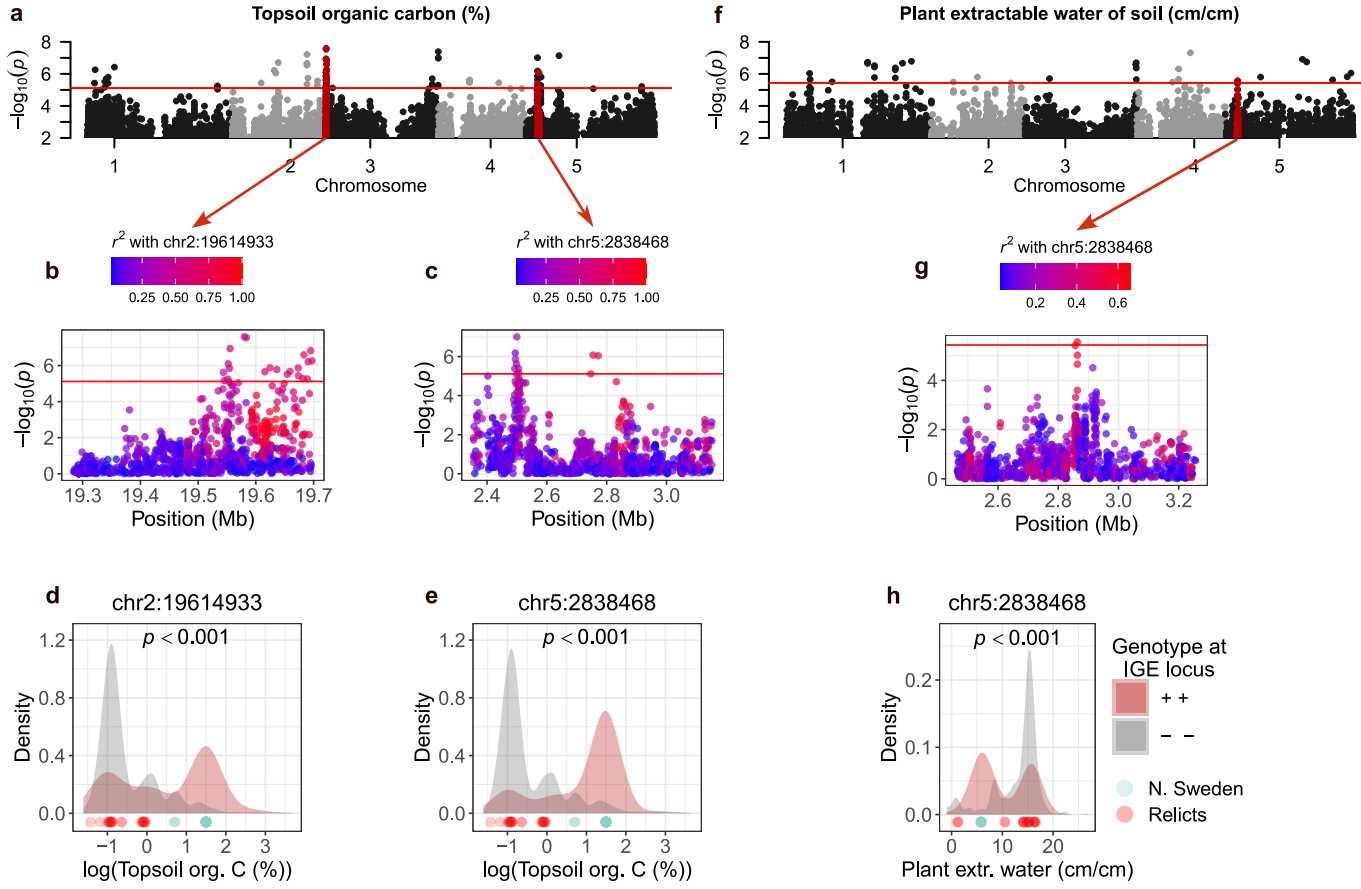

c
Variance component
DGE
IGE
cov(DGE, IGE)
Indirect heritability = 2.70%
% of phenotypic variance
60
40
20
0
Flowering time
Rosette diameter
Biomass

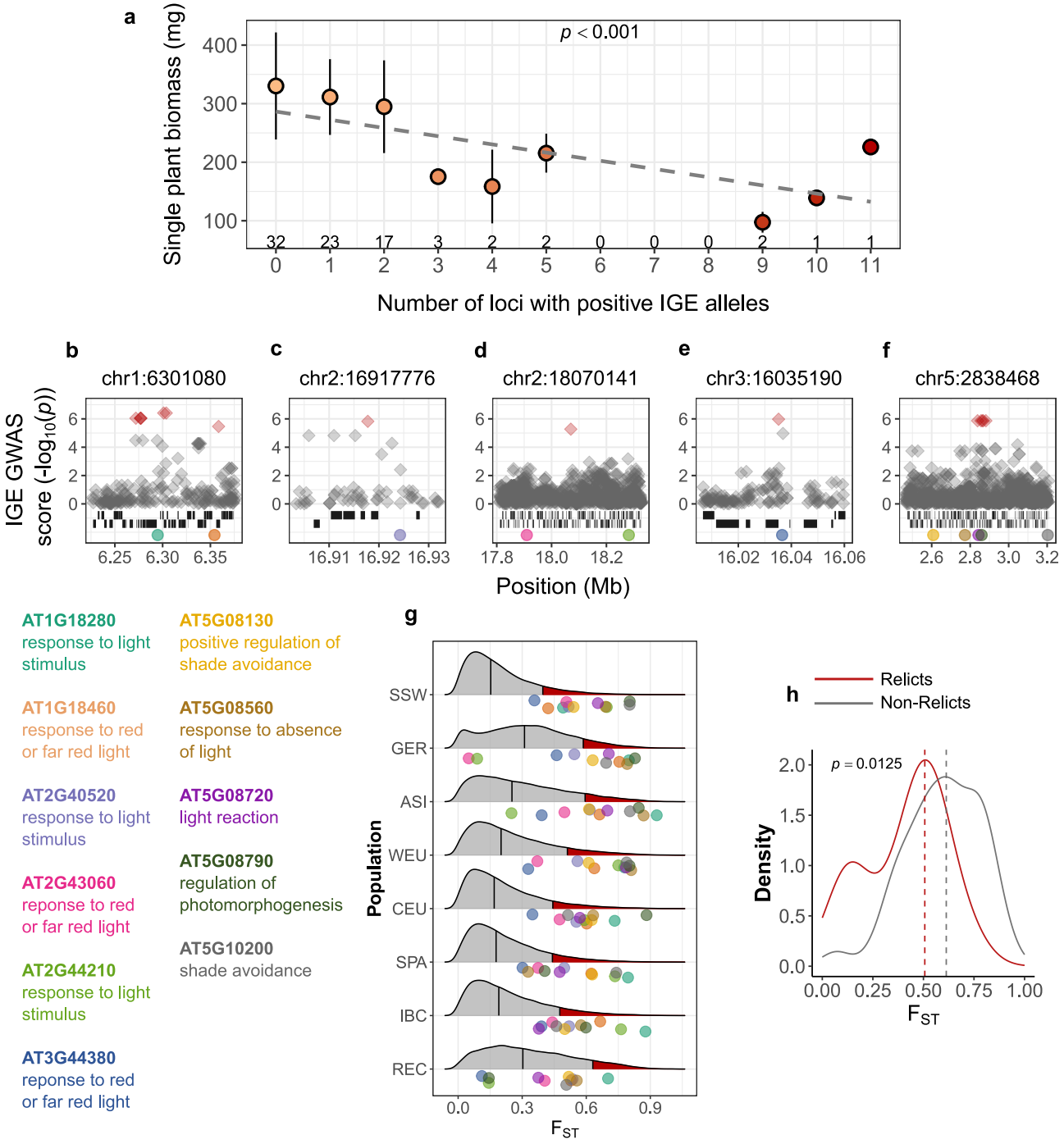

d
p < 0.001
r = − 0.88
Biomass IGE (mg)
100
50
0
-50
-100
-200
0
200
Biomass DGE (mg)


a
chr1:6301080
chr1:20005888
chr2:2258764
chr1:22826979
chr2:16917776
chr2:18070141
chr2:19614933
chr3:188830
chr3:16035190
chr5:2838468
chr5:26039853
$-\log_{10}(p)$
Chromosome
b
% of IGE variance explained
chr5:2838468
chr2:16917776
chr1:6301080
chr5:26039853
chr1:20005888
chr2:19614933
chr2:18070141
chr3:16035190
chr1:22826979
chr3:188830
chr2:2258764
c
14.92%
85.08% (2.30% of phenotypic variance)
Residual
Joint 11 top SNPs
d
$p < 0.001$
Biomass (mg)
1707 538 831 281 46 286 23 24 115 235 69 23
Number of loci with positive IGE alleles in the neighbour


a
70°N
60°N
50°N
40°N
30°N
Latitude
20°W
0°
20°E
40°E
60°E
80°E
Longitude
b
Frequency of positive IGE alleles
0.75
0.50
0.25
0.00
N. Sweden
S. Sweden
Germany
Asia
W. Europe
Admixed
C. Europe
Spain
It.-Balk.-Cauc.
Relicts
Ancestry groups


a
Position on Chromosome 5 (Mb)
chr5:2838468
2.81
2.82
2.83
2.84
2.85
2.86
Relicts
N. Sweden
S. Sweden
Germany
Asia
W. Europe
Admixed
C. Europe
Spain
It.-Balk.-Cauc.
b
chr1:6301080
D
Background
IGE locus
c
chr1:6301080
f_d
d
chr5:2838468
e
chr5:2838468


a
Topsoil organic carbon (%)
−log10(p)
Chromosome
f
Plant extractable water of soil (cm/cm)
r2 with chr2:19614933
r2 with chr5:2838468
b
c
g
Position (Mb)
d
e
h
chr2:19614933
chr5:2838468
p < 0.001
Density
log(Topsoil org. C (%))
Plant extr. water (cm/cm)
Genotype at IGE locus
+ +
− −
N. Sweden
Relicts


a
Single plant biomass (mg)
p < 0.001
400
300
200
100
32 23 17 3 2 2 0 0 0 2 1 1
0 1 2 3 4 5 6 7 8 9 10 11
Number of loci with positive IGE alleles
b
chr1:6301080
c
chr2:16917776
d
chr2:18070141
e
chr3:16035190
f
chr5:2838468
IGE GWAS score (-log10(p))
6.25 6.30 6.35
16.91 16.92 16.93
17.8 18.0 18.2
16.02 16.04 16.06
2.6 2.8 3.0 3.2
Position (Mb)
AT1G18280
response to light stimulus
AT1G18460
response to red or far red light
AT2G40520
response to light stimulus
AT2G43060
reponse to red or far red light
AT2G44210
response to light stimulus
AT3G44380
reponse to red or far red light
AT5G08130
positive regulation of shade avoidance
AT5G08560
response to absence of light
AT5G08720
light reaction
AT5G08790
regulation of photomorphogenesis
AT5G10200
shade avoidance
g
Population
SSW
GER
ASI
WEU
CEU
SPA
IBC
REC
0.0 0.3 0.6 0.9
FST
h
Relicts
Non-Relicts
p = 0.0125
Density
2.0 1.5 1.0 0.5 0.0
0.00 0.25 0.50 0.75 1.00
FST

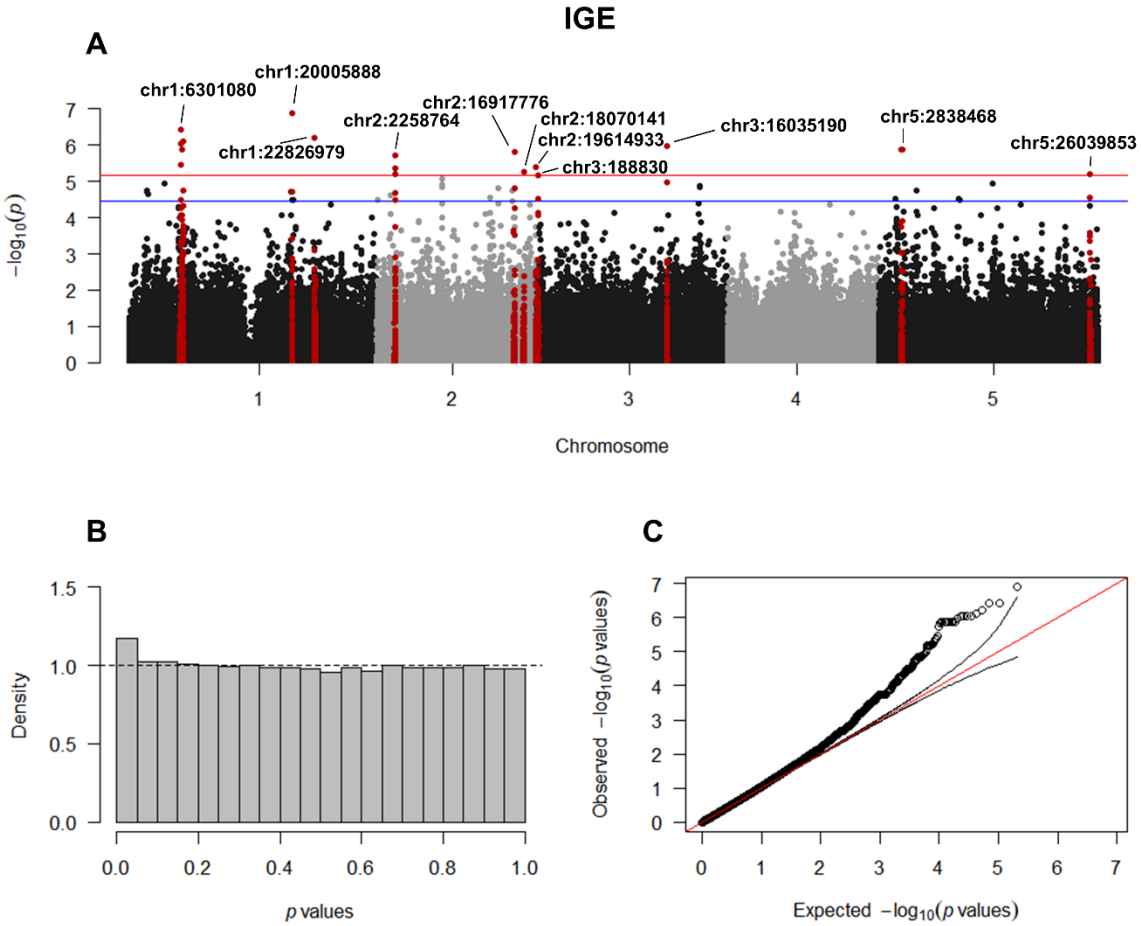
SUPPLEMENTARY INFORMATION

Supplementary Table 1: Model comparison. For each trait, we compared three models of increasing complexity. Model 1 only accounts for direct genetic effects (DGEs). Model 2 accounts for both direct and indirect genetic effects (DGEs + IGEs). Model 3 accounts for both direct and indirect genetic effects as well as for their covariance (DGEs + IGEs + cov(DGEs, IGEs)). The three models were compared using a one-sided Likelihood ratio tests. Here we report degrees of freedom (D.f.), Likelihood ratio statistics (L-R statistics), and associated *p*-values.

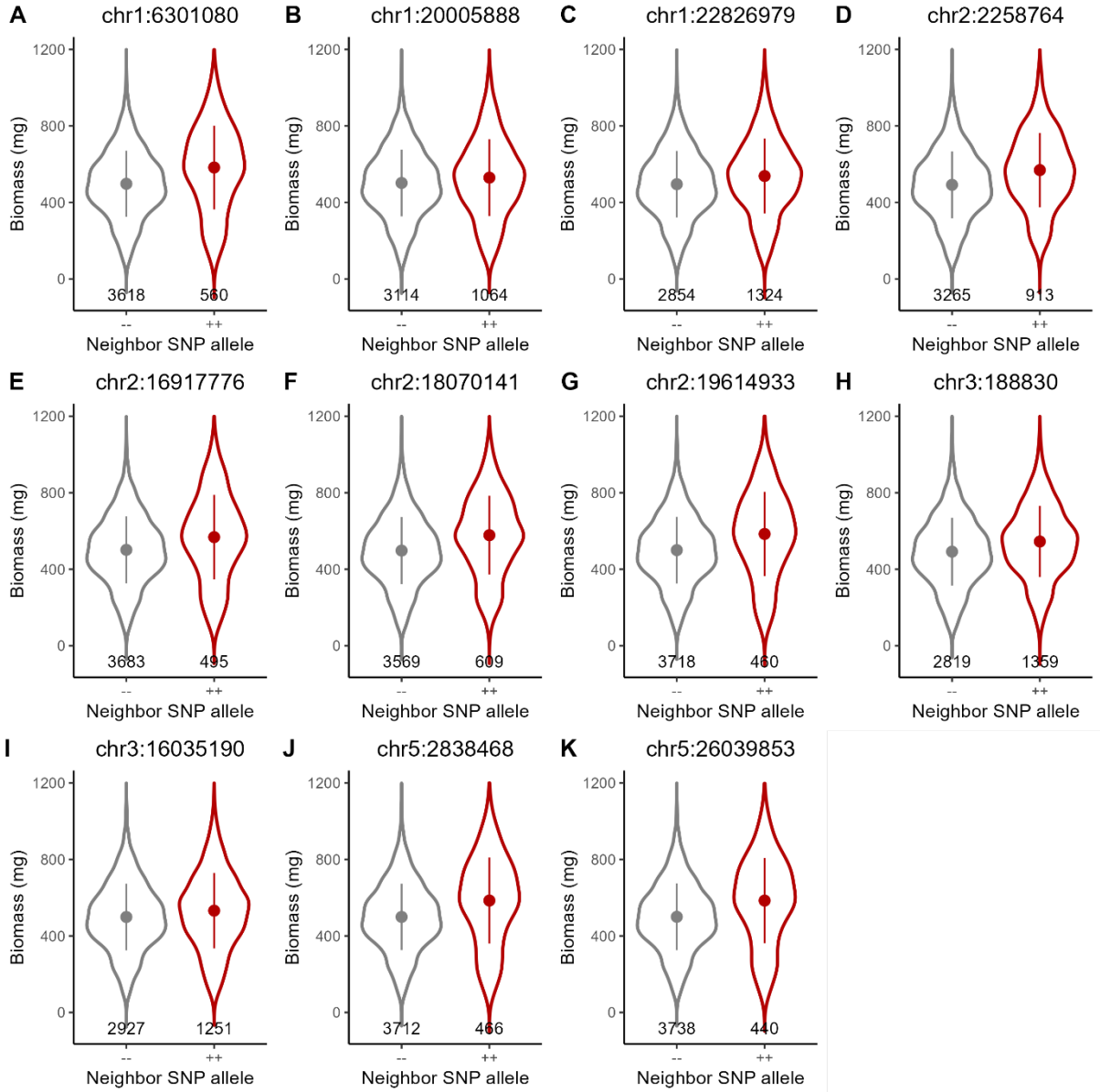
Trait	Comparison	D.f.	L-R statistic	*p*-value
Rosette diameter (cm)	Model 2 vs Model 1	1	0.19	0.3309
	Model 3 vs Model 2	1	-0.05	0.5000
Flowering time (days after sowing)	Model 2 vs Model 1	1	1.88	0.0851
	Model 3 vs Model 2	1	1.87	0.0856
Aboveground biomass (mg)	Model 2 vs Model 1	1	813.17	< 2.2e-16
	Model 3 vs Model 2	1	95.68	< 2.2e-16

Supplementary Table 2: Mixed model estimations. Variance components (± standard errors) estimated from the best model for each trait. $\widehat{\sigma}^2_{A_D}$: direct genetic variance, $\widehat{\sigma}^2_{A_S}$: indirect genetic variance, $\hat{r}_{A_{DS}}$: correlation between direct and indirect genetic effects, $\widehat{\sigma}^2_b$: variance of the block effect, $\widehat{\sigma}^2_e$: residual variance, $\widehat{\rho}_e$: correlation between the residuals of plants that share the same pot.

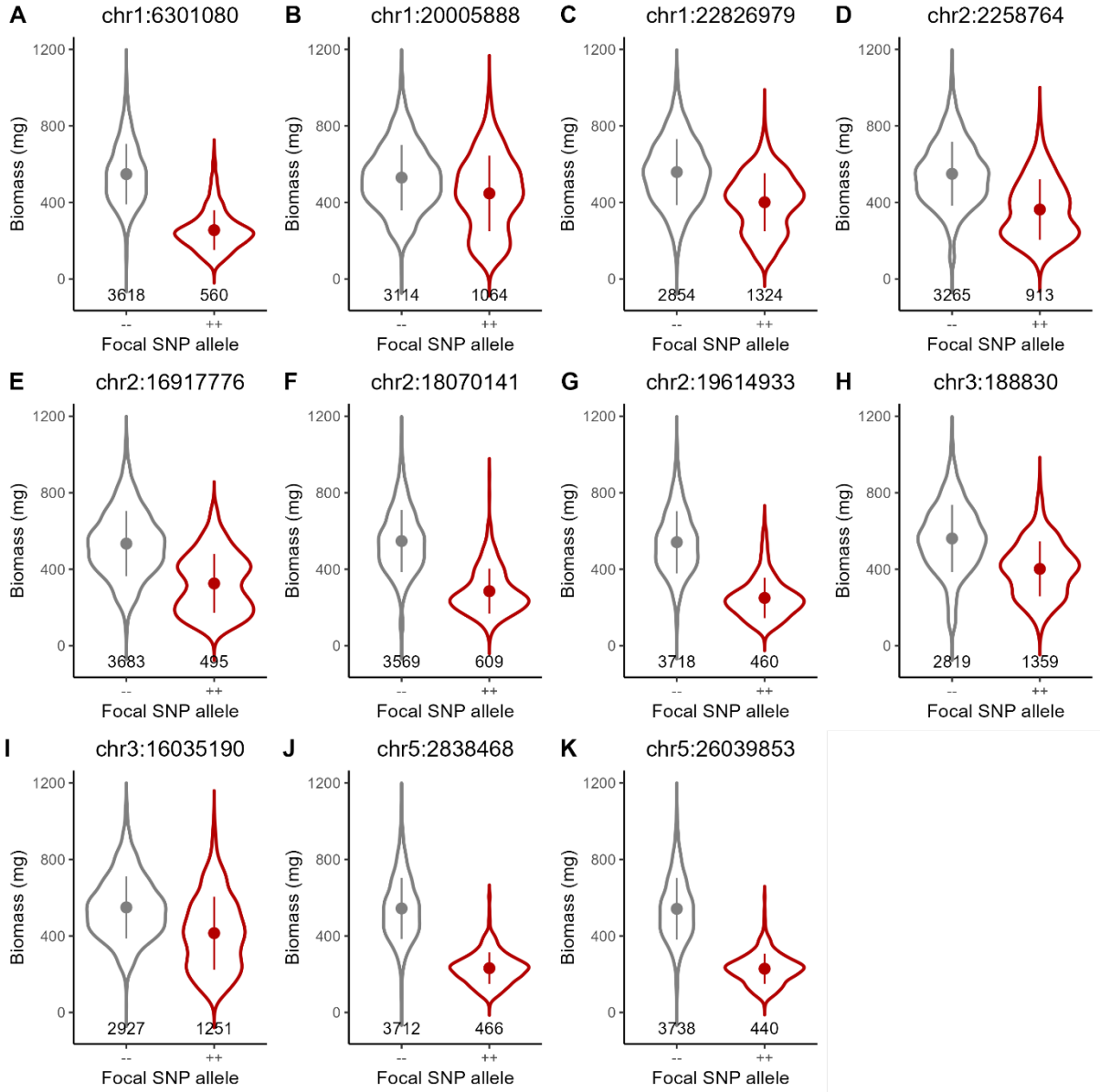
Trait	Best model	$\widehat{\sigma}^2_{A_D}$	$\widehat{\sigma}^2_{A_S}$	$\hat{r}_{A_{DS}}$	$\widehat{\sigma}^2_b$	$\widehat{\sigma}^2_e$	$\widehat{\rho}_e$
Rosette diameter (cm)	Model 1	0.39 ± 0.06	-	-	0.09 ± 0.12	0.76 ± 0.02	0.22 ± 0.02
Flowering time (days after sowing)	Model 1	22.34 ± 4.04	-	-	0.02 ± 0.03	5.35 ± 0.14	0.04 ± 0.03
Aboveground biomass (mg)	Model 3	7285.33 ± 1084.04	895.32 ± 155.60	-0.88 ± 0.03	2153.67 ± 3050.33	8285.33 ± 188.73	-0.18 ± 0.02



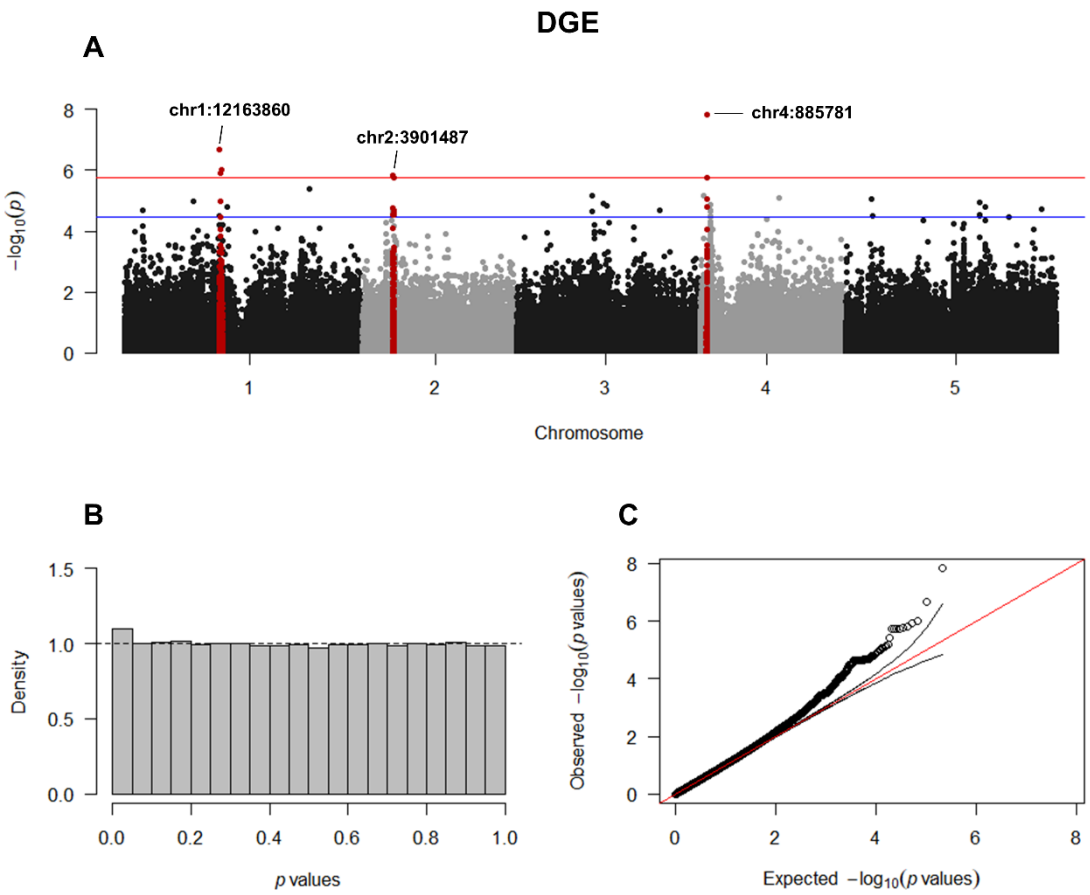
Supplementary Figure 1: Genome-Wide Association Study (GWAS) of Indirect Genetic Effects (IGEs) on plant biomass. A: Manhattan plot reporting *p*-values (-log10 transformed) for the association tests between IGE on biomass and allelic variation at 206,426 SNPs distributed along the genome of *A. thaliana*. The *p*-values correspond to a per SNP two-sided Wald-test corrected for multiple comparison using a False Discovery Rate (FDR) of 5% (10%), here represented with a solid red (blue) line. The points highlighted in red are the most significant SNPs and the surrounding SNPs at ± 300 kb, **B:** Distribution of the 206,426 *p*-values obtained with the genome-wide association test. The dotted line represents the theoretical uniform *p*-value distribution under H_0 (all SNPs effects are null), **C:** Q-Q plot representing the observed vs expected quantiles of the *p*-value distribution. Solid lines show the expected quantiles under the null hypothesis (red) and their 95% confidence interval (black).



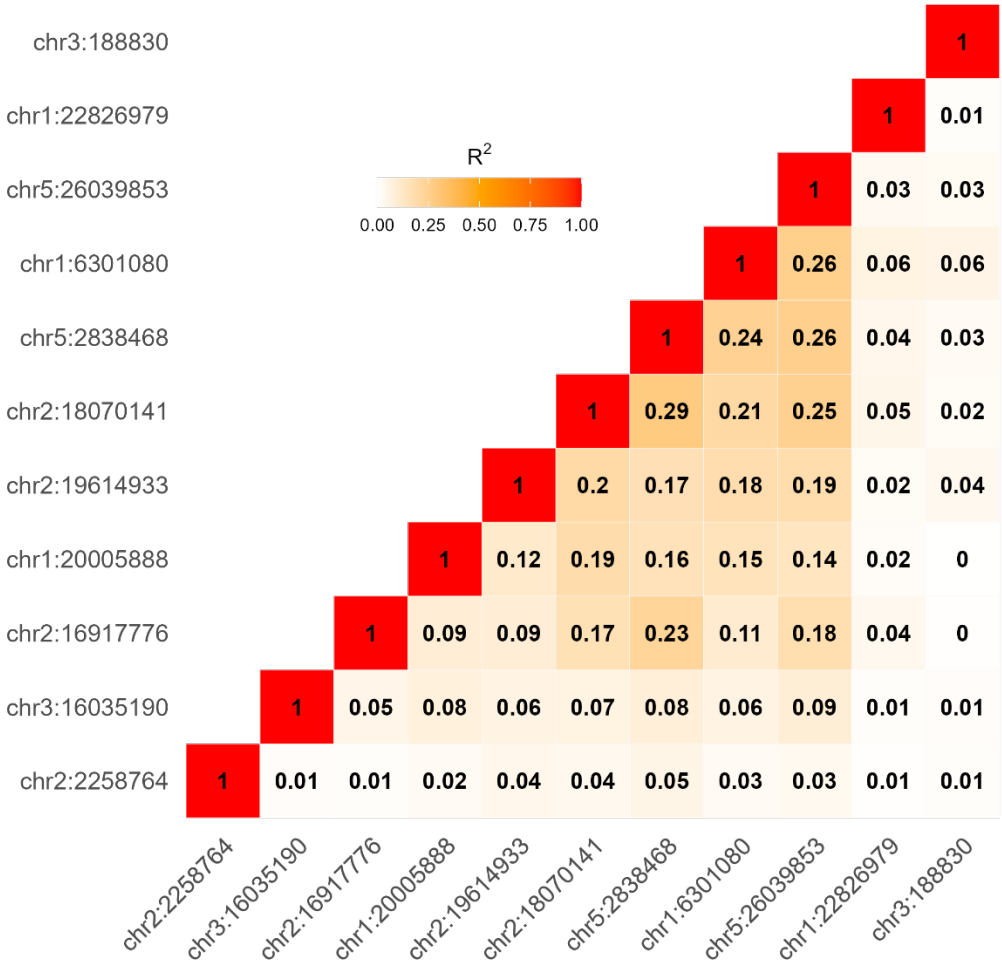
Supplementary Figure 2: Indirect Genetic Effects (IGEs) of the top 11 SNPs associated with IGEs on plant biomass. For each locus (**A-K**) biomass is compared between plants that had a neighbor bearing the negative IGE allele ("- -", grey plot) vs the positive IGE allele ("+ +", red plot). All accessions are diploid inbred lines. Points and error bars represent the mean ± standard deviation. The number of accessions in each category are reported below each violin plot.



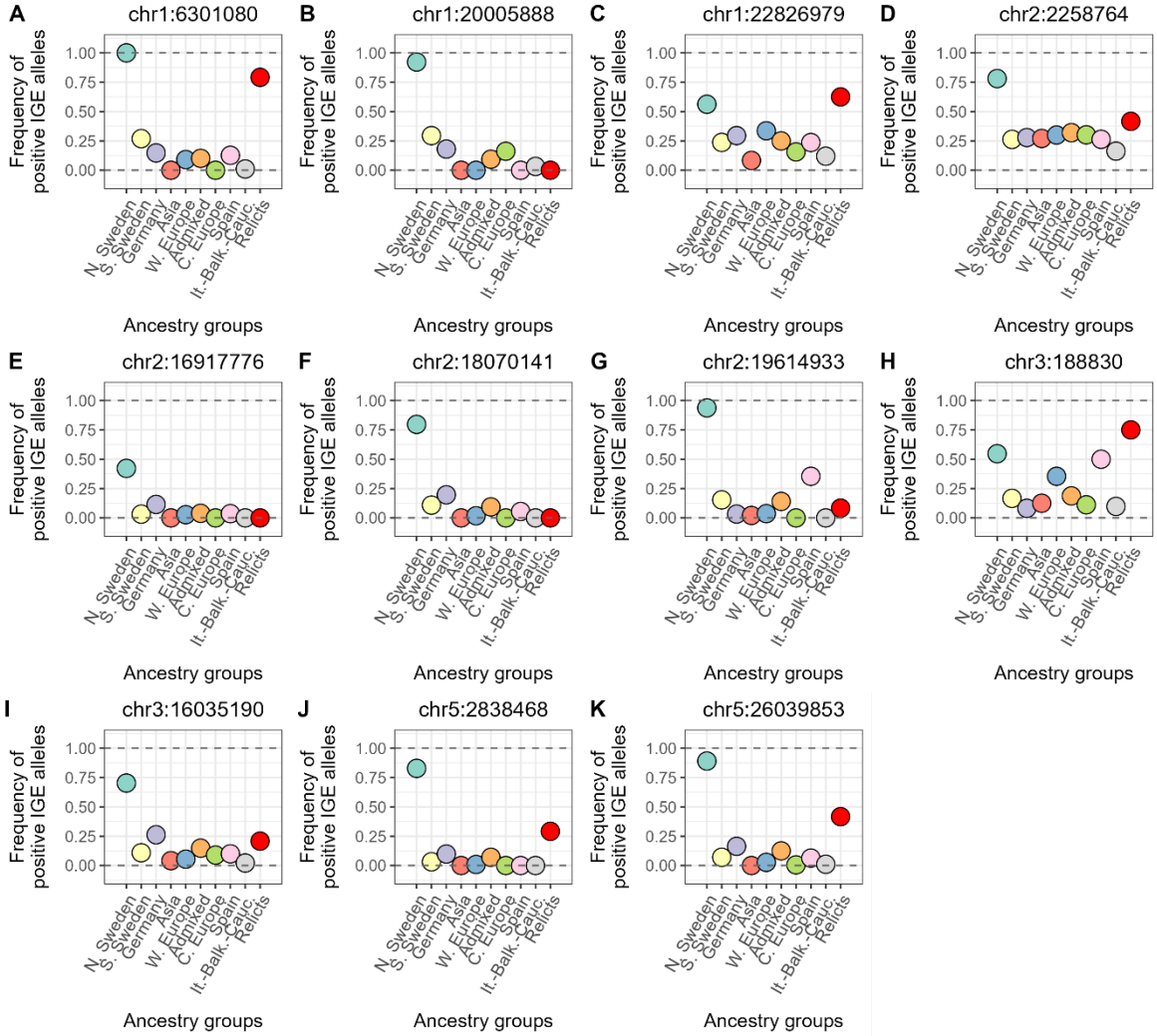
Supplementary Figure 3: Direct Genetic Effects (DGEs) of the top 11 SNPs associated with IGEs on plant biomass. For each locus (**A-K**) biomass is compared between plants bearing the negative IGE allele ("- -", grey plot) vs the positive IGE allele ("+ +", red plot). All accessions are diploid inbred lines. Points and error bars represent the mean ± standard deviation. The number of accessions in each category are reported below each violin plot.



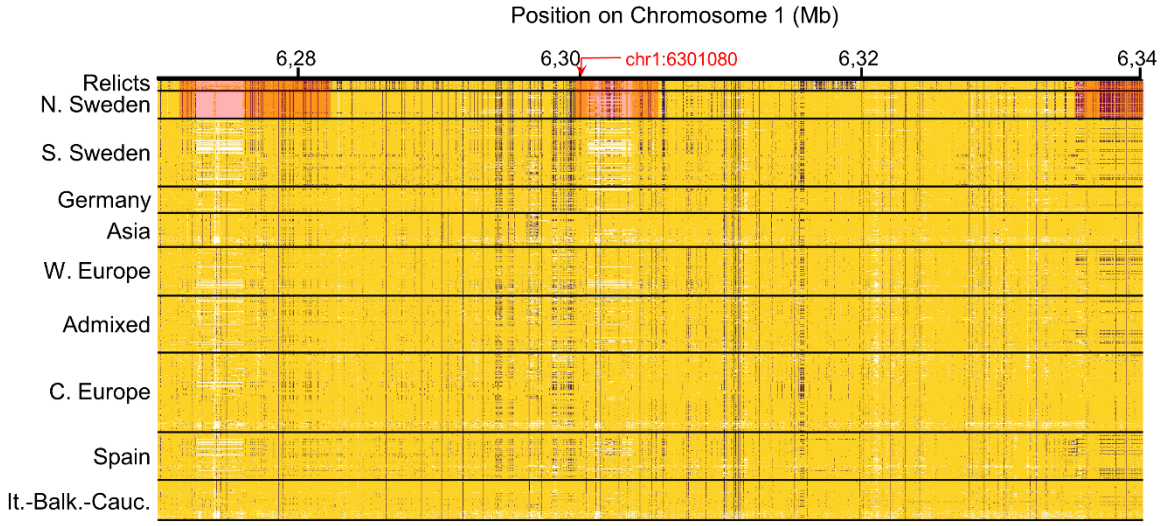
Supplementary Figure 4: Genome-Wide Association Study (GWAS) of Direct Genetic Effects (DGEs) on plant biomass. A: Manhattan plot reporting *p*-values (-log10 transformed) for the association tests between DGEs on biomass and allelic variation at 206,426 SNPs distributed along the genome of *A. thaliana*. The *p*-values correspond to a per SNP two-sided Wald-test corrected for multiple comparison using a False Discovery Rate (FDR) of 5% (10%), here represented with a solid red (blue) line. The points highlighted in red are the most significant SNPs and the surrounding SNPs at ± 300 kb, **B:** Distribution of the 206,426 *p*-values obtained with the genome-wide association test. The dotted line represents the theoretical uniform *p*-value distribution under H_0 (all SNPs effects are null), **C:** Q-Q plot representing the observed vs expected quantiles of the *p*-value distribution. Solid lines show the expected quantiles under the null hypothesis (red) and their 95% confidence interval (black).



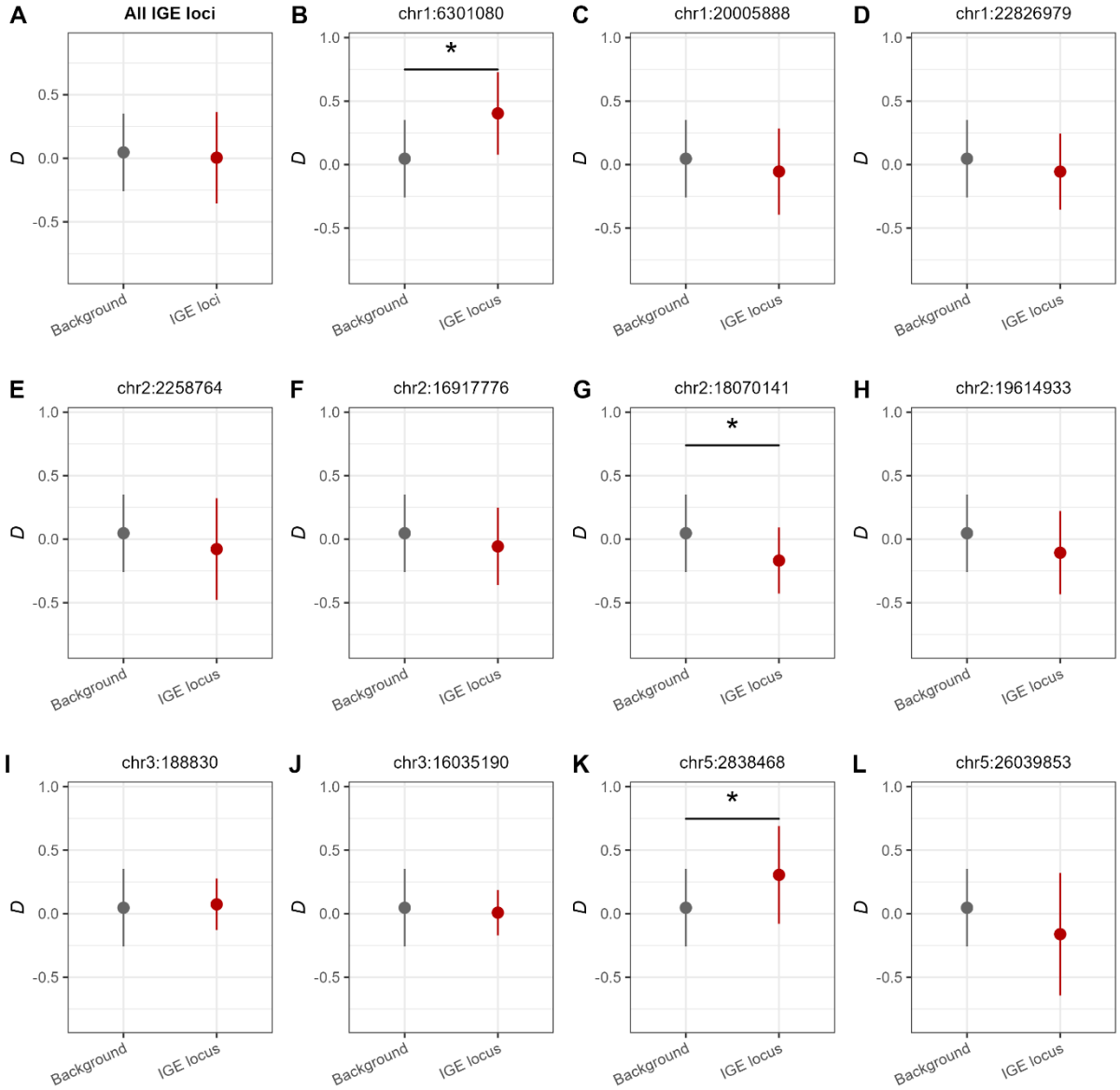
Supplementary Figure 5: Linkage Disequilibrium (LD) between IGE loci. LD was estimated with the r^2 statistic using 972 accessions from the Eurasian native range of *A. thaliana*. The color gradient goes from white to red to indicate low to high LD, respectively.



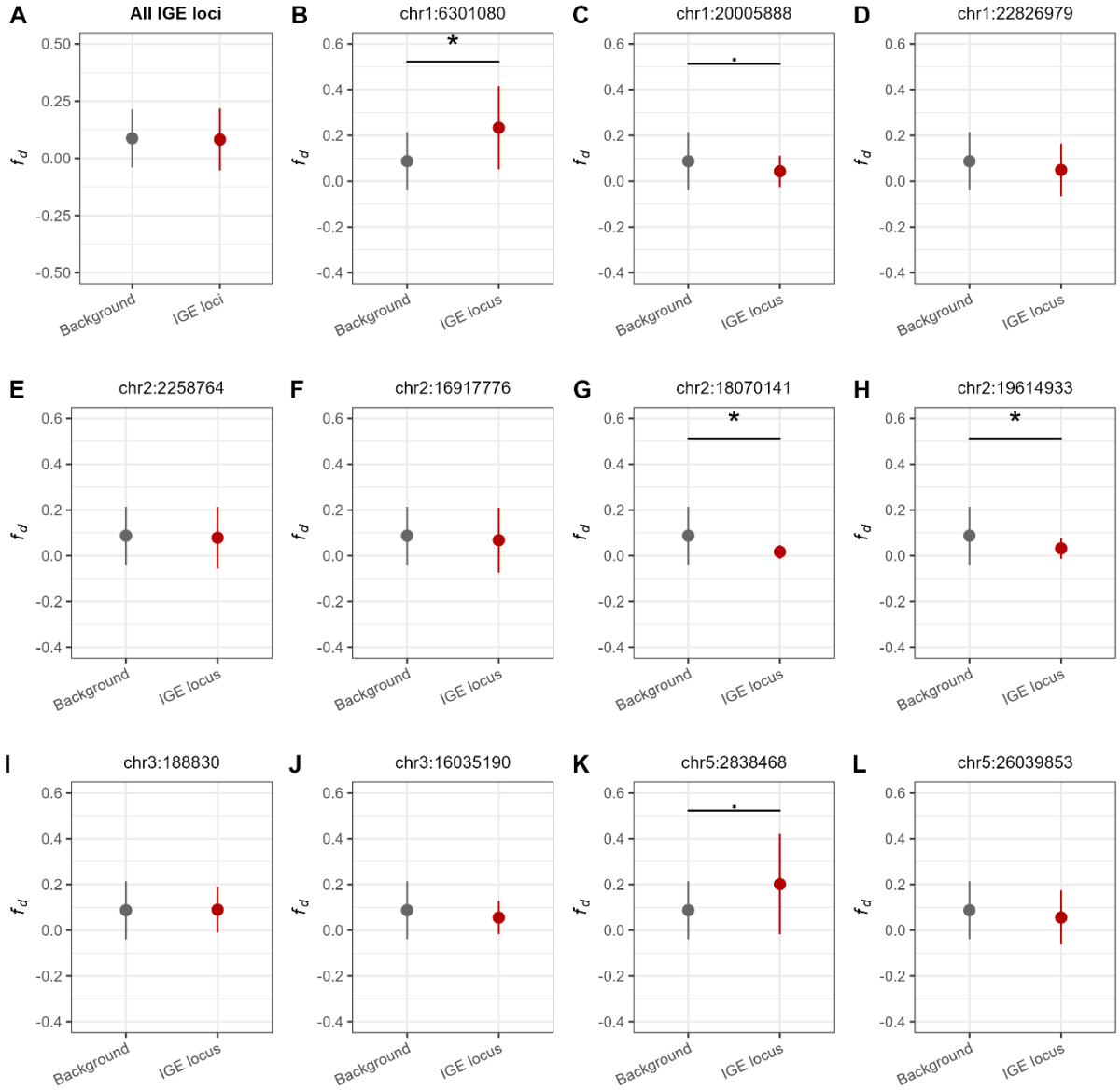
Supplementary Figure 6: Frequency of positive IGE alleles in each ancestry group for the eleven IGE loci (A-K).



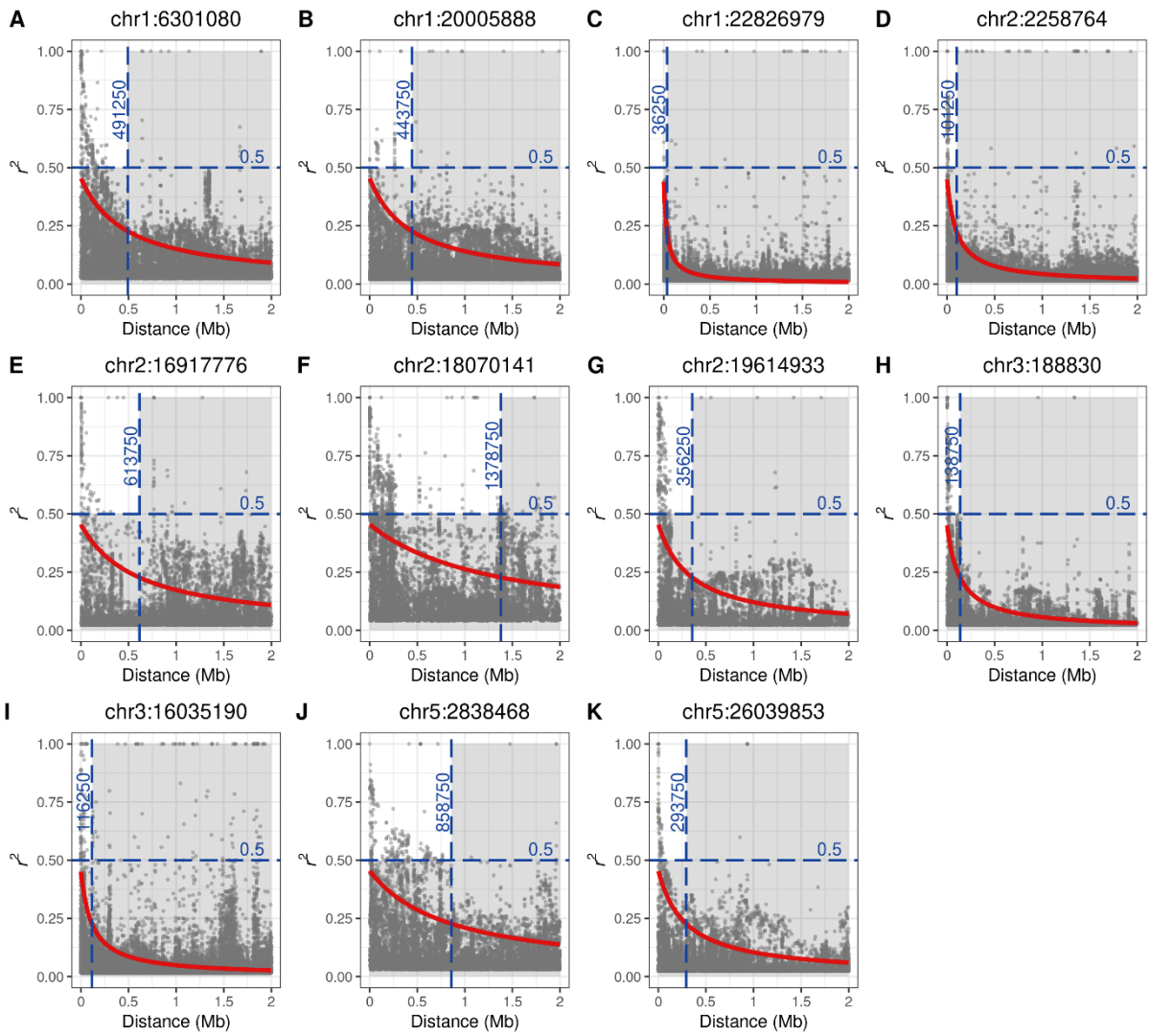
Supplementary Figure 7: Haplotype sharing between north Sweden and relict ancestry groups at chr1:6301080. Patterns of haplotype sharing between ancestry groups in an 80kb region surrounding chr1:6301080 (red arrow). Each row corresponds to an individual, individuals are grouped by ancestry groups, and ancestry groups are separated by black horizontal lines. Each column corresponds to a SNP, with the reference allele colored in yellow, the alternative allele colored in black (all individuals are homozygous), and missing information colored in white. Haplotypes shared between the non-Relicts from north Sweden and the relicts are highlighted in red.



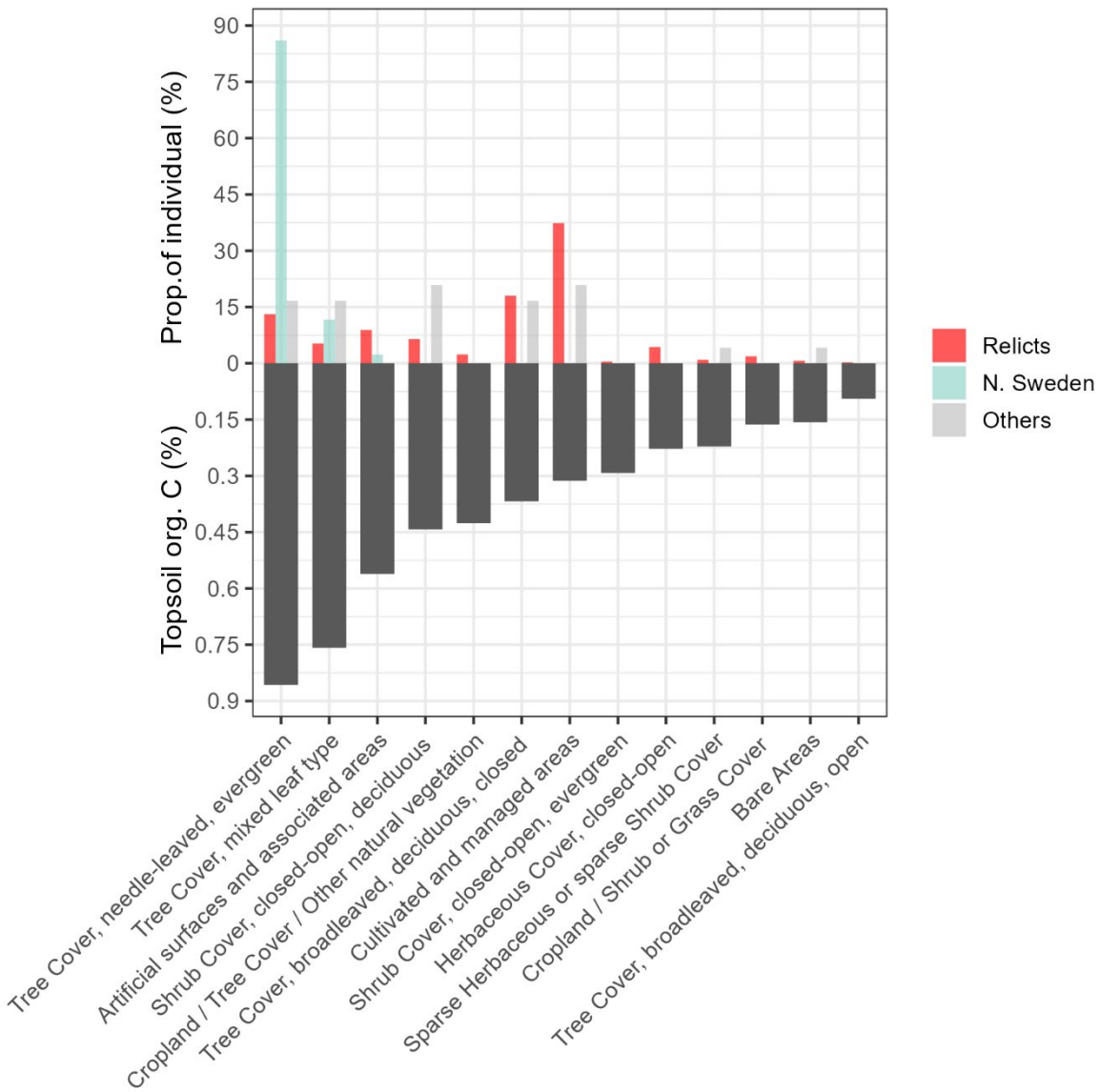
Supplementary Figure 8: Admixture between non-relicts from north Sweden and relicts at IGE loci. **B-E**: Patterson's *D* statistics are compared between background genomic windows (grey) and IGE windows (red). *D* was computed in 20 kb sliding windows along the genome with the following population relationships: ((P1,P2),P3),O) with western Europe as P1, north Sweden as P2, relicts as P3, and *A. lyrata* as the outgroup (O). Points correspond to mean values across genomic windows and error bars to standard deviations. Background windows and IGE windows were compared with two-sided Welch's *t*-tests ("NS": non-significant, ".": $p < 0.1$, "*": $p < 0.05$). **A**: all IGE windows are grouped together ($n = 123$ windows) and compared with the genomic background ($n = 5549$ windows): $t_{125.91} = 1.297$, $p = 0.1971$. **B-L**: IGE windows for each individual IGE locus are compared with the genomic background ($n = 5549$ windows). chr1:6301080: $n = 9$ windows, $t_{8.02} = -3.288$, $p = 0.0110$; chr1:20005888: $n = 11$ windows, $t_{10.03} = 0.985$, $p = 0.3479$; chr1:22826979: $n = 11$ windows, $t_{10.04} = 1.126$, $p = 0.2865$; chr2:2258764: $n = 12$ windows, $t_{11.03} = 1.080$, $p = 0.3029$; chr2:16917776: $n = 12$ windows, $t_{11.05} = 1.180$, $p = 0.2628$; chr2:18070141: $n = 11$ windows, $t_{10.06} = 2.738$, $p = 0.0208$; chr2:19614933: $n = 10$ windows, $t_{9.03} = 1.485$, $p = 0.1717$; chr3:188830: $n = 11$ windows, $t_{10.09} = -0.438$, $p = 0.6706$; chr3:16035190: $n = 12$ windows, $t_{11.14} = 0.743$, $p = 0.4730$; chr5:2838468: $n = 12$ windows, $t_{11.03} = -2.329$, $p = 0.0399$; chr5:26039853: $n = 12$ windows, $t_{11.02} = 1.490$, $p = 0.1642$.



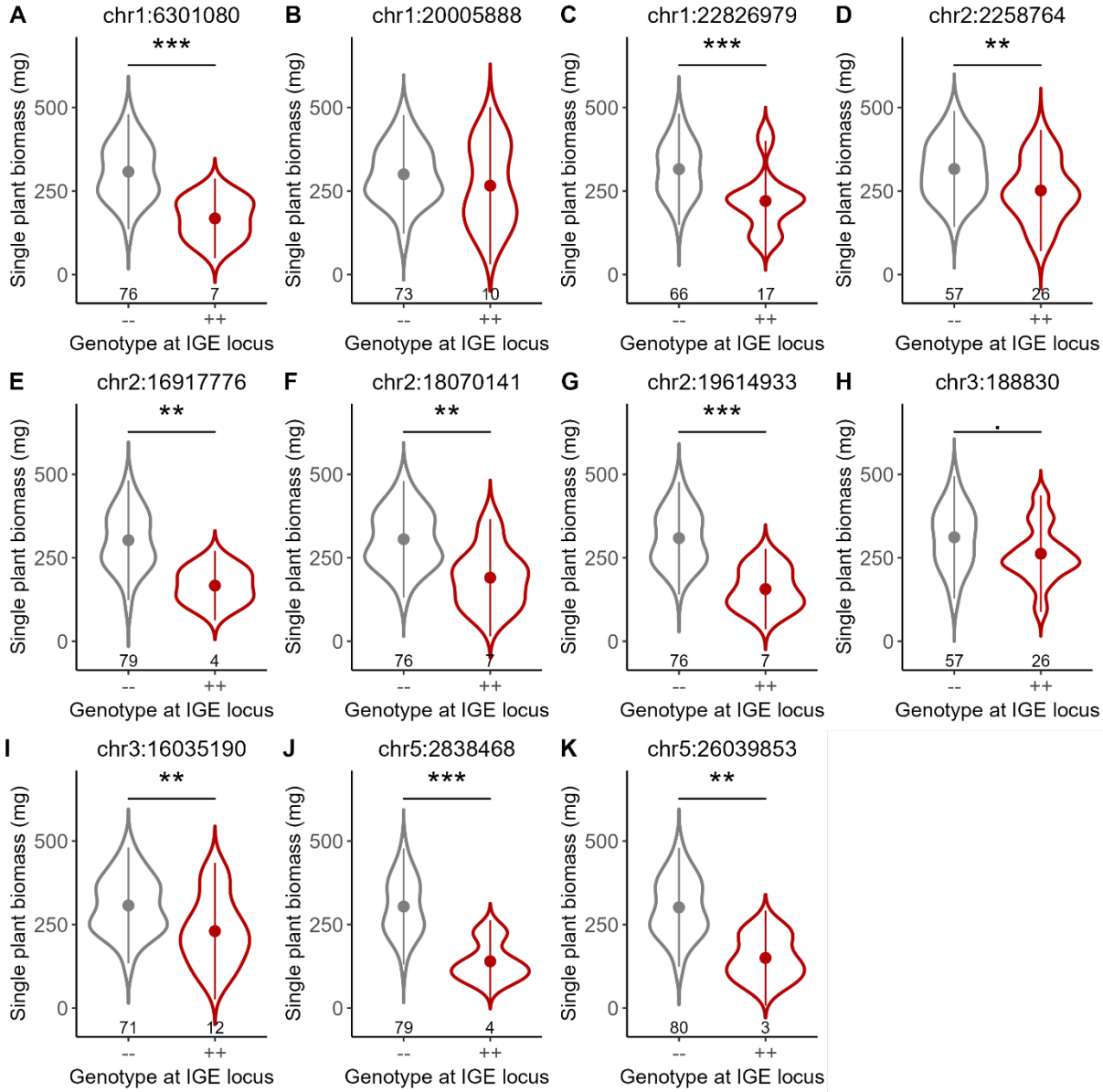
Supplementary Figure 9: Admixture between non-relicts from north Sweden and relicts at IGE loci. **B-E**: f_d statistics are compared between background genomic windows (grey) and IGE windows (red). f_d was computed in 20 kb sliding windows along the genome with the following population relationships: ((P1,P2),P3),O) with western Europe as P1, north Sweden as P2, relicts as P3, and *A. lyrata* as the outgroup (O). Points correspond to mean values across genomic windows and error bars to standard deviations. Background windows and IGE windows were compared with two-sided Welch's *t*-tests ("NS": non-significant, ".": $p < 0.1$, "*": $p < 0.05$). **A**: all IGE windows are grouped together (n = 123 windows) and compared with the genomic background (n = 5549 windows): $t_{126.79} = 0.420$, $p = 0.6754$. **B-L**: IGE windows for each individual IGE locus are compared with the genomic background (n = 5549 windows). chr1:6301080: n = 9 windows, $t_{8.01} = -2.412$, $p = 0.0423$; chr1:20005888: n = 11 windows, $t_{10.14} = 2.139$, $p = 0.0578$; chr1:22826979: n = 11 windows, $t_{10.05} = 1.099$, $p = 0.2975$; chr2:2258764: n = 12 windows, $t_{11.04} = 0.223$, $p = 0.8279$; chr2:16917776: n = 12 windows, $t_{11.04} = 0.474$, $p = 0.6451$; chr2:18070141: n = 11 windows, $t_{10.70} = 7.589$, $p = 1.261e-05$; chr2:19614933: n = 10 windows, $t_{9.25} = 3.794$, $p = 0.0040$; chr3:188830: n = 11 windows, $t_{10.06} = -0.073$, $p = 0.9435$; chr3:16035190: n = 12 windows, $t_{11.15} = 1.529$, $p = 0.1542$; chr5:2838468: n = 12 windows, $t_{11.02} = -1.799$, $p = 0.0994$; chr5:26039853: n = 12 windows, $t_{11.06} = 0.930$, $p = 0.3725$.



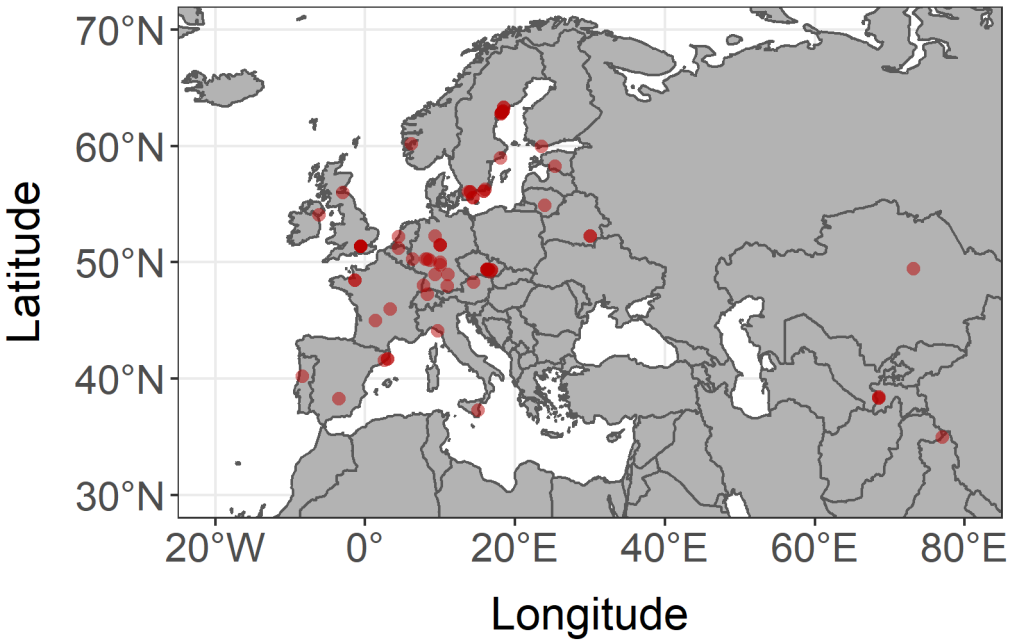
Supplementary Figure 10: Linkage Disequilibrium (LD) decay around the eleven top SNPs associated with IGEs. For each of the eleven SNPs (**A-K**), LD (here measured with r^2) was plotted against the physical distance between the IGE SNPs and all other SNPs in a 2Mb window (grey points). We removed all SNPs with LD values lower than the interchromosomal LD, which was computed as the quadratic mean of the LD between the top IGE SNP and all other SNPs located on different chromosomes. We then fitted a non-linear regression model (red line) following the equation provided by Hill and Weir (1988). To fit the model, we first computed r^2 in 2.5kb sliding windows, and we only retained the 0.95 percentile r^2 value in each window. We then computed the half-LD decay distance (i.e., the distance at which LD has halved), using the fitted values of the regression (vertical blue dashed line). We used this distance to investigate associations between IGE loci and environmental variables (i.e., Genome-Environment Associations) and to investigate candidate genes. For these two analyses, we only considered SNPs with distance lower than the half-LD decay distance and with $r^2 \geq 0.5$ (horizontal blue dashed line). The set of SNPs left after these two filters are in the top left unshaded quarter of the plots.



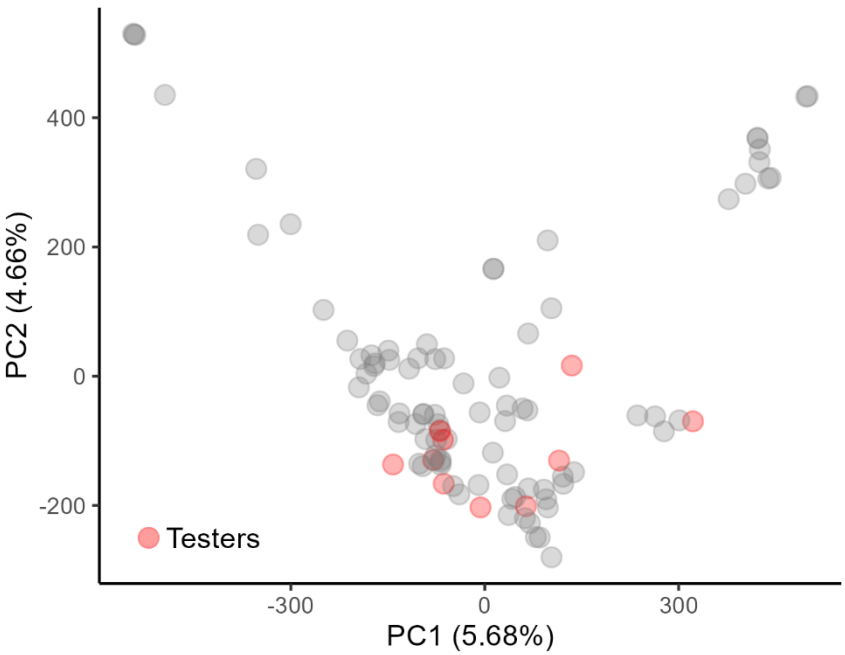
Supplementary Figure 11: Distribution of *A. thaliana* accessions per land cover types (top) and average topsoil organic carbon content per land cover types (bottom). *A. thaliana* accessions were clustered in three groups: Relicts (red), non-Relicts from North Sweden (blue), and other Non-Relict groups (grey). Land cover types were obtained from the Global Land Cover 2000 database[85] accessed through CLIMtools[52].



Supplementary Figure 12: Association between the eleven top SNPs associated with IGEs and biomass measured at the single plant level and in the absence of intraspecific interactions. For each locus (**A-K**) biomass is compared between plants that had the negative IGE allele ("- -", grey plot) vs the positive IGE allele ("+ +", red plot). All accessions are diploid inbred lines. Points and error bars represent the mean ± standard deviation. The number of accessions in each category are reported below each violin plot. For each locus, "- -" vs "+ +" comparisons were tested using a two-sided Wald-test (".": $p < 0.1$, "**": $p < 0.01$, "***": $p < 0.001$). chr1:6301080: $W_1 = 15.88$, $p = 6.761e-05$; chr1:20005888: $W_1 = 1.00$, $p = 0.3174$; chr1:22826979: $W_1 = 15.22$, $p = 9.56e-05$; chr2:2258764: $W_1 = 7.88$, $p = 0.0050$; chr2:16917776: $W_1 = 7.27$, $p = 0.0070$; chr2:18070141: $W_1 = 9.28$, $p = 0.0024$; chr2:19614933: $W_1 = 20.72$, $p = 5.32e-06$; chr3:188830: $W_1 = 3.31$, $p = 0.0690$; chr3:16035190: $W_1 = 6.71$, $p = 0.0096$; chr5:2838468: $W_1 = 12.02$, $p = 0.0005$; chr5:26039853: $W_1 = 6.98$, $p = 0.0082$.



Supplementary Figure 13: Geographic distribution of the 98 *A. thaliana* natural accessions used in the experiment.



Supplementary Figure 14: PCA (Principal Component Analysis) of the 98 accessions used in the experiment. PCA was performed with 206,416 SNPs. The ten testers are colored in red.