

# Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*

Moises Exposito-Alonso<sup>1</sup>, François Vasseur<sup>1§</sup>, Wei Ding<sup>1</sup>, George Wang<sup>1¶</sup>, Hernán A. Burbano<sup>1,2</sup>, Detlef Weigel<sup>1\*</sup>.

<sup>1</sup>Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

<sup>2</sup>Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

\*Corresponding author. Email: [weigel@weigelworld.org](mailto:weigel@weigelworld.org)

§Current address: CNRS, UMR5175, Centre d'Ecologie Fonctionnelle et Evolutive, F-34000 Montpellier, France (FV). ¶Current address: Computomics, Davis, California, United States (GW).

Keywords: climate change, polygenic adaptation, GWA, environmental niche models, random forest, drought, *Arabidopsis thaliana*, image processing.

One-sentence summary: "Future genetic changes in *A. thaliana* populations can be forecast by combining climate change models with genomic predictions based on experimental phenotypic data."

Short title: "Genetic adaptation to extreme drought in *A. thaliana*"

20

21 **Because earth is currently experiencing unprecedented climate change, it is important to predict**  
22 **how species will respond to it. However, geographically-explicit predictive studies frequently**  
23 **ignore that species are comprised of genetically diverse individuals that can vary in their degree**  
24 **of adaptation to extreme local environments; properties that will determine the species' ability**  
25 **to withstand climate change. Because an increase in extreme drought events is expected to**  
26 **challenge plant communities with global warming, we carried out a greenhouse experiment to**  
27 **investigate which genetic variants predict surviving an extreme drought event and how those**  
28 **variants are distributed across Eurasian *Arabidopsis thaliana* individuals. Genetic variants**  
29 **conferring higher drought survival showed signatures of polygenic adaptation, and were more**  
30 **frequently found in Mediterranean and Scandinavian regions. Using geoenvironmental models,**  
31 **we predicted that Central European populations might lag behind in adaptation by the end of**  
32 **the 21<sup>st</sup> century. Further analyses showed that a population decline could nevertheless be**  
33 **compensated by natural selection acting efficiently over standing variation or by migration of**  
34 **adapted individuals from populations at the margins of the species' distribution. These findings**  
35 **highlight the importance of within-species genetic heterogeneity in facilitating an evolutionary**  
36 **response to a changing climate.**

37 Ongoing climate change has already shifted latitudinal and altitudinal distributions of many plant  
38 species (1). Future changes in distributions by local extinctions and migrations are most commonly  
39 inferred from niche models that are based on current climate across species ranges (2, 3). Such  
40 approaches, however, ignore that an adaptive response can occur also *in situ* if there is sufficient  
41 variation in genes responsible for local adaptation (4–6). The plant *Arabidopsis thaliana* is found  
42 under a wide range of contrasting climates, making it distinctively suited to study evolutionary  
43 adaptation to a changing climate (7–9). For the next 50 to 100 years, it is predicted that extreme  
44 drought events, potentially one of the strongest climate change-related selective pressures (10),  
45 will become pervasive across the Eurasian range of *A. thaliana* (2, 11). An attractive hypothesis is  
46 that populations from the southern edge of the species' range (12) provide a reservoir of genetic  
47 variants that can make individuals resistant to future, more extreme, climate conditions (12, 13). To  
48 investigate the potential of *A. thaliana* to adapt to extreme drought events, we first linked genetic  
49 variation to survival under an experimental extreme drought treatment (14–16). By combining  
50 genome-wide association (GWA) techniques that capture signals of local and/or polygenic  
51 adaptation (17, 18) with environmental niche models (8, 19), we then predict genetic changes of

52 populations under future climate change scenarios.

53 We began by exposing a high-quality subset of 211 geo-referenced natural inbred *A.*  
54 *thaliana* accessions (18) to an experimental extreme drought event during the vegetative phase,  
55 which killed the plants before they could reproduce (Table S1). After two weeks of normal growth,  
56 plants were challenged by a terminal severe drought for over six weeks and imaged every 2-4  
57 days (Fig. 1A) (see Supplementary Online Materials [SOM]). A polynomial linear mixed model was  
58 fit to the time-series data to quantify the rate of leaf decay (Fig. 1B-D, Video S1). The genotype  
59 deviations from the mean quadratic-term in the model provided the best estimate of this  
60 survivorship trait in late stages (Fig. S3, see details in SOM), ranging from  $-5$  to  $+5 \times 10^{-4}$  green  
61 pixels/day<sup>2</sup>. The most sensitive plants survived only about 32 days, while the most resilient plants  
62 survived about 15 days longer.

63 The amount of water available during the drought experiment translates to only about  
64 30-40 mm of monthly rainfall, and as expected, accessions with higher survival come from  
65 regions with low precipitation during the warmest season (correlation with climate variable bio18  
66 [[www.worldclim.org](http://www.worldclim.org), ref. (20)]: Pearson correlation,  $r=-0.19$ ,  $p=0.005$ ), and specifically with low  
67 precipitation during May and June ( $r \leq -0.19$ ,  $p \leq 0.005$ ) (see Fig. 2A) (21). To further exploit current  
68 climatic data, we used 19 bioclimatic variables and random forest models (22) for environmental  
69 niche modeling (ENM) to predict the geographic distribution of the drought-survival index across  
70 Europe (Fig. 1C). Surprisingly, we found that individuals with higher drought survival were not only  
71 from the Mediterranean, but also from the opposite end of the species' range in Sweden (Fig. 1C,  
72 ENM cross-validation accuracy=89%, Table S10) (21). In contrast to the warm-dry Mediterranean  
73 climate, Scandinavian dry periods occur on average at freezing temperatures (Fig. S12).  
74 Consequently, precipitation might occur as snow, which is not accessible for plants and produces a  
75 physiological drought response (23).

76 We then studied whether the different populations of *A. thaliana* are locally adapted (5) to  
77 low precipitation regimes via an increased drought-survival. Using an extended panel of 762 *A.*  
78 *thaliana* accessions (Table S1) we carried out genetic clustering (24) and studied population  
79 trajectories (25) (Fig. 2). This corroborated the existence of a so-called 'relict' group (12) and ten  
80 other derived groups of relict (e.g. Spanish groups) or other (e.g. Central Europe) origin; likely of the  
81 result of complex migration and admixture processes (26). A generalized linear model indicated  
82 that genetic group membership explains a significant amount of drought-survival variance (GLM:  
83  $R^2=12.8\%$ ;  $p=4 \times 10^{-5}$ ), with the North (N) Swedish and Northeastern (NE) Spanish groups each  
84 having on average higher survival than the other groups (t-test  $p \leq 0.01$ ). A population graph

84 estimated by Treemix (27) suggested a gene flow edge between the Mediterranean and  
85 Scandinavian drought-resistant genetic groups, potentially indicative of historical sharing of  
86 drought survival alleles (Fig. 2D). Finally, running an ENM of the genetic group membership with  
87 climatic variables from the origin of plants confirmed that the most important predictive variable  
88 is precipitation during the warmest quarter (bio18), followed by mean temperature of the driest  
89 quarter (bio9), and minimum temperature of the coldest month (bio6) (ENM accuracy > 95%. Fig.  
90 S8B and Table S10). As our results indicate that the deepest genetic structure parallels the local  
91 precipitation regimes and the ability of populations to survive drought, we expect that areas with  
92 the strongest decline in rainfall will see the most turnover in genetic diversity (see Fig.12 Fig. S8)  
93 (11).

94 Because the potential of populations to adapt to drought will depend on the genetic  
95 architecture of the selected trait, we identified drought-associated loci with EMMAX (28), a  
96 genome-wide association (GWA) method. Although genotype-associated variance (28)  $h^2$  was  
97 50%, no individual SNP was significantly associated with drought survival (minimum  $p \sim 10^{-7}$ , after  
98 FDR or Bonferroni corrections  $p > 0.05$ ) (Fig. S5, Table S3). Significant associations in multiple  
99 phenotypes have been detected in similarly powered *A. thaliana* experiments (29). While multiple  
100 testing adjustment can over-correct p-values and obscure true associations, the absence of  
101 significant associations may also be due to (i) polygenic trait architecture, with many small-effect  
102 loci (30), and/or (ii) confounding by strong population structure, consistent with the association of  
103 drought survival with genetic group membership.

104 To test for polygenic adaptation, we repeated the GWA analyses with a model that  
105 specifically handles both oligo- and polygenic architectures, BSLMM (31). This model estimates,  
106 among other parameters, the probability that each SNP comes from a group of major-effect loci.  
107 Around half of the top non-significant EMMAX SNPs were found to have over 99% probability of  
108 belonging to such a major-effect group (Fisher's exact test of overlap,  $p=3 \times 10^{-7}$ ; see SOM). We  
109 further tested the polygenic hypothesis using the population genetic approach of Berg & Coop  
110 (32). The test is based on the principle that if populations diverge in drought-survival due to many  
111 loci, there should be an orchestrated shift in their allele frequency. After testing some 60 groups  
112 of EMMAX SNP hits of variable size and at different ranks, we detected the most significant signal  
113 of polygenic adaptation with the group that included the 151 top SNPs (Table S9). The signal was  
114 lost for ranks below the top 300-400 EMMAX SNPs (Table S9). We then compared summary  
115 statistics of the top 151 SNPs with background SNPs matched in frequency to avoid GWA  
116 discovery biases. The top 151 SNPs showed high  $F_{ST}$  values, consistent with allele frequency



117 differentiation between populations (Fig. S5). Tajima's D values were positive (U Mann-Whitney  
118 p-value < 0.05), indicating intermediate allele frequencies at the GWA loci (Fig. S5), which could  
119 be a result of selection favoring alternative alleles in different ecological niches of the species (33).  
120 The top SNPs did not show any evidence for precipitous reductions of haplotypic diversity, as  
121 would be expected for hard selective sweeps (34) (Fig. S5). Together these patterns fit the  
122 expectations of local adaptation from a polygenic trait controlled by some hundreds loci (35) –  
123 theoretically expected to enable a fast response to a new environmental shift (36)

124 During local adaptation, the relevant loci diverge due to natural selection across  
125 populations, which generates a statistical correlation with population groups (37). In this situation,  
126 the default correction of population structure applied in GWA might obscure some of the true  
127 associations. While  $F_{ST}$  scans can be useful to identify overly divergent loci across populations,  
128 elevated genome-wide  $F_{ST}$  due to strong population structure can difficult outlier detection (37),  
129 as it is in our case (Fig. S4). In order to recover relevant variants that are deeply divergent across  
130 populations, we can study the ancestry of each SNP. Using ChomoPainter (38), which relies on  
131 linkage disequilibrium information, we segment each genome in question into its different  
132 population ancestries (here 11 groups). The first outcome of this analysis was that individuals from  
133 NW and NE Spain and, to lesser extent, the Southern Mediterranean (Fig. 2A), have inherited  
134 many DNA segments from relict individuals (Fig. S7). Then, in a generalized linear model  
135 framework, we test whether the ancestries of individuals at a SNP coincide with the observed  
136 phenotypic differences in drought-survival. Performing this “ancestry” genome-wide association  
137 (aGWA) and using a permutation correction of p-values (see SOM), we detected 8 distinct peaks  
138 ( $p < 0.001$ , Fig. 3A) including over 1,000 significant SNPs (70 SNPs after linkage disequilibrium  
139 pruning) (Table S4). The most prominent peak was located on chromosome 5 and explained over  
140 20% of the variance in drought survival (Table S4). There was no overlap in top SNPs between  
141 GWA and aGWA because they search for different association signals. Our aGWA resembles other  
142 admixture mapping techniques (39), and might be most useful for associations in scenarios of  
143 adaptive introgression and local adaptation. To understand the origin of aGWA-identified SNPs,  
144 we constructed trees for all concatenated aGWA SNPs and for genome-wide background SNPs.  
145 Although the individuals from both the warm (Iberia and relicts) and cold (Scandinavia) edges of  
146 the species distribution are far apart in genome-wide SNPs, they are closely related in  
147 drought-associated SNPs (Fig. 3B). Overall, this is consistent with a common Mediterranean origin  
148 of drought-adaptive genetic variants of both Northern and Southern individuals (Fig. 2D, Fig. 3B),  
149 and highlights the relevance of populations at the latitudinal extremes of the species range as a

150 possible genetic reservoir for future climate change adaptation (12).

151 Depending on the nature of the stress, different mechanisms for drought adaptation can  
152 be most advantageous (23, 40, 41). Annual plants, including *A. thaliana*, typically adapt to water  
153 stress deficit by accelerating the transition from germination to flowering (escape strategy) (14–16,  
154 41) instead of increasing water use efficiency (avoidance strategy). Previous drought experiments  
155 with *A. thaliana* showed variation in both strategies but concluded it predominantly utilizes the  
156 drought escape strategy. Our extreme drought experiment focused in characterising the  
157 avoidance strategy by means of the drought-survival index, which was linearly associated to  
158 precipitation regimes (Fig. S11, Table S6). This trait was not correlated with flowering time of the  
159 accessions in unstressed conditions (Pearson correlation,  $r=0.07$ ,  $p=0.12$ ). However, we found a  
160 positive correlation between drought-survival and flowering time GWA summary statistics of the  
161 top 151 SNPs (Pearson correlation,  $r=0.51$ ,  $p=1 \times 10^{-11}$ , see SOM) – suggesting a weak genetic  
162 trade-off (16). Interestingly, we did not find any associated between GWA or aGWA top SNPs and  
163 known flowering time QTLs (14–16), but rather a weak enrichment with membrane transporters  
164 (see SOM). Adjustment of osmotic balance through cell membrane transport is a drought  
165 avoidance mechanism (42) that might also confer cross-tolerance to other abiotic stresses (43),  
166 therefore it might be of relevance for Scandinavian *A. thaliana* accessions or other populations in  
167 the niche extremes (Fig. S12) (21).

168 Increased survival to extreme abiotic stresses should confer an evolutionary advantage  
169 given the predicted increase in drought frequency and intensity both around the Mediterranean  
170 and in Europe, which will constitute a critical hazard for many plants, including *A. thaliana* (2, 11).  
171 Environmental niche models (ENM), which have been developed to relate species distributions to  
172 climate variables, can be used to predict future changes to species' ranges (2, 3). Ignoring  
173 adaptation from standing variation (44–46), however, could lead to overestimates of extinction  
174 rates (47–49). By fitting ENM of current climate with SNP data (19), using a similar rationale as in  
175 Hancock and colleagues' "climate GWA" (7), we can predict the most likely genetic makeup under  
176 current and future climate conditions. Using such an approach, we trained ENMs with 762  
177 accessions and produced maps of the present distributions of the 151 GWA and 70 aGWA  
178 drought-associated SNPs (all ENM 5CV accuracy >92%; Table S3–4, Fig. S13–16). Concatenating  
179 the 221 maps, we inferred the most likely individual genotype at each location. At present,  
180 individuals from both Northern and Southern edges of the distribution are predicted to harbor  
181 more drought-survival alleles than those located in between (Fig. 3C, Fig. S15–16, with the  
182 quadratic term in a regression of allele count on latitude being positive at  $p=10^{-3}$ ), corroborating

183 our previous observations. Then, using the trained ENM, we forecast the distribution of the 221  
184 drought-survival alleles in 2070 (rpc 8.5, IPCC, [www.ipcc.ch](http://www.ipcc.ch), ref. (20)). While it was expected that  
185 populations in the Mediterranean Basin would need to become more drought resistant (11), we  
186 predicted a more robust increase in the total number of drought-survival alleles for Central Europe  
187 (Fig. 3, Fig. S14-15). This is because rainfall in Central Europe will likely become more similar to that  
188 in the Mediterranean by 2070 (2, 11) (Fig. S12).

189 Because some of the drought-survival alleles are currently not yet present in Central  
190 Europe, we speculated that gene migration might be necessary to facilitate adaptation to future  
191 conditions (50). An underlying assumption of the ENM is that allele presence only depends on  
192 environmental variables, but this assumption, “universal migration”, may not be realistic for future  
193 predictions if present distributions are geographically narrow. We therefore included two  
194 geographic boundary conditions in the ENM to generate two models that were either more or less  
195 “migration-limited” (see SOM). After fitting all possible models and predicting allele distributions  
196 with future climate, we calculated the difference of predicted presence per map grid cell between  
197 the naïve, free migration ENM and the two geographically constrained ones (Fig. 3D-E). If an allele  
198 has currently a narrow distribution or is specific to a certain genetic background, its future  
199 presence in an area might not be predicted by the constrained models, even though the climate  
200 variables coincide with the SNP’s environmental range. Such a scenario seems to apply to Central  
201 Europe, as the deficit in drought-survival alleles predicted by the free over the constrained models  
202 was 8-30% (18-66 out of 221) (Fig. 3E; with the quadratic term in a regression of the allele count  
203 difference on latitude being negative at  $p < 10^{-10}$ ). Central European populations may therefore be  
204 under threat of lagging adaptation by the end of the 21<sup>st</sup> century.

205 In the end, for a population to persist, not only the number of drought-survival alleles has  
206 to increase, but it has to do so in actual individuals (51). The chance of this occurring will depend  
207 on local allele frequencies and the natural selection favouring the drought-survival alleles.  
208 Therefore, we studied current allele frequencies at three representative locations with the highest  
209 sampling density in our dataset (40 samples within a 50 kilometer area): Madrid (Spain), Tübingen  
210 (Germany) and Malmö (Sweden), which are at the southern edge, center and northern edge of the  
211 range, respectively. Based on ENM predictions, we calculated allele frequencies from present to  
212 2070. Frequencies are predicted to increase significantly only in the Tübingen population  
213 (Student’s t test,  $p < 10^{-16}$ , Table S11), but not in Madrid and Malmö, indicating that these two  
214 populations might be already adapted to the future local climate. Because the Tübingen  
215 population already has most drought-associated alleles (53% of 70 aGWA SNPs and 90% of 151

216 GWA SNPs), increasing the number of total favorable alleles in individual genotypes should be  
217 feasible, especially since there are single genotypes that have 63% (aGWA) and 90% (GWA) of  
218 those alleles already present (see SOM). Starting 50-generations simulations at the present  
219 Tübingen frequency of independent drought-survival alleles and assuming a range of selection  
220 coefficients, we estimated that a 1-3% of fitness advantage on average would be necessary to  
221 increase frequencies to match those of the adapted Madrid and Malmö populations (Fig. S17, see  
222 SOM). Such selection could take place efficiently in large populations like the ones of a  
223 highly-reproductive weed (51, 52).

224 Leveraging the model organism *A. thaliana*, we have begun to address key questions to  
225 understand the burning issue of climate change effects on biodiversity. We provide evidence for  
226 the possibility of adaptive genetic variation to extreme drought events. Harnessing the power of  
227 methods that allow polygenic genetic architecture and testing evolutionary hypotheses of natural  
228 selection, we detected that relevant genetic variants had been under polygenic local adaptation  
229 and were more abundant at the edges of the species range. Extreme adaptation at range edges  
230 might indeed be critical for a species' persistence under climate change. Although many aspects of  
231 future adaptation are not considered here, namely non-drought related or seasonal climate  
232 change (51), biotic interactions, phenotypic plasticity, or novel adaptive mutations (53), our  
233 spatially explicit analyses emphasize the potential of adaptive evolution from standing variation  
234 to ameliorate climate change's detrimental effects.

## 235 REFERENCES

- 236 1. C. Parmesan, G. Yohe, A globally coherent fingerprint of climate change impacts across  
237 natural systems. *Nature*. **421**, 37–42 (2003).
- 238 2. W. Thuiller, S. Lavorel, M. B. Araújo, M. T. Sykes, I. C. Prentice, Climate change threats to plant  
239 diversity in Europe. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8245–8250 (2005).
- 240 3. T. Jezkova, J. J. Wiens, Rates of change in climatic niches in plant and animal populations are  
241 much slower than projected climate change. *Proc. R. Soc. B.* **283**, 20162104 (2016).
- 242 4. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**,  
243 38–44 (2008).
- 244 5. G. Turesson, The Plant Species in Relation to Habitat and Climate. *Hereditas.* **6**, 147–236  
245 (1925).
- 246 6. J. Hereford, A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173**,  
247 579–588 (2009).
- 248 7. A. M. Hancock *et al.*, Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*.

- 
- 249           **334**, 83–86 (2011).
- 250       8.   A. Fournier-Level *et al.*, A map of local adaptation in *Arabidopsis thaliana*. *Science*. **334**,  
251       86–89 (2011).
- 252       9.   J. R. Lasky *et al.*, Characterizing genomic variation of *Arabidopsis thaliana*: the roles of  
253       geography and climate. *Mol. Ecol.* **21**, 5512–5529 (2012).
- 254       10.  A. M. Siepielski *et al.*, Precipitation drives global variation in natural selection. *Science*. **355**,  
255       959–962 (2017).
- 256       11.  A. Dai, Increasing drought under global warming in observations and models. *Nat. Clim.*  
257       *Chang.* **3**, 52–58 (2012).
- 258       12.  A. Hampe, R. J. Petit, Conserving biodiversity under climate change: the rear edge matters.  
259       *Ecol. Lett.* **8**, 461–467 (2005).
- 260       13.  J. A. Lee-Yaw *et al.*, A synthesis of transplant experiments and ecological niche models  
261       suggests that range limits are often niche limits. *Ecol. Lett.* (2016), doi:10.1111/ele.12604.
- 262       14.  J. A. Bac-Molenaar, C. Granier, J. J. B. Keurentjes, D. Vreugdenhil, Genome-wide association  
263       mapping of time-dependent growth responses to moderate drought stress in *Arabidopsis*.  
264       *Plant Cell Environ.* **39**, 88–102 (2016).
- 265       15.  M. El-Soda, W. Kruijjer, M. Malosetti, M. Koornneef, M. G. M. Aarts, Quantitative trait loci and  
266       candidate genes underlying genotype by environment interaction in the response of  
267       *Arabidopsis thaliana* to drought. *Plant Cell Environ.* **38**, 585–599 (2015).
- 268       16.  A. M. Kenney, J. K. McKay, J. H. Richards, T. E. Juenger, Direct and indirect selection on  
269       flowering time, water-use efficiency (WUE,  $\delta^{13}C$ ), and WUE plasticity to drought in  
270       *Arabidopsis thaliana*. *Ecol. Evol.* **4**, 4505–4521 (2014).
- 271       17.  B. Brachi *et al.*, Investigation of the geographical scale of adaptive phenological variation and  
272       its underlying genetics in *Arabidopsis thaliana*. *Mol. Ecol.* **22**, 4222–4240 (2013).
- 273       18.  1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in  
274       *Arabidopsis thaliana*. *Cell*. **166**, 481–491 (2016).
- 275       19.  J. A. Banta *et al.*, Climate envelope modelling reveals intraspecific relationships among  
276       flowering phenology, niche breadth and potential range size in *Arabidopsis thaliana*. *Ecol. Lett.*  
277       **15**, 769–777 (2012).
- 278       20.  R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated  
279       climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- 280       21.  J. P. Mojica *et al.*, Genetics of water use physiology in locally adapted *Arabidopsis thaliana*.  
281       *Plant Sci.* **251**, 12–22 (2016).
- 282       22.  L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
- 283       23.  J. Ingram, D. Bartels, The Molecular Basis of Dehydration Tolerance in Plants. *Annu. Rev. Plant*  
284       *Physiol. Plant Mol. Biol.* **47**, 377–403 (1996).
- 285       24.  D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated

- 286 individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 287 25. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple  
288 genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
- 289 26. C.-R. Lee *et al.*, On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat.*  
290 *Commun.* **8**, 14458 (2017).
- 291 27. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide  
292 allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- 293 28. H. M. Kang *et al.*, Variance component model to account for sample structure in  
294 genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
- 295 29. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana*  
296 inbred lines. *Nature.* **465**, 627–631 (2010).
- 297 30. G. Gibson, Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2011).
- 298 31. X. Zhou, P. Carbonetto, M. Stephens, Polygenic modeling with bayesian sparse linear mixed  
299 models. *PLoS Genet.* **9**, e1003264 (2013).
- 300 32. J. J. Berg, G. Coop, A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet.* **10**,  
301 e1004412–e1004412 (2014).
- 302 33. P. W. Hedrick, Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics.  
303 *Annu. Rev. Ecol. Evol. Syst.* **37**, 67–93 (2006).
- 304 34. P. Pavlidis, D. Živkovic, A. Stamatakis, N. Alachiotis, SweeD: likelihood-based detection of  
305 selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
- 306 35. J. K. Pritchard, J. K. Pickrell, G. Coop, The genetics of human adaptation: hard sweeps, soft  
307 sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–15 (2010).
- 308 36. K. Jain, W. Stephan, Rapid adaptation of a polygenic trait after a sudden environmental shift.  
309 *arXiv [q-bio.PE]* (2016), (available at <http://arxiv.org/abs/1610.05478>).
- 310 37. E. B. Josephs, J. R. Stinchcombe, S. I. Wright, What can genome-wide association studies tell  
311 us about the evolutionary forces maintaining genetic variation for quantitative traits? *New*  
312 *Phytol.* (2017), doi:10.1111/nph.14410.
- 313 38. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense  
314 haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
- 315 39. D. Shrinier, A. Adeyemo, E. Ramos, G. Chen, C. N. Rotimi, Mapping of disease-associated  
316 variants in admixed populations. *Genome Biol.* **12**, 223 (2011).
- 317 40. F. Tardieu, Any trait or trait-related allele can confer drought tolerance: just design the right  
318 drought scenario. *J. Exp. Bot.* **63**, 25–31 (2012).
- 319 41. F. Roux, P. Touzet, J. Cuguen, V. Le Corre, How to be early flowering: an evolutionary  
320 perspective. *Trends Plant Sci.* **11**, 375–381 (2006).
- 321 42. K. M. Jarzyniak, M. Jasiński, Membrane transporters and drought resistance - a complex issue.



- 322 *Front. Plant Sci.* **5**, 687 (2014).
- 323 43. W. R. Swindell, The association among gene expression responses to nine abiotic stress  
324 treatments in *Arabidopsis thaliana*. *Genetics*. **174**, 1811–1824 (2006).
- 325 44. S. U. Pauls, C. Nowak, M. Bálint, M. Pfenninger, The impact of global climate change on  
326 genetic diversity within populations and species. *Mol. Ecol.* **22**, 925–946 (2013).
- 327 45. J. L. Brown *et al.*, Predicting the genetic consequences of future climate change: The power of  
328 coupling spatial demography, the coalescent, and historical landscape changes. *Am. J. Bot.*  
329 **103**, 153–163 (2016).
- 330 46. M. C. Fitzpatrick, S. R. Keller, Ecological genomics meets community-level modelling of  
331 biodiversity: mapping the genomic landscape of current and future environmental  
332 adaptation. *Ecol. Lett.* **18**, 1–16 (2015).
- 333 47. R. A. Catullo, S. Ferrier, A. A. Hoffmann, Extending spatial modelling of climate change  
334 responses beyond the realized niche: estimating, and accommodating, physiological limits  
335 and adaptive evolution. *Glob. Ecol. Biogeogr.* **24**, 1192–1202 (2015).
- 336 48. C. Moritz, R. Agudo, The future of species under climate change: resilience or decline? *Science*.  
337 **341**, 504–508 (2013).
- 338 49. A. A. Hoffmann, C. M. Sgrò, Climate change and evolutionary adaptation. *Nature*. **470**,  
339 479–485 (2011).
- 340 50. S. N. Aitken, M. C. Whitlock, Assisted Gene Flow to Facilitate Local Adaptation to Climate  
341 Change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).
- 342 51. A. Fournier-Level *et al.*, Predicting the evolutionary dynamics of seasonal adaptation to novel  
343 climates in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2812–21 (2016).
- 344 52. F. Roux, S. Giancola, S. Durand, X. Reboud, Building of an experimental cline with *Arabidopsis*  
345 *thaliana* to estimate herbicide fitness cost. *Genetics*. **173**, 1023–1031 (2006).
- 346 53. M. Exposito-Alonso *et al.*, The rate and effect of de novo mutations in a colonising lineage of  
347 *Arabidopsis thaliana*. *bioRxiv* (2016), p. 050203.

348

## 349 SUPPLEMENTAL MATERIALS

350 Methods and any associated references are available in the online version of the paper. [[link](#)]

351 Supplementary text

352 Figs. S1 to S17

353 Video S1

354 Tables S1 to S1



355 References (\* to \*)

356 **URLs** Code for image analysis pipeline available at  
357 <http://github.com/MoisesExpositoAlonso/hippo>. Code for ancestryGWA available at  
358 <http://github.com/MoisesExpositoAlonso/aGWA>. Bioclimatic data used in the paper accessible  
359 through an R package stored in <http://github.com/MoisesExpositoAlonso/rbioclim>. Phenotypic  
360 datasets available at dryad [[link here](#)]. Genomes are available at  
361 <http://1001genomes.org/data/GMI-MPI/releases/v3.1/>.

362 **Acknowledgements** We thank I. Henderson for the recombination map, R. Wedegärtner for  
363 assistance with the greenhouse drought experiment, the Petrov, Coop, Ross-Ibarra, Gaut and  
364 Schmitt labs for discussions. We thank J. Lasky, X. Picó, A. Hancock, H. Thomassen, T.  
365 Mitchell-Olds, J. Mujica, P. Lang, and D. Seymour for comments and the Weigel and Burbano labs  
366 for discussion. This work was supported by ERC Advanced Grant IMMUNEMESIS to DW and the  
367 President's Fund of the Max Planck Society, project "Darwin" to HAB.

368 **Author contribution** MEA conceived and designed the project. GW and FV helped and advised on  
369 image phenotyping and FV provided additional phenotypes. MEA and WD performed  
370 chromosome painter analyses. MEA performed the drought experiment, processed the image  
371 data, and designed and carried out the statistical analyses. DW and HAB advised and oversaw the  
372 project. MEA wrote the first draft and together with HAB and DW wrote the final manuscript with  
373 input from all authors.

374

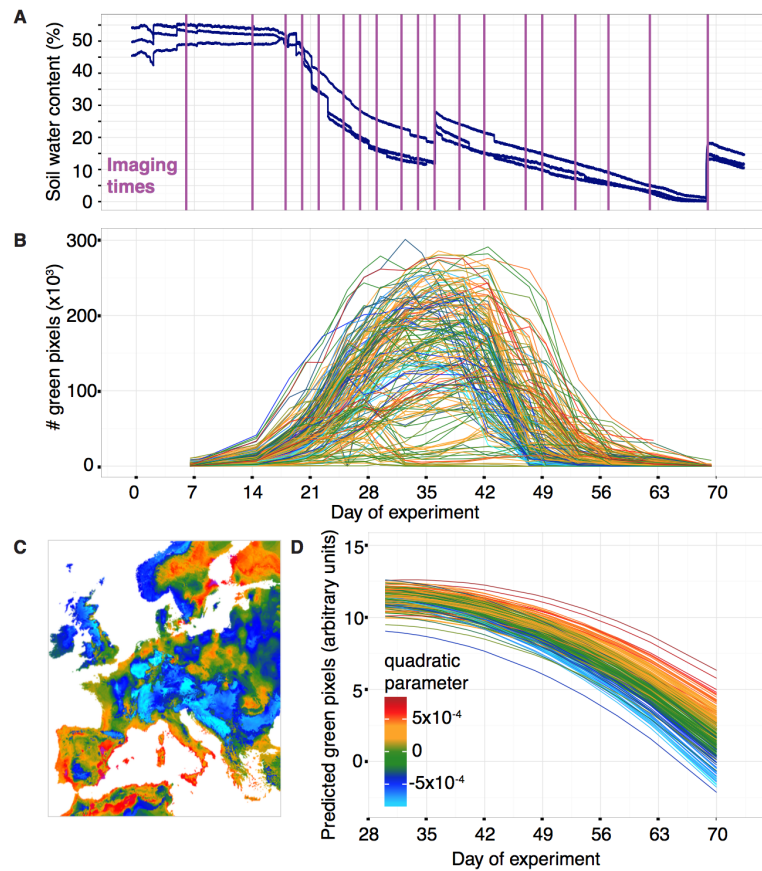
375

376

377

## FIGURE LEGENDS

378



379

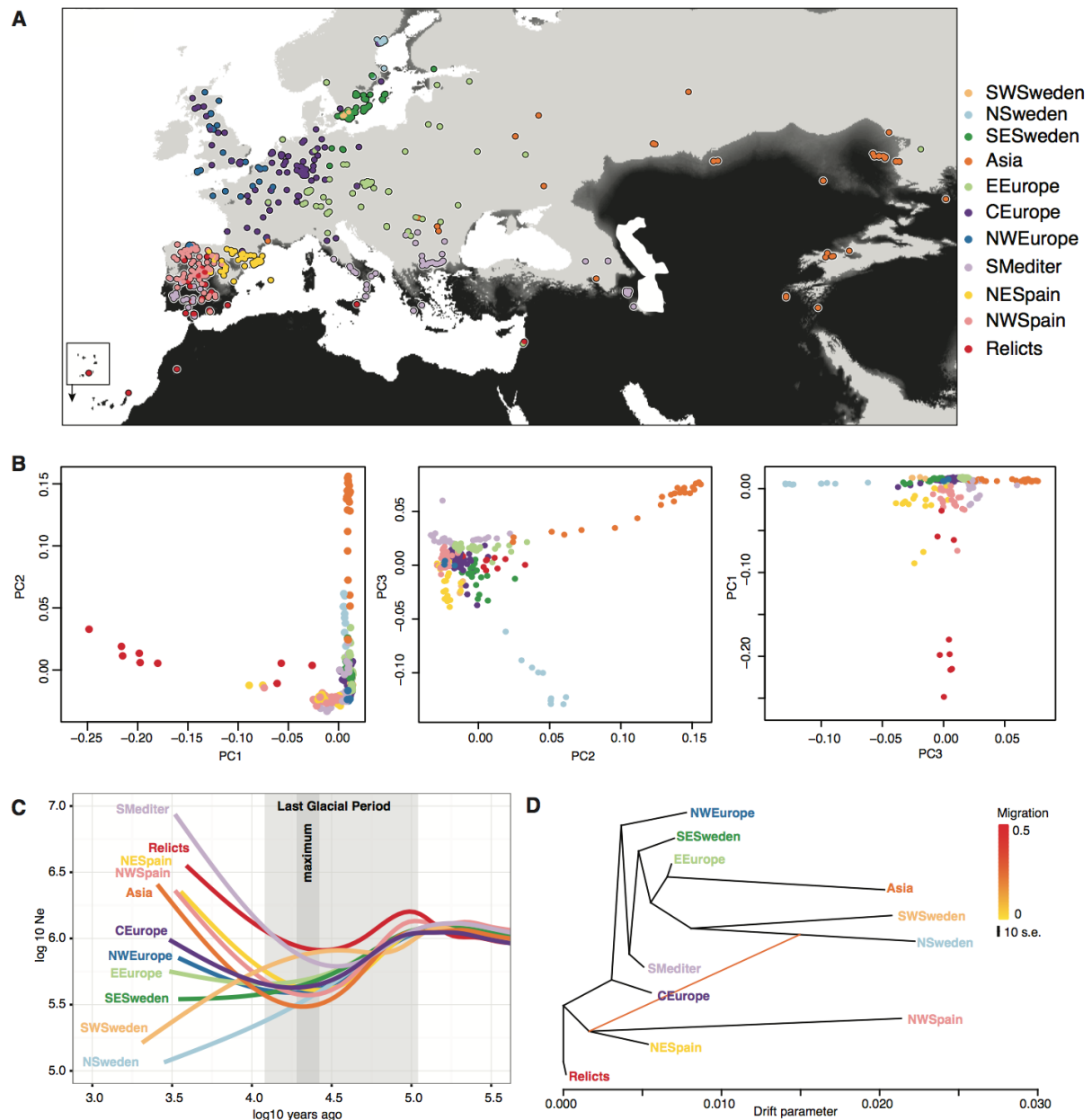
### Figure 1. Terminal drought treatment and phenotyping of 211 accessions.

380 (A) Soil water content from three sensors placed in three experimental trays regularly distributed  
381 in the greenhouse. Purple lines indicate dates of image acquisition. (B) Trajectories of total rosette  
382 area of 200 randomly chosen pots (see [Video S1](#)). Color index according to quadratic parameter in  
383 (D). (C) Map projection of the environmental niche model prediction of the quadratic parameter  
384 (the drought-survival index) in (D). (D) Decay trajectory modeled with a polynomial regression,  
385 with genotypes as random factors, from the average maximum day of green pixels until the end of  
386 the experiment. Each line corresponds to one genotype.

387

388

389



390

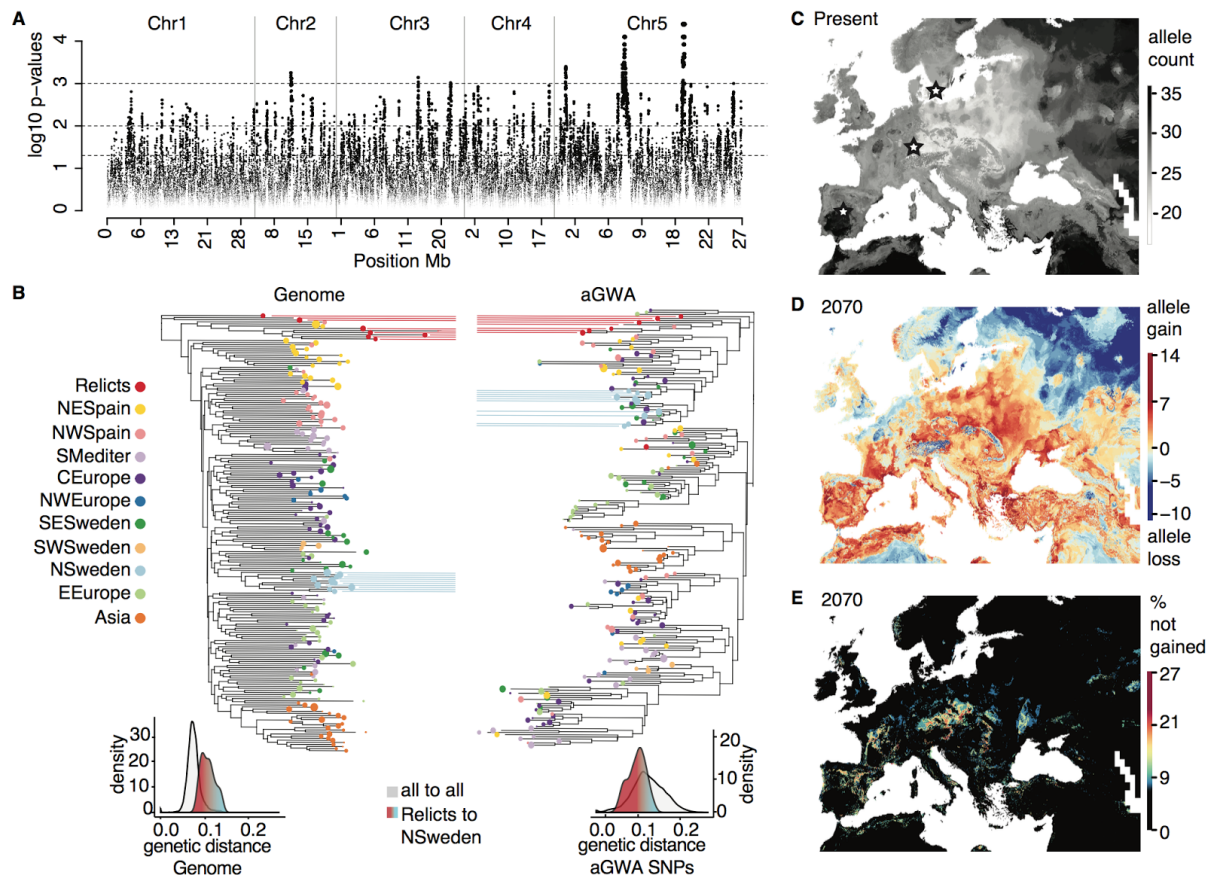
**Figure 2. Population structure and history of 762 high-quality genomes.**

391 **(A)** Geographic locations and 11 genetic clusters estimated by ADMIXTURE (K11 showed the lowest  
 392 cross-validation error). Black indicates less than 40 mm of June rainfall (1960 to 1990 average),  
 393 which corresponds to the amount of water provided in our drought experiment (Fig. 1). Note the  
 394 presence of black areas in the Mediterranean basin and along the coast in Scandinavia (partially  
 395 obscured by colored circles). Cape Verde Islands are shown as inset. **(B)** Principal Component

396 Analysis of genome-wide SNPs. (C) Effective population sizes in time estimated from MSMC. (D)  
 397 Population ancestral graph and the first migration trajectory using Treemix.

398

399



400

**Figure 3. Ancestry GWA of drought survival and environmental predictions.**

401 (A) Manhattan plot of SNPs from ancestry GWA (aGWA) after permutation correction of p-values.  
 402 Dashed lines indicate significant thresholds at 0.05, 0.01, and 0.001. (B) Top, neighbour Joining  
 403 phylogeny of 1,000 concatenated genome-wide SNPs compared with a phylogeny of all  
 404 significant aGWA SNPs (ca 1,000). Colors indicate population clusters (Fig. 2). Relicts and N.  
 405 Swedish groups are highlighted. Bottom, genetic distances for all or aGWA SNPs. (C)  
 406 Environmental niche models of 70 top aGWA SNPs (after LD pruning), trained with averages from  
 407 1960-1990, and then (D) used to forecast gain or loss of alleles in 2070 under free migration. (E)  
 408 The bottom indicates the discrepancy of gained alleles between the geographically constrained  
 409 (PCA control) model relative to the free migration model.

410

411

1  
2  
3  
  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

Supplementary Information for:

# Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*

Moises Exposito-Alonso, François Vasseur, Wei Ding, George Wang, Hernán A. Burbano,  
Detlef Weigel.

Correspondence to: [weigel@weigelworld.org](mailto:weigel@weigelworld.org)

This PDF file includes:

Supplementary Text

Figures S1 to S17

Video S1

Tables S1 to S11

28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67

## Table of content

<b>SUPPLEMENTARY TEXT</b>	<b>4</b>
<b>1. Experimental design and biological material</b>	<b>4</b>
1.1 Choice of accessions from the 1001 Genomes resource	4
1.2 Greenhouse terminal drought experiment	4
1.3 Experiment under optimal growing conditions	5
<b>2. Drought phenotyping</b>	<b>5</b>
2.1 Image analysis pipeline	5
2.2. Drought index	5
<b>3. Quantitative genetics analyses</b>	<b>7</b>
3.1 Population structure	7
3.1.1 Association of genetic group membership with drought	7
3.2 Coalescent rates over time	7
3.3 Genome wide associations (GWA)	8
3.3.1 Linear Mixed Models (LMMs) with EMMAX	8
3.3.2 Bayesian Sparse Linear Mixed Models (BSMLMMs) with GEMMA	8
3.4 Multivariate analyses of phenotypes and GWA summary statistics	9
3.4.1 All pairwise correlations	9
3.4.2 Canonical Correlation Analysis (CCA)	9
3.5 Polygenic adaptation signal	10
3.5.1 Classic Qst-Fst comparison	10
3.5.2 Berg & Coop methodology	10
3.6 ChromoPainter and ancestry GWA	11
3.6.1 Global proportion of ancestral chromosome segments	11
3.6.2 aGWA for admixture mapping	11
3.6.2.1 Phylogeny of aGWA SNPs	13
3.7 Test for annotation enrichment	13
<b>4. Environmental and forecasting analyses</b>	<b>14</b>
4.1 Environmental data	14
4.2 Environmental Niche Models (ENM)	14
4.2.1 Geographic areas used and niche limits	14
4.2.2 Random Forest models	15
4.2.3 ENM of genetic groups	15
4.2.4 Genome Environment Models (GEMs)	16
4.2.4.1 Migration assumptions	16
4.2.4.2 Allele frequency change predicted by GEMs	17
4.2.4.3 Possible genetic trade offs of drought survival and flowering time	17
4.2.4.4 Population genetics simulations	18
4.2.4.5 Considerations regarding recombination	19

68	<b>4.2.4.6 GEM limitations</b>	<b>20</b>
69		
70	<b>REFERENCES</b>	<b>21</b>
71		
72	<b>SUPPLEMENTARY FIGURES &amp; MEDIA</b>	<b>26</b>
73	Figure S1. Extent of <i>Arabidopsis thaliana</i> distribution	26
74	Figure S2. Environmental ranges of <i>Arabidopsis thaliana</i>	27
75	Figure S3. Correlation between raw green pixels in plant images and model parameters	28
76	Figure S4. GWA with drought survival and population genetic statistics	29
77	Figure S5. Cross-coalescent rates between populations inferred by MSMC	30
78	Figure S6. Treemix with different migration rates	32
79	Figure S7. Genomic ChromoPainter chunks per population	33
80	Figure S8. Environmental niche model of genetic diversity and population structure	35
81	Figure S9. Environmental niche model of drought index	38
82	Figure S10. Environmental niche model of flowering time	40
83	Figure S11. Profile of phenotypic change under climate change	41
84	Figure S12. Maps of the most important climatic variables	43
85	Figure S13. aGWA Genome Environment Models (GEM)	45
86	Figure S14. GWA Genome Environment Models (GEM)	47
87	Figure S15. aGWA GEM residuals	49
88	Figure S16. GWA GEM residuals	51
89	Figure S17. Population genetics simulations	52
90	Video S1. Example of segmentation	53
91		
92	<b>SUPPLEMENTARY TABLES &amp; DATASETS</b>	<b>54</b>
93	Table S1. Accession information.	54
94	Table S2. ADMIXTURE cross-validation for all possible groups from 2 to 20.	54
95	Table S3. Selection signatures and annotation of top SNP hits from GWA.	54
96	Table S4. Annotation of top SNP hits from aGWA.	54
97	Table S5. Information on phenotypic and climate traits.	54
98	Table S6. Correlations between climate and phenotype variables per accession.	54
99	Table S7. Correlations between different GWA effects of the 150 polygenic SNPs.	54
100	Table S8. Canonical Correlation Analysis.	54
101	Table S9. Polygenic model at different top SNPs groups.	54
102	Table S10. Importance of variables in Random Forest analyses.	55
103	Table S11. Allele frequency change	55



104

105 SUPPLEMENTARY TEXT

## 106 1. Experimental design and biological material

### 107 1.1 Choice of accessions from the 1001 Genomes resource

108 The 1001 Genomes project has released resequencing data for 1,135 natural inbred lines, also  
109 called accessions (<http://1001genomes.org>). We applied several filters to select the most  
110 informative, least biased accessions for our experiment. (i) The first filter removed 176 accessions  
111 with low quality genome information, < 10X genome coverage and < 90% congruence of SNPs  
112 called from Max Planck Institute and Gregor Mendel Institute pipelines (1). (ii) The second filter  
113 removed 244 nearly-identical accessions, many from N. America. For this, we calculated pairwise  
114 genome-wide identity-by-state differences using PLINK v1.9 (2). When pairs differed in less than <  
115 0.01 changes per polymorphic site, we randomly removed one member of the pair. The overlap  
116 between (i) and (ii) was 762 accessions ([Fig. S1](#), [S2](#), Table S1). For geographic analyses in the native  
117 range (e.g. environmental niche models), we used the 729 accessions that were within 50°W to  
118 100°E longitude, i.e. Eurasia (see [section 4.2.1](#)). For the terminal drought experiment, we used 211  
119 of these 729 accessions. The seeds were progeny of the 1001 Genomes collection seed stocks  
120 obtained from the ABRC Stock Center (CS78942).

### 121 1.2 Greenhouse terminal drought experiment

122 The 211 accessions included both vernalization-requiring, slow-flowering and  
123 vernalization-independent, fast-flowering ones. Because the phenotype of interest in our study  
124 was the variance in survival during vegetative stages under terminal drought conditions, we  
125 applied a terminal severe drought without prior vernalization. Because of the difficulties  
126 associated with disentangling drought-induced mortality and reproduction-associated  
127 senescence at the end of the plant life cycle, our study focused on drought stress during the  
128 vegetative stage. (Note that onset of flowering, or flowering time, was not a confounding factor.  
129 See [section 2.2](#)).

130 Seeds were aliquoted in Eppendorf tubes, suspended in 1% agar solution, stratified in a  
131 4°C dark room for 5 days to promote germination, and then pipetted into pots filled with sieved  
132 soil (CL-P, Einheitserde Werkverband e.V., Deutschland). When multiple seeds germinated per  
133 pot, all but one were removed at random. We sowed 8 replicates per genotype in 49 trays of 8x5  
134 cells (5.5 x 5.5 x 10 cm) using a randomized incomplete block design. We excluded corner cells,

135 where edge effects are strongest.

136 During the first two weeks after sowing (defined as day 0), trays were watered close to soil  
137 saturation once every 3 days, with temperature maxima from 20 to 25°C under 16 hours natural  
138 and supplemental light. After this period, seedlings were challenged with a terminal drought, with  
139 “recovery waterings” after 3 and 6 weeks, in order to increase the variance in survival. The overall  
140 watering during the drought period (4 l in each tray of 40 x 60 cm), corresponded to  
141 approximately 33 mm of rainfall ( $4,000 + 4,000 \text{ cm}^3 \text{ water} / 2400 \text{ cm}^2 \text{ surface} = 3.3 \text{ cm}$ ). We  
142 monitored water content using moisture sensors (Parrot SA, Paris, France) (see water content  
143 graph in Fig. 1A). We monitored rosette green area by imaging at 20 time points (Fig. 1A) using a  
144 customized system (see below).

### 145 **1.3 Experiment under optimal growing conditions**

146 In a first experiment, we grew the same 211 genotypes under optimal watering and nutrient  
147 conditions and monitored vegetative growth by image analysis (Vasseur et al. submitted [*link to*  
148 *preprint or in press article will be added here*]) (see Table S5 for a description of the 24 traits  
149 extracted from the images). This set of traits was used to investigate whether variation of  
150 drought-survival index was correlated with growth under optimal conditions.

## 151 **2. Drought phenotyping**

### 152 **2.1 Image analysis pipeline**

153 Images were taken using a Panasonic DMC-TZ61 digital camera and a customized closed black box  
154 at a distance of 40 cm from the tray. This produced very consistent images in terms of  
155 illumination (only from in-camera flash) and focal distance to the plants.

156 We extracted leaf area per plant over time using the imaging module Open Computer  
157 Vision in Python (3) ([Video S1](#)), with these steps: (i) 5 pixel mean denoise of the whole-tray image.  
158 (ii) Fixed Hue Saturation Value (HSV) segmentation of “green” values. The threshold values were  
159 determined heuristically by comparing the HSV ranges of leaves at five timepoints (i.e. early and  
160 mature). (iii) Cropping of each pot to extract individual plant images. (iv) Counting of green pixels  
161 (for more specific details and code visit <http://github.com/MoisesExpositoAlonso/hippo>). Pots with  
162 green pixels but without plants were excluded after careful visual inspection of all images.

### 163 **2.2. Drought index**

164 After determining the peak of green area for the majority of pots, we modeled the daily

165 number of green pixels per pot. Several different models, including up to third order polynomial  
166 models, several error correction factors, either raw or genotype averages, were tested. All models  
167 were ranked based on parameter convergence in an MCMC walk and AIC values. The final chosen  
168 model was a generalized linear mixed model with Poisson link of the form:

$$169 \quad y = i + s t + q t^2 + \epsilon_{gi} + \epsilon_{gs} + \epsilon_{gq} + \epsilon_{tray} + \epsilon_{pos} + \epsilon$$

170 where green area,  $y$ , was the response variable, and an intercept ( $i$ ), slope ( $s$ ), and quadratic  
171 coefficients ( $q$ ) with time ( $t$ ) were fitted as fixed effects. Genotypes were treated as random  
172 factors, that is allowed to deviate from the main trends, following a normal distribution ( $0$ ,  
173  $\sigma_g$ ). Tray block and position within the tray grid were fitted as random factors following also a  
174 normal distribution. To estimate these parameters, we performed 10,000 iterations in a Monte  
175 Carlo Markov Chain (MCMC) and 1,000 burn-in using the glmmMCMC R package (4).

176 The variance from all genotype-dependent components relative to the total phenotypic  
177 variance was ~10%. This might be an underestimate since certain variance is due to inaccurate  
178 phenotyping: either optical and illumination distortion, or overlap of neighboring plants. Genotype  
179 values of the three parameters of interest (intercept, slope, and quadratic coefficient) were used  
180 for Genome Wide Association (GWA) and downstream analyses. Additive genetic variance was  
181 estimated from linear mixed models using a kinship matrix (see GWA section (3.3)). The intercept,  
182 slope and quadratic deviations had narrow-sense heritabilities ( $h^2$ , or kinship-associated variance)  
183 of 0%, 0%, and 50%; respectively. We chose the latter as the drought-survival index. This  
184 parameter informed about survival during the late stage of the experiment, as can be observed  
185 from a high correlation between the drought-survival index and the the raw green pixels in the  
186 final monitored days (Fig. S3).

187 Because the drought-survival index could depend on the developmental stage of the  
188 plants when the drought treatment started, we computed the pixel decay polynomial model with  
189 and without a covariate of flowering time under optimal conditions (indicative of developmental  
190 speed; see source of the phenotype in section 1.3 and phenotypic correlations in section 3.4.1).  
191 The small effect of flowering time in the model (0.01% of the variance) confirmed that the  
192 measurements were not biased, therefore we removed flowering time from the final model.

193 In order to provide an intuitive understanding of the drought survival index, we looked at  
194 the relationship between the index and the last day on which a plant was clearly alive, defined as  
195 the last day with at least 5,000 green pixels left. The relationship between the drought-survival  
196 index and the last living day was highly significant ( $p < 10^{-16}$ ). The most sensitive plants survived

197 for 32 days, and the most resistant were alive for 15 days longer.

### 198 **3. Quantitative genetics analyses**

#### 199 **3.1 Population structure**

200 From the vcf file containing SNP calls from the 1001 Genomes project (  
201 <http://1001genomes.org/data/GMI-MPI/releases/v3.1/>), we identified SNPs with a genotype calling  
202 rate >95%, which resulted in ~4M SNPs in 762 accessions ([section 1.1](#)). We defined genetic clusters  
203 with ADMIXTURE v1.2 (5) (Table S2). As a model-free alternative to ADMIXTURE, we used PCA  
204 implemented in PLINK v1.9 (2). The first three axes explained 16.0%, 9.6% and 7.9% of the total  
205 genetic variance. ADMIXTURE clustering and PCA were used to understand population structure  
206 and to relate it to phenotype variables. We assessed population splits and migrations with a  
207 population ancestral graph using TreeMix v. 1.12 (6), a tree based on genome-wide allele frequency  
208 differences across populations. Additionally, we calculated a proxy of local genetic diversity (7) per  
209 location using the number of polymorphic sites in the geographically closest genome.

##### 210 **3.1.1 Association of genetic group membership with drought**

211 Using the ADMIXTURE membership probabilities of each genome, we carried out univariate linear  
212 regressions with the drought survival phenotype. The groups that yielded positive relationships  
213 were NE Spanish ( $p < 0.05$ ), Mediterranean ( $p = 0.06$ ), and the N. Swedish groups ( $p < 0.001$ ). The  
214 groups negatively associated were Central Europe ( $p = 0.06$ ), Asia ( $p < 0.001$ ), and E. Europe  
215 ( $p < 0.001$ ). This broadly coincided with the map of drought-survival prediction (Fig. 1D, [S11](#)). We  
216 found that only PC3 was significantly associated with the drought survival phenotype (GLM  
217  $R^2 = 0.076$ ;  $p = 5.15 \times 10^{-5}$ ). The N. Swedish and NE Spanish groups showed particularly low values in  
218 PC3 than the rest ([Fig. S8](#)).

#### 219 **3.2 Coalescent rates over time**

220 Only the accessions with  $\geq 90\%$  of membership probability in one of the genetic groups were  
221 used. Using MSMC software (8) (<http://github.com/stschiff/msmc>), we performed within- and  
222 cross-genetic group coalescent analyses by contrasting two pairs from different genetic groups. In  
223 total 333 analyses were performed, with each genetic group being tested at least 3 times. The  
224 results were summarized using a smoothed generalized additive model in R (Fig. 2C).

## 225 3.3 Genome wide associations (GWA)

### 226 3.3.1 Linear Mixed Models (LMMs) with EMMAX

227 We used 879,654 biallelic SNPs with a minimum allele frequency (MAF) of 5% for genome wide  
228 association (GWA) using EMMAX (9). We carried out GWA for all climatic variables and 11  
229 phenotypes (Table S5). The GWA is based on linear mixed models that test, one by one, each of  
230 the SNPs, and correct the results by population structure using a random factor with a  
231 variance/covariance kinship matrix built from genome-wide SNPs. This is an appropriate method  
232 to correct for coancestry in *Arabidopsis thaliana* (10).

233 To rule out the possibility that drought survival measurements were dependent on the  
234 developmental stage of the plant during the experiment, we carried out the GWA with and  
235 without a covariate of flowering time that had been scored in controlled conditions (proxy of  
236 developmental speed; [section 1.3](#)). The top SNP hits were the same with or without this covariate,  
237 and we only show results without the covariate. To account for familywise error in GWA we used  
238 Bonferroni correction ( $p$  value  $\times$  number of SNPs) and the Benjamin-Hochberg false discovery  
239 rate correction (11). The kinship-associated variance of drought-survival – an approximation of  
240 narrow sense heritability,  $h^2$ , was 49%. When we fit a kinship calculated from only the 151 top  
241 polygenic GWA SNPs (see [section 3.5.2](#)), the estimate of  $h^2$  was 52%. This is probably a better  
242 estimate than that from the genome-wide-based kinship matrix, as the putatively causal SNPs are  
243 better “tagged” in the 151 SNPs kinship matrix.

244 After calculating genome-wide  $F_{st}$  (12) and Tajima’s D (13) with PLINK v1.9 (2) and  
245 likelihood of a selective sweep with SweeD (14), we investigated the enrichment of the top SNPs  
246 in the upper tail of the distributions of those statistics ([Fig. S4](#)) (Table S3, rank columns).

### 247 3.3.2 Bayesian Sparse Linear Mixed Models (BSLMMs) with GEMMA

248 The Bayesian Sparse Linear Mixed model (BSLMM) implemented in GEMMA (15) accommodates  
249 both poly- and oligogenic architectures in a GWA framework. It models two effect  
250 hyperparameters, a basal effect, *alpha*, that captures the fact that many SNPs contribute to the  
251 phenotype, and an extra effect, *beta*, that captures the stronger effect of only a subset of SNPs.  
252 The parameter measuring the probability of having another extra effect, *gamma*, can be used to  
253 prioritize SNPs (personal instructions from the author X. Zhou). Over 40% of the top 151 SNPs  
254 from EMMAX were found to have over 99% percentile of the gamma inclusion probability in  
255 GEMMA (Fisher’s exact test odds ratio = 17.21,  $p=3 \times 10^{-7}$ ). The estimate of realized heritability with  
256 BSLMM was 50%, which is in agreement with the EMMAX analyses. The 95% highest posterior

257 density (95%HPD) from 1,000 MCMC steps ranged from 25-85%.

### 258 **3.4 Multivariate analyses of phenotypes and GWA summary statistics**

259 For a description and sources of all variables used, see Table S5.

260

#### 261 3.4.1 All pairwise correlations

262 We computed all-against-all Pearson product-moment correlation coefficients among accession  
263 line means (n= 211 accessions) of phenotypic and climate variables (Table S5, S6). To study genetic  
264 correlations, we performed the same analyses with SNP effect sizes (n= 151 drought-associated  
265 SNPs) estimated from multiple GWA (Table S7).

266 The phenotype correlations (Table S6) showed that the drought-survival index was  
267 negatively correlated with reproductive allocation and number of seeds ( $r < -0.16$ ,  $p < 0.02$ ),  
268 suggesting a fitness trade-off between stressful and optimal growth environments.  
269 Drought-survival was not correlated with flowering time ( $r=0.07$ ,  $p=0.12$ ) nor plant size (rosette  
270 area and dry mass,  $r < 0.12$ ,  $p > 0.07$ ).

271 Drought-survival SNP effects negatively correlated with the SNP effect sizes of most  
272 precipitation variables separately ( $r < -0.4$ ;  $p < 10^{-8}$ , Table S7), indicating that alleles that increased  
273 drought survival were found in more arid geographic regions, i.e. regions with high temperatures  
274 and lower precipitation at different times of the year. Drought-survival SNP effects were also  
275 positively correlated with SNP effects of rosette area, dry mass, and flowering time (Table S7).  
276 These analyses have two-fold interests: (1) GWA-estimated effect have been corrected by  
277 population structure, thus correlations should not reflect phenotypic differences caused by drift of  
278 populations. (2) SNPs can have pleiotropic effects and this can limit adaptation due to genetic  
279 constraints (see [section 4.2.4.3](#)) (16).

#### 280 3.4.2 Canonical Correlation Analysis (CCA)

281 We further utilized Canonical Correlation Analysis (CCA) to decompose environment-phenotype  
282 associations of SNP effects. This was done for all genome-wide SNPs (n=800,000) and for the  
283 151 drought-associated SNPs (Table S8).

284 CCA of genome-wide SNPs revealed the first canonical correlation axis (CC1) to be driven  
285 by lower flowering time (T\_repro, loading=-0.77), lower rosette dry mass (loading=-0.76) and  
286 higher annual temperature (bio1, loading=0.5). CC2 indicates that lower plant photosystem stress  
287 (FvFm, loading=0.60) is related to higher mean temperature of the wettest quarter and higher



288 precipitation seasonality (bio8, bio15, loadings>0.25). CC3 shows that lower drought survival  
289 (loading=-0.58) effects are related to higher precipitation in the driest (bio17, loading=0.44) and  
290 warmest quarters (bio18, loading=0.35).

291 CCA of 151 top GWA SNPs yielded a first canonical correlation coefficient of 0.99, with a  
292 phenotype canonical variate driven by lower drought survival, higher rosette area and dry mass  
293 (loadings >0.75), and a climatic canonical variate dominated by higher precipitation during the  
294 wettest month (bio13) and wettest quarter of year (bio16) (loadings >0.75).

### 295 **3.5 Polygenic adaptation signal**

#### 296 **3.5.1 Classic $Q_{st}$ - $F_{st}$ comparison**

297  $Q_{st}/F_{st}$  ratios have been proposed as an appropriate indicator of local adaptation in *A. thaliana* (17)  
298 and related species (18). Genome-wide  $F_{st}$  across the eleven population was computed from 211  
299 genomes using vcfTools (v0.1.12b) (19). We estimated the mean and confidence intervals based on  
300 the standard error of the mean, obtaining a mean  $F_{st} = 0.042$  (95% cumulative distribution =  
301 0.360). We calculated  $Q_{st}$  for the drought-survival index as the between-genetic group variance  
302 divided by the total variance. We used the MCMCglmm function in R with a 10,000 steps chain,  
303 1,000 steps burn-in, and fitting the genetic group as random effect. This resulted in a  $Q_{st} = 0.143$   
304 (90%HPD = 0.052 - 0.338). When the variance across populations was done using the NE  
305 Spanish and the NSweden (population groups hypothetically under local adaptation) against the  
306 rest, we obtained  $Q_{st} = 0.377$  (0.047 - 0.987). We thus concluded that a significant  $Q_{st} > F_{st}$   
307 signal is only observable at the individual level when the hypothetical populations that underwent  
308 local adaptation were oriented in the calculation of the variance.

#### 309 **3.5.2 Berg & Coop methodology**

310 We tested for a polygenic adaptation signature following Berg & Coop (20), an extension of the  
311  $Q_{st}/F_{st}$  ratio test based on SNP frequency per population and effect sizes as estimated from a  
312 GWA analysis. We used different groups and numbers of ranked SNPs after pruning linked SNPs  
313 ( $r^2 > 0.6$ ), to learn about the robustness of this test and the apparent number of SNPs that  
314 contribute to the signal (Table S9). Since this test does not use direct phenotypes but calculates  
315 the average phenotype per population based on allele frequencies of GWA SNPs, we could  
316 perform the test with 762 high quality accessions. Since results did not vary between 762 and 211  
317 sample analyses, we only report the analyses with the 762 genomes (Table S9).



### 318 **3.6 ChromoPainter and ancestry GWA**

319 We ran ChromoPainter version 2.0.7 (available at <http://paintmychromosomes.com>) (21) on the  
320 762 genomes dataset, after imputing missing genotypes with Beagle version 3.3.2 (22) using  
321 default parameters.

322 ChromoPainter analyses require a “training” run to estimate several hyperparameters. We  
323 ran 10 expectation maximization iterations on chromosome 2 (the smallest chromosome). We  
324 informed ChromoPainter with a published recombination map of *Arabidopsis thaliana* (23) that we  
325 reshaped to our SNP dataset. We used the command:

```
326 ChromoPainterv2 -i 10 -in -iM -j -g haplotypefile -r recombinationfile -a 0 0 -t labelfile
```

327 We used the output hyperparameters to run ChromoPainter on all chromosomes in an  
328 unsupervised all-to-all genomes mode, with the command:

```
329 ChromoPainterv2 -n 4.737068 -M 0.000421 -j -g haplotypefile -r recombinationfile -a 0 0 -t  
330 labelfile
```

#### 331 **3.6.1 Global proportion of ancestral chromosome segments**

332 To study the ancestry relationships of each of the genetic groups, we counted the number  
333 of chromosomal segments (termed “chunks” in the original ChromoPainter paper (21)) that each  
334 genome “received” from all other genomes. The segment varied in size depending on local  
335 recombination rates and between genomes, but *a posteriori* analyses indicated that the median  
336 size was in the order of magnitude of kilo and megabases. To make the counts more  
337 informative, we show boxplots per ADMIXTURE group rather than counts per individual (Fig. S7  
338 A-K). This showed, for example, that NW Spain, NE Spain and S. Mediterranean (the latter to a  
339 lower degree), were “painted” mostly by relict DNA segments. Next, we tried to infer how well the  
340 drought survival of an individual correlated with the number of segments inherited from a certain  
341 ancestry. This indicated that only N. Sweden and relicts passed DNA segments that were  
342 correlated with the drought-survival index of the receiving individual (Fig. S7L). The Pearson  
343 correlation coefficient was calculated excluding the individuals from the same admixture group as  
344 the predictor.

#### 345 **3.6.2 aGWA for admixture mapping**

346 If populations are locally adapted,  $F_{st}$  outlier scans can be used to identify genetic variants under  
347 divergent selection (24, 25). However, when populations get isolated and diverge genetically, as it

348 is the case in *Arabidopsis thaliana*,  $F_{st}$  values are shifted to high values across the entire genome  
349 even when subsequent admixture happen, making the identification of outliers difficult (Fig. S4)  
350 (24). Thus, we must rely on linkage disequilibrium and identity by descent to find DNA segments  
351 characteristic of the different populations. If subsequent but incomplete admixture occurred  
352 between the locally adapted populations, it is expected that the individuals that retained the DNA  
353 segments responsible for local adaptation, would show the largest phenotypic differences. This is  
354 the principle of admixture mapping (26).

355 With the above rationale, we developed an admixture mapping technique (26)  
356 repurposing the output of ChromoPainter. The “painted” genome matrix produced by  
357 ChromoPainter has 762 states (one per individual in the analysis) and we repainted it into a  
358 genome matrix of 11 states (the genetic groups from ADMIXTURE analysis, which are  
359 geographically and environmentally separated). We then computed a regression of the  
360 drought-survival phenotype on the population group specific to a SNP as:

$$361 \quad y_i = \mu + Ab + \epsilon$$

362 Where  $y$  is a vector of  $i=1...211$  individual's phenotypes,  $\mu$  is the mean phenotype,  $B$  is the 211  
363 x 11 design sparse matrix of the ancestry states,  $b$  is a 1 x 11 vector of effects of each ancestry has in  
364 the mean phenotype, and  $\epsilon$  is the uncorrelated random residuals assumed to be normal. This  
365 model was repeated for each SNP in our dataset (~2 million imputed and ‘painted’ SNPs, see  
366 [section 3.6](#)). We report R square of and p-value of each SNP model (Table S4). Since we already  
367 knew that the phenotype is associated with the membership assigned per individual, we expected  
368 that the membership of any random SNP would be on average also associated, because of linkage  
369 resulting from common ancestry. Therefore, we implemented an empirical p-value distribution  
370 correction to only detect those SNPs whose ancestry explained an even larger proportion of  
371 variance than the whole-genome ancestry. The permutation was done within each individual  
372 genome shuffling the SNP states at a distance of 1,000 to 10,000 SNP positions — defined from  
373 analysing the typical size of “homogeneously painted DNA segments” (code are available at  
374 <http://github.com/MoisesExpositoAlonso/aGWA>). We permuted the dataset 1,000 times and  
375 repeated this “aGWA” analysis to build p-value distributions. Since the nature of the associations is  
376 very different from that of a standard GWA analysis, we did not expect and did not find any  
377 overlap of top aGWA SNPs with the top SNPs from regular GWA. The closest was a regular GWA  
378 SNP that was 8 kb away from an aGWA SNP. The closest gene to both encodes a defensin-like  
379 protein; a family of proteins with broad anti-fungal and anti-bacterial activity (27).

380 Our approach is conceptually related to admixture mapping in humans, which has focused

381 on local enrichment of Neandertal- and Denisovan-like variants, and which has led to the  
382 identification of in TLR immunity genes (28) as adaptive. It has also helped to increase the power  
383 for detection of background-dependent disease risk in humans with mixed ancestries, e.g.  
384 African-American individuals (29), or other more complex mixtures (30). Such approaches  
385 constitute a powerful tool for understanding the genetic basis of local adaptation when complex  
386 demographic scenarios of admixture exist.

### 387 3.6.2.1 Phylogeny of aGWA SNPs

388 To learn about the distribution and shared ancestries of the drought-related alleles, we computed  
389 a neighbour joining phylogeny of all concatenated SNP hits from aGWA ( $p < 0.001$ ) and compared  
390 it with a genome background phylogeny of 1,000 randomly chosen SNPs (Fig 3B). This revealed  
391 that the N. Swedish groups and Mediterranean relicts were across the genome more distant from  
392 each other than the average between-group distance (Student's t test,  $p < 2 \times 10^{-16}$ ) (Fig. 3B), much  
393 closer than the average when considering only the aGWA SNPs (Student's t test,  $p < 2 \times 10^{-16}$ ). The  
394 same analyses showed also higher affinity of N. Swedish and NE Spanish populations (Student's t  
395 test,  $p < 10^{-10}$ ).

## 396 3.7 Test for annotation enrichment

397 Using the TAIR10 gene annotation of *Arabidopsis thaliana* (available at  
398 [arabidopsis.org/portals/genAnnotation/functional\\_annotation/](http://arabidopsis.org/portals/genAnnotation/functional_annotation/)), we tested whether there was a  
399 specific annotation class enriched in our GWA and aGWA hits. The most suggestive genes  
400 overlapping with the 151 top GWA hits were the nitrate transporter gene *NRT1.8*, which among  
401 other functions mediates cadmium tolerance and is related to ABA transport (31–33), the  
402 *CATION/CARNITINE TRANSPORTER 4 (OCT4)*, which mediates homeostasis of metabolites and  
403 promotes lateral root formation (34), and the sugar transporter gene *SWEET8*, which is  
404 upregulated during salt stress (35). On the other hand, the strongest peak fell inside the *CATION*  
405 *EXCHANGER 9*, a gene important for homeostasis of  $K^+$ ,  $Na^+$  and  $Mn^{++}$  that confers salt tolerance  
406 when introduced into yeast (36). An empirical distribution test based on random draws of genes  
407 showed, however, only marginal enrichment. The 30 genes defined by the 151 top GWA SNPs  
408 were weakly enriched for cell membrane transport (6/30;  $p = 0.01$ ), and the 23 genes defined by  
409 the 70 top aGWA SNPs were only very marginally enriched for membrane transport (7/23;  
410  $p = 0.06$ ). Testing for overrepresentation with PANTHER ([www.pantherdb.org](http://www.pantherdb.org)) and including genes  
411 adjacent to the GWA and aGWA SNPs revealed weak enrichment of aGWA genes for ferredoxin  
412 metabolic process ( $p = 0.03$ ) and vesicle-mediated transport ( $p = 0.05$ ), and of GWA genes for  
413 growth-related functions ( $p = 0.0007$ ) and metabolite biosynthetic processes ( $p = 0.0002$ ). It is

414 difficult to know what to conclude from this, but the most noteworthy finding is probably that  
415 there was no link to flowering time, in contrast to previous QTL and GWA studies of *A. thaliana*  
416 response to drought (37–39).

## 417 **4. Environmental and forecasting analyses**

### 418 **4.1 Environmental data**

419 The environmental data comprised the Last Glacial Period (LPG, ~22,000 years ago), recent  
420 averages from 1960-1990, and two 2070 climate projections of contrasting socio-economic  
421 scenarios, the 2.6 and 8.5 CO<sub>2</sub> representative concentration pathways (rcp) (40, 41). The data was  
422 retrieved from [www.worldclim.com](http://www.worldclim.com) v.1.4 (ref. (42)). It consists of 19 bioclimatic variables at 2.5  
423 minutes geographic resolution (code to retrieve and process data available at  
424 <http://github.com/MoisesExpositoAlonso/rbioclim>).

### 425 **4.2 Environmental Niche Models (ENM)**

426 We carried out ENM with a number of response variables (for summary statistics see Table S10),  
427 namely the drought-survival phenotype, flowering time, the genomic principal component axes,  
428 the discrete population groups, the local genetic diversity, and the SNPs identified in GWA and  
429 aGWA analyses .

#### 430 4.2.1 Geographic areas used and niche limits

431 To train ENM, we removed accessions from Japan and from N. America, as they are recent  
432 introductions (43) and might not reflect long-term climate adaptation. In addition, the sampled  
433 locations used to trained the models were within 15 to 63° N and 23°W to 88°E longitude, but we  
434 only predicted in a reduced area, from 34 to 63° N and -10.5 to 35°E, to avoid extrapolation of  
435 data. Predictions for the last glacial maxima were masked in those areas that were likely tundra or  
436 ice sheet at the time (<5°C and <0°C annual temperature, respectively), as they are presumably  
437 artifacts.

438 Because the sampling in the 1001 Genomes project was not even across the species  
439 range, predictions for underrepresented regions such as N. Africa, the Middle East, or Russia, must  
440 be taken with caution. In order to be explicit about for which areas we could make the most robust  
441 predictions, we show the sampling density per 1° x 1° latitude x longitude grid, which varies from  
442 around 1 to 60 individuals (Fig. S8D), and plot trends of predicted values against other variables,  
443 such as latitude or climate variables (e.g. Fig. S9-S11), only at those locations where there is at least

444 one sample.

445 Finally, it is worth noting that even for the most pessimistic climate change scenario (rcp  
446 8.5), the values of annual precipitation (bio1) and the precipitation during the warmest season  
447 (bio18) were always above the present minimum precipitation values where *A. thaliana* is currently  
448 found (see [Fig. S12](#)). Therefore, we expect that transgressive phenotypes are not required to  
449 survive future climates.

#### 450 4.2.2 Random Forest models

451 After trying different methodologies, including generalized linear models, MaxEnt, and  
452 linear discriminant models, we opted for random forest models because they are nonparametric,  
453 nonlinear, allow both continuous and discrete response variables, and are computationally  
454 efficient (44). Additionally, the implementation of an “importance” parameter of each predictor  
455 variable available in the *randomForest* R package makes ranking of variables straightforward. To  
456 mitigate the overfitting problem typical of machine learning methods, a 5-fold cross validation  
457 procedure was used. We randomly divided the dataset in five parts, used four parts as training  
458 dataset and one part as testing dataset, and repeated this five times. Reported accuracy from  
459 cross validation was the  $R^2$  of a linear model between observed and predicted values for  
460 continuous variables, and the rate of successful assignment of categories relative to the total  
461 number of observations for discrete variables. To build the final forest, a total of 50 classification  
462 or regression trees per cross-validation set were used, and six variables were tested for each  
463 classification split.

#### 464 4.2.3 ENM of genetic groups

465 We modeled the presence of population structure as a discrete response variable in ENM;  
466 either using eleven genetic groups as states, or the two relict and non-relict states.

467 In order to formally quantify the relevance of genetic group membership, we calculated  
468 the percentage of map grid cells that each genetic group occupies. For this we only considered  
469 areas where at least one genome per 1° x 1° latitude x longitude was observed ([Fig. S8A](#)) and where  
470 tundra or ice sheet are not expected (important for LGM comparisons).

471 When we used the present-data trained relict/non-relict ENM with past climate data from  
472 the last glacial maxima, we found that relicts likely occupied almost a quarter of the non-glaciated  
473 areas, compared to less than 2% today ([Fig. S8D](#)), in agreement with genomic inferences of  
474 higher effective population size in the past ([Fig. 2C](#)). The reason that the relicts' environmental  
475 niche is predicted with 100% accuracy under 5-fold cross-validation (5CV) is likely that the local

476 number of relicts individuals is low, 26 accessions out of 762, and because their niche is very  
477 restricted.

478 Under a future high CO<sub>2</sub> increase socio-economic scenario, the ENM with 11 genetic  
479 groups predicts that the S. Mediterranean group will expand most dramatically into  
480 Central-European areas, replacing groups currently occupying these areas (Fig. S8 C, F). Although  
481 these models are not mechanistic, they illustrate that genetic groups from the Mediterranean and  
482 from temperate areas have contrasting environmental niches and thus might replace each other  
483 under future climate warming.

#### 484 4.2.4 Genome Environment Models (GEMs)

485 All 151 GWA and 70 aGWA SNPs were modeled as a bivariate discrete variable (drought-sensitive  
486 and drought-survival allele) in a random forest. The prediction accuracy and the most important  
487 predictor for each model below is shown per SNP in Table S3 and S4.

488 After modeling presence/absence of each drought-survival SNP, we projected the present  
489 inferred allele distributions in a map and then summarized all maps by intersecting them. In this  
490 way, we generated a continuous map surface of the total number of drought-resistant alleles in a  
491 given location. Ancestral GWA and GWA models showed overall similar patterns (Fig. S13-14), but  
492 the latter were more biased to Western areas. This is likely due to GWA SNPs suffering from  
493 high-frequency bias, making them more likely to be present in geographic areas with more  
494 samples (Fig. S8). After we had trained the models with present data, we used them to predict  
495 allele distributions in 2070 under low and high CO<sub>2</sub> increase scenarios. While patterns were  
496 similar in both scenarios, for further analyses we used the most “pessimistic” high CO<sub>2</sub> increase  
497 scenario to be able to show main trends more clearly.

##### 498 4.2.4.1 Migration assumptions

499 For each SNP we trained three models in order to overcome the “universal (or free) migration”  
500 assumption, implicit when using a current climate-trained ENMs with future climate data (e.g.  
501 (45)). Although typically the free-migration model may not be entirely appropriate for predictions,  
502 it might be more realistic for cosmopolitan species with continental-scale migrations in the recent  
503 past, as is the case of *A. thaliana* (43). Nevertheless, we designed two additional models to  
504 account for limited migration. The free model includes only the 19 bioclimatic variables as  
505 predictors of the drought-survival alleles. The first geographically-controlled model includes in  
506 addition the first three PC genomic axes as predictors (Fig. S8G-I), in an attempt to limit  
507 prediction allele presence to geographic areas where the genomic background that they reside on



508 is present today. The second geographically-controlled model, which is even more restrictive,  
509 includes latitude and longitude together with the 19 bioclimatic variables. For all models we not  
510 only show the predicted maps (Fig. S13-14) but also provide residuals of predicted vs observed  
511 (empirical) number of alleles in the locations where we have a sample. We also show their  
512 relationship with latitude (Fig. S15-16).

#### 513 4.2.4.2 Allele frequency change predicted by GEMs

514 We took 40 individuals approximately within 50 km of each other at three locations with the  
515 highest density of samples in our dataset: Madrid (Spain), Tübingen (Germany) and Malmö  
516 (Sweden) (Fig. 3C, Fig. S8). We tested overall future allele frequency changes of all SNPs per  
517 population as well as SNP-specific allele frequency changes.

518 First, we calculated the mean allele frequency differences between future (rcp 8.5, 2070)  
519 and present predictions. This proved to be significant in most locations and models (Table S11),  
520 although the direction of change was different between the two edge populations, Madrid and  
521 Malmö, and the Tübingen population from the center of the range. The former showed a decrease  
522 or a steady state in allele frequency, and the latter showed a highly significant increase in all  
523 models and SNP subsets (Table S11).

524 Secondly, we calculated the differences in frequency ( $F$ ) between present (*pres.*) and  
525 future (2070) populations per SNP and tested the difference using a Student's t test and a pooled  
526 standard error ( $se$ ) of the frequency measurements:

$$527 \quad t = \frac{F_{SNP\ 2070} - F_{SNP\ pres}}{\sqrt{se_{SNP\ 2070}^2 + se_{SNP\ pres}^2}}$$

528 This not only revealed the main frequency change trend, but also the distribution of  
529 differences in alleles (see histograms in Fig. S13-14). It corroborated the general trend observed for  
530 all SNPs (Table S11) and in addition showed that the global distribution of allele frequency changes  
531 in Tübingen is skewed to the right in some SNPs (increase of drought allele frequency).

#### 532 4.2.4.3 Possible genetic trade offs of drought survival and flowering time

533 Contrary to our expectations, there were areas in the Mediterranean that were predicted to lose  
534 drought-survival alleles under climate change (Fig. S9-10). These are areas that already suffer  
535 today from low precipitation (reached the zero in summer, Fig. S12) and will probably not become  
536 much drier in summer. On the other hand, temperatures will keep increasing, which likely will



537 demand an acceleration of flowering time (in trade-off with drought avoidance). Predictions at the  
538 phenotypic level (Fig. S9-10) showed this trend: drought-survival will increase only in the  
539 transition areas from Mediterranean and temperate regions (Fig. S9) and might decrease in areas  
540 that are already dry (Fig. S11). On the other hand, flowering time was predicted to decrease in the  
541 Mediterranean (Fig. S10-11). We note that the SNP effects on drought survival and flowering time  
542 were positively correlated, as disclosed by Canonical Correlation Analyses (section 3.4).

#### 543 4.2.4.4 Population genetics simulations

544 The predicted an allele in 2070 does not directly inform about the actual possibility of adaptation.  
545 This depends on (i) the frequency of the alleles and haplotypes in the population, (ii) the  
546 recombination rate, and (iii) the strength and efficiency of selection. Indeed, geographic  
547 predictions of alleles are probably bad predictors of allele frequency because random forest  
548 models tend to predict either one allele or another in a certain region. That is why we do not  
549 compare empirical present allele frequencies with frequency calculated from future predicted  
550 presence of alleles in different locations of Tübingen, Madrid and Malmö.

551 To obtain further insights into population dynamics required for adaptation, we simulated  
552 allele frequency changes in a Wright-Fisher population under a mutation-selection balance with  
553 inbreeding, as *Arabidopsis thaliana* is a selfer (code available at  
554 <http://github.com/MoisesExpositoAlonso/popgensim>). We started the simulations with the  
555 present frequencies of drought-related alleles of the 221 aGWA/GWA SNPs, with (codominant)  
556 selection coefficients ( $s$ ) ranging from 0.01 to 20% fitness advantage. We considered SNPs as  
557 independent, that is, we did not include linkage disequilibrium information nor a recombination  
558 rate (see next section).

559 We carried out forward-in-time simulations for 50 generations, the approximate number  
560 of generations that natural populations of *A. thaliana* from today until 2070 at an average  
561 generation time of around 1.5 years (46). We assumed a mutation rate ( $\mu$ ) calculated from  
562 laboratory mutation accumulation lines (43, 47) and a selfing coefficient ( $\psi$ ) of 99%, a  
563 conservative lower bound estimate from natural populations' heterozygosity (48). The population  
564 size ( $N$ ) was estimated from the genomic diversity in our dataset: The 40 genomes within  
565 Tübingen area had a genome-wide nucleotide diversity of 0.004 (section 3.1). With the equation:

$$566 \quad 4N_e \times \mu = \pi$$

567 ,we solved for effective population size ( $N_e$ ) and transformed it into population size

568 following the the relationship (49):

$$569 \quad N_e = \frac{1}{1 + F} \times N = \frac{2 - \psi}{\psi} \times N$$

570 This yielded a  $N = 300,000$  plants, which might be reasonable given that we consider an  
571 area of 50 km around Tübingen.

572 After running the simulations, we asked what selection coefficient would be needed to  
573 reach quasi-fixation frequencies of each allele (>0.9 frequency) or to match the drought-allele  
574 frequencies in Madrid or Malmö (assuming that those populations are adapted in comparison with  
575 the Tübingen one). When a specific allele frequency was higher in the Tübingen population than in  
576 Madrid or Malmö, we assumed selection would not be necessary and the coefficient was assumed  
577 zero. The results indicated that selection coefficients should be strong (but see (50)) for alleles to  
578 become fixed (Fig. S17). However, the distribution of selection coefficients were centered around  
579 1-3% fitness advantage for Tübingen allele frequencies to match Malmö or Madrid (Fig. S17) (but  
580 see next section 4.2.4.5). We did not simulate different degrees of drift in our analyses as when  
581 the inequality:  $N_e s > 1$ , holds, the weight of drift relative to selection is thought to be  
582 imperceptible (7).

#### 583 4.2.4.5 Considerations regarding recombination

584 As stated above, assessing whether a population can adapt depends on the frequency of  
585 drought-resistant individuals in the population, the rate of recombination to shuffle advantageous  
586 alleles, and the strength of selection. In our simulations above we did not work with haplotypes of  
587 SNPs in linkage disequilibrium as they are found in individuals, but considered SNPs to be  
588 independent. This can be seen *a priori* as a strong assumption. Simulations including whole  
589 haplotypes could inform about processes such as Hill-Robertson effect, hitchhiking, or  
590 background selection, which could be achieved with more complex approaches in the future (51).

591 Of relevance is that even in Tübingen there are already some individuals that already have  
592 many of the 151 GWA drought-resistant alleles, with one exceptional individual having 123/151  
593 drought-survival alleles. The three next best individuals have 107, 105, and 99 alleles. Fifteen of  
594 the 28 drought-resistance alleles are not present in the Tübingen population and will have to be  
595 imported by migration. Therefore, to produce a hypothetical “fully adapted haplotype” with 136  
596 alleles from the current standing variation (123 alleles are 90% of all 136 present alleles), only 13  
597 drought-resistant alleles would have to be recombined and introgressed in the already  
598 advantageous haplotype. Such introgression events might not be limited by low frequencies of

599 the advantageous alleles, as some were found at intermediate or as high as 90-95% frequency.  
600 Furthermore, in a scenario with a haplotype in the population with 123 alleles already present,  
601 simple individual-based simulations show that already with selection coefficients on the order of  
602 0.5% advantage per allele, the 123 alleles haplotype will become completely fixed in the  
603 population within 50 generations (simulations not shown). Results of aGWA indicate similar  
604 patterns (24 alleles are 63% of all 32 present alleles) but more alleles are missing in the Tübingen  
605 population, as their frequency is lower and geographic distributions are narrower than GWA  
606 alleles.

607 We also used a series of approximate calculations to ascertain how many recombination  
608 events are required to generate a hypothetical “fully adapted haplotype”. In *A. thaliana*, there are  
609 on average 1.4 meiotic crossovers for each of the five chromosomes (52). Together with  
610 independent segregation of the five chromosomes, the parental haplotypes are rearranged at  
611 around 12 positions. A population of ~300,000 individuals (N) with a lower bound outcrossing  
612 rate of 1% (=1-F) over 50 generations (g), could thus undergo around ~2 million recombination  
613 events (=  $N * (1-F) * r * g$ ), or about 1 event per 50 bp. Note that this could be a conservative  
614 estimate as rates of outcrossing in geographically close plants can be above 10% (53). This result  
615 suggests that recombination might be less limiting than expected *a priori*.

#### 616 4.2.4.6 GEM limitations

617 There is a long list of factors that we did not take into account and that will influence future plant  
618 response to climate change. We briefly enumerate them here:

- 619 A. We only focus on adaptation to drought, but other environmental stresses could have  
620 similar detrimental effects such as extremely high temperatures or ecosystem destruction.  
621 In addition, fluctuation in selection gradients and seasonal environmental variation are  
622 other possible consequences of climate change (54, 55).
- 623 B. We only can explain ca. 50% of the drought survival variance with 221 SNPs.
- 624 C. We evaluated drought survival in a controlled greenhouse experiment, but the  
625 extrapolation to natural conditions may be difficult. This would require field experiments  
626 assessing fitness *in situ* to confirm that the identified SNPs actually report a fitness  
627 advantage (56).
- 628 D. Because the high narrow sense heritability suggests a mostly-additive genetic architecture,  
629 we carry out predictions with allele counts. However, we acknowledge that there is  
630 variation of the magnitude of the SNP effects (Table S3-4), and epistatic effects might  
631 exist.

- 632 E. All of our predictions are based on existing diversity, but de novo mutations are likely to  
633 make a contribution as well, especially in species with high reproduction rate, short  
634 generation time, and large population sizes (43, 57, 58).
- 635 F. Biotic interactions can also play a relevant role of population dynamics, which we ignored  
636 (59–61).
- 637 G. In addition, although long-term evolution should be driven by genetic adaptation, it is  
638 expected that phenotypic plasticity will partially buffer the detrimental consequences of  
639 environmental change (62).
- 640 H. The existence of a seed bank in *A. thaliana* (46, 63) would cause longer generation times  
641 and overlapping generations, and both alter the speed and dynamics of allele frequencies  
642 (64).
- 643 I. Although our rough calculations suggest that recombination would not be a limitation for  
644 future adaptation in *A. thaliana* populations, we have not incorporated such processes in  
645 our modeling, as it is not a trivial matter (51, 65). This ignores phenomena such as  
646 background selection or hitchhiking effects that could arise from phenotypic trade-offs  
647 and the currently realized composition of haplotypes in the population.

## 648 REFERENCES

- 649
- 650
- 651 1. 1001 Genomes Consortium, 1,135 Genomes Reveal the Global Pattern of Polymorphism in  
652 *Arabidopsis thaliana*. *Cell*. **0** (2016), doi:10.1016/j.cell.2016.05.063.
  - 653 2. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage  
654 analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
  - 655 3. G. Bradski, The opencv library. *Doctor Dobbs Journal*. **25**, 120–126 (2000).
  - 656 4. J. D. Hadfield, MCMC methods for multi-response generalized linear mixed models: the  
657 MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
  - 658 5. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated  
659 individuals. *Genome Res.* **19**, 1655–1664 (2009).
  - 660 6. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide  
661 allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
  - 662 7. B. Charlesworth, D. Charlesworth, *Elements of Evolutionary Genetics* (Roberts and Company  
663 Publishers, 2010).
  - 664 8. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple  
665 genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
  - 666 9. H. M. Kang *et al.*, Variance component model to account for sample structure in  
genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).

- 667 10. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana*  
668 inbred lines. *Nature*. **465**, 627–631 (2010).
- 669 11. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful  
670 approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.*, 289–300 (1995).
- 671 12. B. S. Weir, C. C. Cockerham, Estimating F-Statistics for the Analysis of Population Structure.  
672 *Evolution*. **38**, 1358–1370 (1984).
- 673 13. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA  
674 polymorphism. *Genetics*. **123**, 585–595 (1989).
- 675 14. P. Pavlidis, D. Živkovic, A. Stamatakis, N. Alachiotis, SweeD: likelihood-based detection of  
676 selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
- 677 15. X. Zhou, P. Carbonetto, M. Stephens, Polygenic modeling with bayesian sparse linear mixed  
678 models. *PLoS Genet.* **9**, e1003264 (2013).
- 679 16. T. Mitchell-Olds, Pleiotropy Causes Long-Term Genetic Constraints on Life-History Evolution  
680 in *Brassica rapa*. *Evolution*. **50**, 1849–1858 (1996).
- 681 17. E. Porcher, T. Giraud, I. Goldringer, C. Lavigne, Experimental demonstration of a causal  
682 relationship between heterogeneity of selection and genetic differentiation in quantitative  
683 traits. *Evolution*. **58**, 1434–1445 (2004).
- 684 18. C.-R. Lee, T. Mitchell-Olds, Complex trait divergence contributes to environmental niche  
685 differentiation in ecological speciation of *Boechera stricta*. *Mol. Ecol.* **22**, 2204–2217 (2013).
- 686 19. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).
- 687 20. J. J. Berg, G. Coop, A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet.* **10**,  
688 e1004412–e1004412 (2014).
- 689 21. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense  
690 haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
- 691 22. B. L. Browning, S. R. Browning, Genotype Imputation with Millions of Reference Samples. *Am.*  
692 *J. Hum. Genet.* **98**, 116–126 (2016).
- 693 23. K. Choi *et al.*, *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at  
694 gene promoters. *Nat. Genet.* **45**, 1327–1336 (2013).
- 695 24. E. B. Josephs, J. R. Stinchcombe, S. I. Wright, What can genome-wide association studies tell  
696 us about the evolutionary forces maintaining genetic variation for quantitative traits? *New*  
697 *Phytol.* (2017), doi:10.1111/nph.14410.
- 698 25. F. C. Jones *et al.*, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*.  
699 **484**, 55–61 (2012).
- 700 26. D. Shriener, A. Adeyemo, E. Ramos, G. Chen, C. N. Rotimi, Mapping of disease-associated  
701 variants in admixed populations. *Genome Biol.* **12**, 223 (2011).
- 702 27. M. Tesfaye *et al.*, Spatio-Temporal Expression Patterns of *Arabidopsis thaliana* and *Medicago*  
703 *truncatula* Defensin-Like Genes. *PLoS One.* **8**, e58992 (2013).

- 704 28. M. Dannemann, A. M. Andrés, J. Kelso, Introgression of Neandertal- and Denisovan-like  
705 Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.*  
706 **98**, 22–33 (2016).
- 707 29. D. C. Rife, Populations of hybrid origin as source material for the detection of linkage. *Am. J.*  
708 *Hum. Genet.* **6**, 26–33 (1954).
- 709 30. H. Qin *et al.*, Interrogating local population structure for fine mapping in genome-wide  
710 association studies. *Bioinformatics.* **26**, 2961–2968 (2010).
- 711 31. J.-Y. Li *et al.*, The Arabidopsis nitrate transporter NRT1.8 functions in nitrate removal from the  
712 xylem sap and mediates cadmium tolerance. *Plant Cell.* **22**, 1633–1646 (2010).
- 713 32. A. Gojon, F. Gaymard, Keeping Nitrate in the Roots: An Unexpected Requirement for  
714 Cadmium Tolerance in Plants. *J. Mol. Cell Biol.* **2**, 299–301 (2010).
- 715 33. K. M. Jarzyniak, M. Jasiński, Membrane transporters and drought resistance - a complex issue.  
716 *Front. Plant Sci.* **5**, 687 (2014).
- 717 34. C. Lelandais-Brière *et al.*, Disruption of AtOCT1, an organic cation transporter gene, affects  
718 root development and carnitine-related responses in Arabidopsis. *Plant J.* **51**, 154–164 (2007).
- 719 35. S. Ma, Q. Gong, H. J. Bohnert, Dissecting salt stress pathways. *J. Exp. Bot.* **57**, 1097–1107  
720 (2006).
- 721 36. M. P. Apse, J. B. Sottosanto, E. Blumwald, Vacuolar cation/H<sup>+</sup> exchange, ion homeostasis, and  
722 leaf development are altered in a T-DNA insertional mutant of AtNHX1, the Arabidopsis  
723 vacuolar Na<sup>+</sup>/H<sup>+</sup> antiporter. *Plant J.* **36**, 229–239 (2003).
- 724 37. J. T. Lovell *et al.*, Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation.  
725 *Proceedings of the Royal Society of London B: Biological Sciences.* **280**, 20131043 (2013).
- 726 38. J. A. Bac-Molenaar, C. Granier, J. J. B. Keurentjes, D. Vreugdenhil, Genome-wide association  
727 mapping of time-dependent growth responses to moderate drought stress in Arabidopsis.  
728 *Plant Cell Environ.* **39**, 88–102 (2016).
- 729 39. F. Vasseur, T. Bontpart, M. Dauzat, C. Granier, D. Vile, Multivariate genetic analysis of plant  
730 responses to water deficit and high temperature revealed contrasting adaptive strategies. *J.*  
731 *Exp. Bot.* **65**, 6457–6469 (2014).
- 732 40. R. H. Moss *et al.*, The next generation of scenarios for climate change research and  
733 assessment. *Nature.* **463**, 747–756 (2010).
- 734 41. J. Guiot, W. Cramer, Climate change: The 2015 Paris Agreement thresholds and  
735 Mediterranean basin ecosystems. *Science.* **354**, 465–468 (2016).
- 736 42. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated  
737 climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- 738 43. M. Exposito-Alonso *et al.*, The rate and effect of de novo mutations in a colonising lineage of  
739 Arabidopsis thaliana. *bioRxiv* (2016), p. 050203.
- 740 44. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
- 741 45. W. Thuiller, S. Lavorel, M. B. Araújo, M. T. Sykes, I. C. Prentice, Climate change threats to plant



- 742 diversity in Europe. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8245–8250 (2005).
- 743 46. M. Falahati-Anbaran, S. Lundemo, H. K. Stenøien, Seed dispersal in time can counteract the  
744 effect of gene flow between natural populations of *Arabidopsis thaliana*. *New Phytol.* **202**,  
745 1043–1054 (2014).
- 746 47. S. Ossowski *et al.*, The rate and molecular spectrum of spontaneous mutations in *Arabidopsis*  
747 *thaliana*. *Science.* **327**, 92–94 (2010).
- 748 48. A. Platt *et al.*, The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* **6**,  
749 e1000843 (2010).
- 750 49. M. Nordborg, P. Donnelly, The Coalescent Process with selfing. *Genetics.* **146**, 1185–1195 (1997).
- 751 50. F. Roux, S. Giancola, S. Durand, X. Reboud, Building of an experimental cline with *Arabidopsis*  
752 *thaliana* to estimate herbicide fitness cost. *Genetics.* **173**, 1023–1031 (2006).
- 753 51. B. C. Haller, P. W. Messer, SLiM 2: Flexible, Interactive Forward Genetic Simulations. *Mol. Biol.*  
754 *Evol.* **34**, 230–240 (2017).
- 755 52. P. A. Salomé *et al.*, The recombination landscape in *Arabidopsis thaliana* F2 populations.  
756 *Heredity* . **108**, 447–455 (2012).
- 757 53. K. Bomblies *et al.*, Local-scale patterns of genetic variability, outcrossing, and spatial structure  
758 in natural stands of *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1000890–e1000890 (2010).
- 759 54. A. Fournier-Level *et al.*, Predicting the evolutionary dynamics of seasonal adaptation to novel  
760 climates in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2812–21 (2016).
- 761 55. L.-M. Chevin, M. E. Visser, J. Tufto, Estimating the variation, autocorrelation, and  
762 environmental sensitivity of phenotypic selection. *Evolution.* **69**, 2319–2332 (2015).
- 763 56. P. de Villemereuil, O. E. Gaggiotti, M. Mouterde, I. Till-Bottraud, Common garden experiments  
764 in the genomic era: new perspectives and opportunities. *Heredity.* **116**, 249–254 (2016).
- 765 57. R. I. Colautti, J. A. Lau, Contemporary evolution during invasion: evidence for differentiation,  
766 natural selection, and local adaptation. *Mol. Ecol.* **24**, 1999–2017 (2015).
- 767 58. K. M. Dlugosch, S. R. Anderson, J. Braasch, F. A. Cang, H. D. Gillette, The devil is in the details:  
768 genetic variation in introduced populations and its contributions to invasion. *Mol. Ecol.* **24**,  
769 2095–2111 (2015).
- 770 59. T. L. Karasov *et al.*, The long-term maintenance of a resistance polymorphism through diffuse  
771 interactions. *Nature.* **512**, 436–440 (2014).
- 772 60. K. Jakob *et al.*, *Pseudomonas viridiflava* and *P. syringae*—Natural Pathogens of *Arabidopsis*  
773 *thaliana*. *Mol. Plant. Microbe. Interact.* **15**, 1195–1203 (2002).
- 774 61. M. B. Araújo, M. Luoto, The importance of biotic interactions for modelling species  
775 distributions under climate change. *Glob. Ecol. Biogeogr.* **16**, 743–753 (2007).
- 776 62. L. M. Chevin, R. Lande, When do adaptive plasticity and genetic evolution prevent extinction  
777 of a density-regulated population? *Evolution.* **64**, 1143–1150 (2010).
- 778 63. S. Lundemo, M. Falahati-Anbaran, H. K. Stenøien, Seed banks cause elevated generation

- 779 times and effective population sizes of *Arabidopsis thaliana* in northern Europe. *Mol. Ecol.* **18**,  
780 2798–2811 (2009).
- 781 64. B. Charlesworth, Selection in populations with overlapping generations. VI. Rates of change  
782 of gene frequency and population growth rate. *Theor. Popul. Biol.* **6**, 108–133 (1974).
- 783 65. N. Li, M. Stephens, Modeling linkage disequilibrium and identifying recombination hotspots  
784 using single-nucleotide polymorphism data. *Genetics.* **165**, 2213–2233 (2003).
- 785 66. J. J. Berg, G. Coop, A Coalescent Model for a Sweep of a Unique Standing Variant. *Genetics.*  
786 **201**, 707–725 (2015).
- 787
- 788
- 789
- 790
- 791
- 792

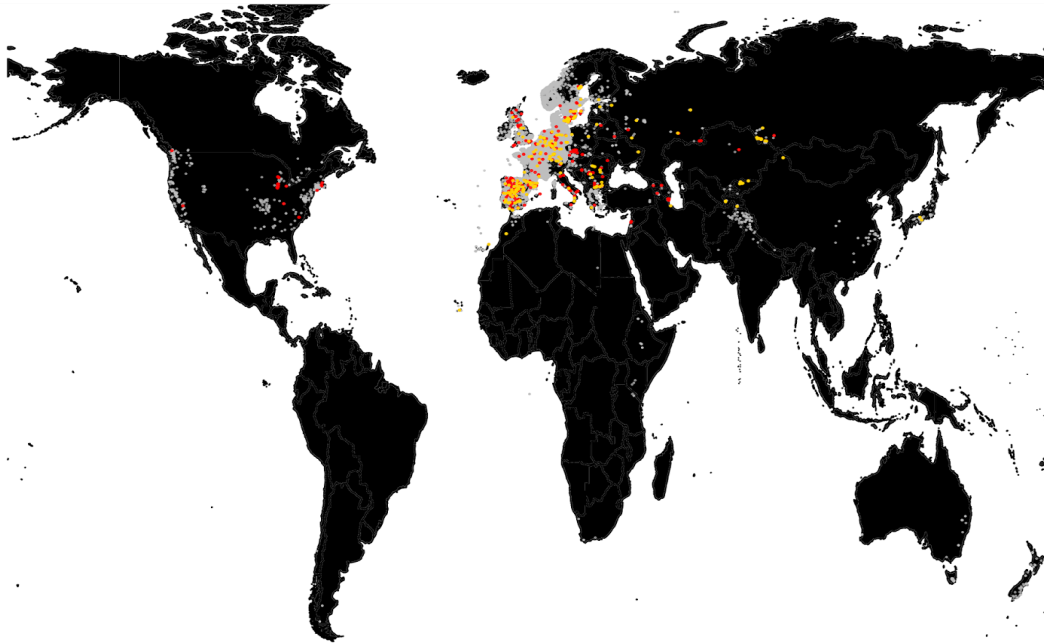
793

794

## SUPPLEMENTARY FIGURES & MEDIA

795

796



797

### Figure S1. Extent of *Arabidopsis thaliana* distribution

798

The global map shows ca. 80,000 records from the Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org)) (grey), the 762 global accessions used for genetic analyses (red), and 210 accessions used for phenotyping experiments (yellow).

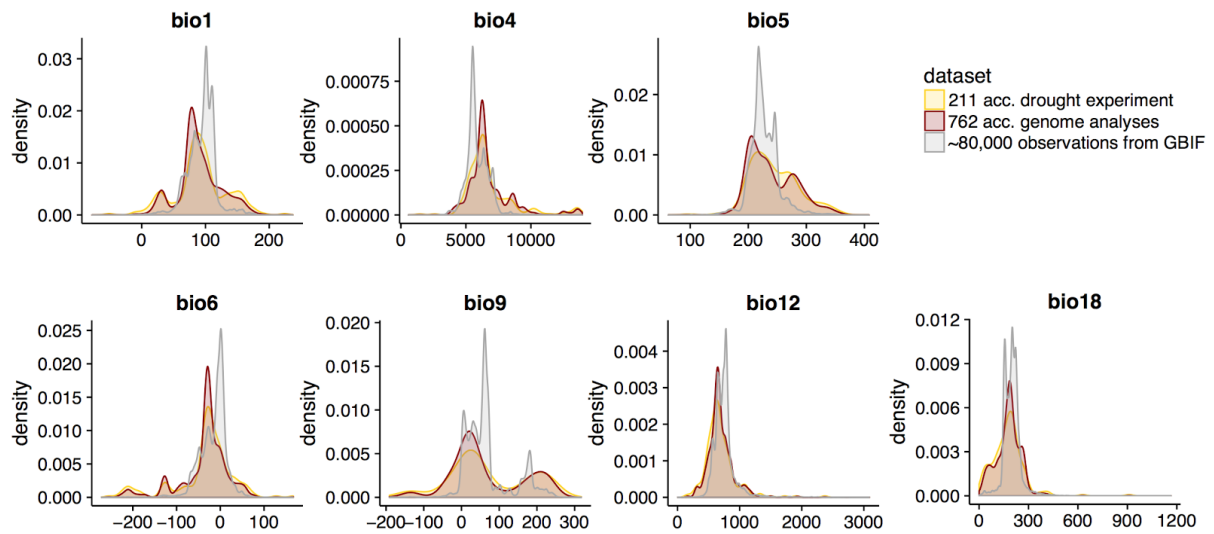
800

801

802

803

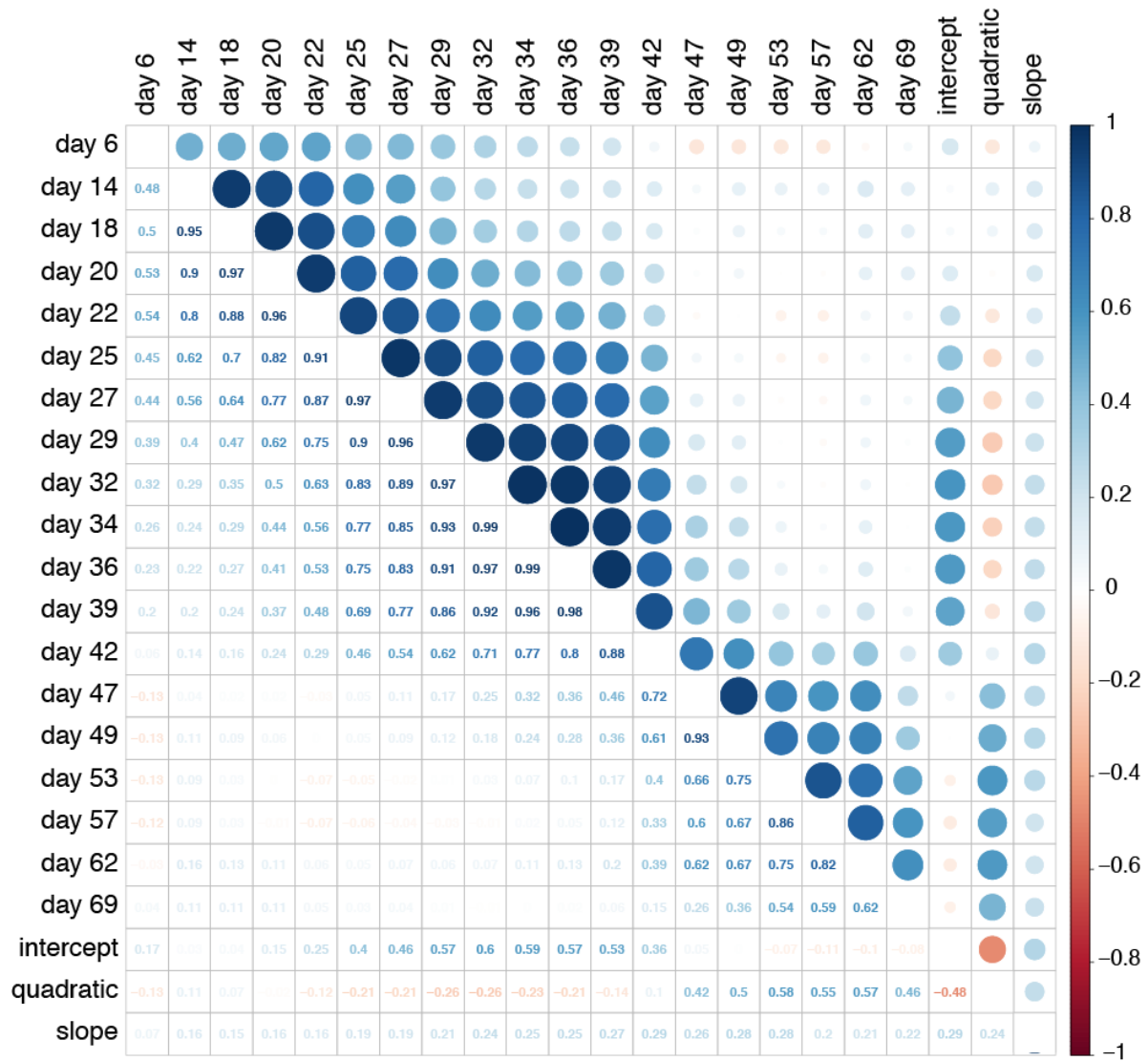
804  
805



806 **Figure S2. Environmental ranges of *Arabidopsis thaliana***

807 We show the range in key environmental variables for the three datasets in Fig. S1. The set  
808 of accessions used in our analyses not only covered the range of the species as estimated from GBIF  
809 data, but also showed that these accessions have a more even distribution throughout the  
810 environmental ranges. The bioclimatic variables are: annual precipitation (bio12), precipitation of  
811 the warmest quarter (bio18), annual mean temperature (bio1), temperature seasonality (bio4),  
812 maximum temperature of the warmest month (bio5), minimum temperature of the coldest  
813 month (bio6), and mean temperature of the driest quarter (bio9). See Table S5 for more  
814 information.  
815

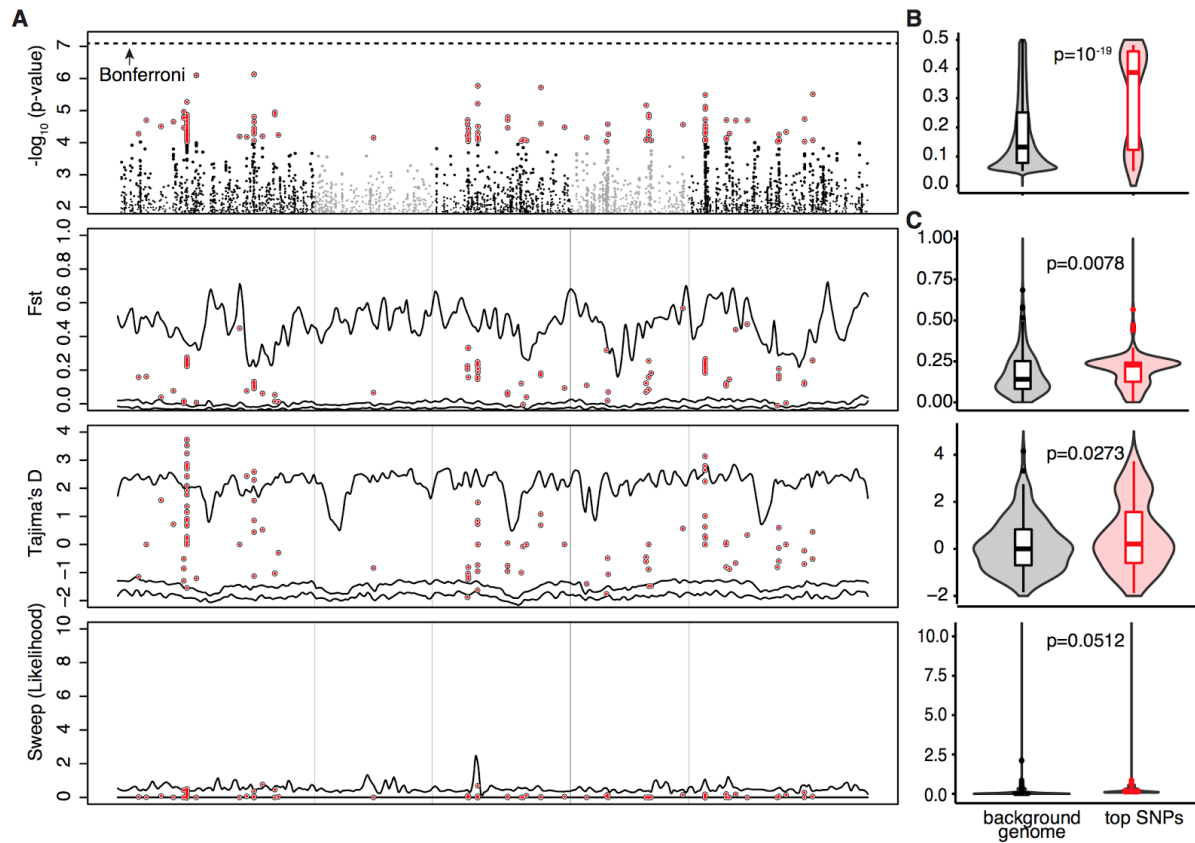
816



817 **Figure S3. Correlation between raw green pixels in plant images and model parameters**  
 818 Pearson product-moment correlation coefficients between the three drought-index parameters  
 819 and the 'raw' number of green pixels per pot. Sizes of circles indicate strength and colors signs of  
 820 association, shown as numbers in the lower triangle.

821

822



823

**Figure S4. GWA with drought survival and population genetic statistics**

824

(A) Manhattan plot of drought resistant GWA,  $F_{st}$ , Tajima's D, and selective sweeps. (B) Violin and box plots of allele frequency, and (C)  $F_{st}$ , Tajima's D, and selective sweeps of the top 150 SNPs (red) vs frequency-matched 150 SNPs from a random genome background (grey). GWA was calculated using EMMAX.  $F_{st}$  across populations (see Fig. 1) and Tajima's D were calculated using PLINK. Sweep likelihood was calculated using SWEED software. Median p-values from Wilcoxon tests with 100 bootstrap replicates are indicated.

827

828

829

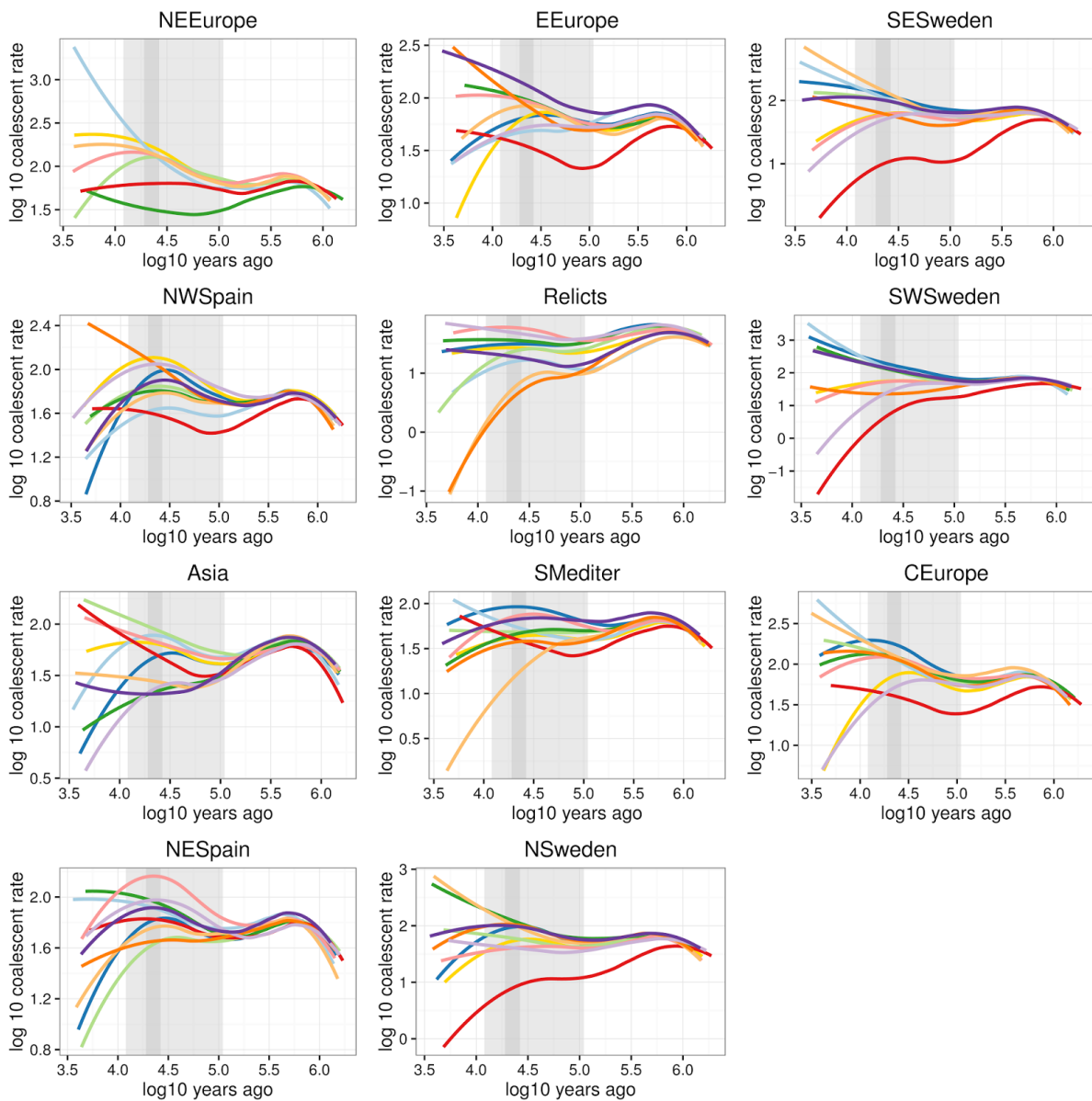
830

831



832

833



834

**Figure S5. Cross-coalescent rates between populations inferred by MSMC**

835

Joint coalescent rates of each of the 11 ADMIXTURE genetic groups are (see Fig. 1 and Fig. S12)

836

compared to the other groups. Each line is a smoothed loess of 6 replicated runs. Light grey area

837

indicates the extent of the last glacial maxima (100-10 kya) and dark grey area the peak of the last

838

glaciation (22 kya). Analyses between certain groups failed (e.g. NE Europe with Asia), likely due to

839

proximity between genomes. Note that the N. Swedish group is the first to separate from the

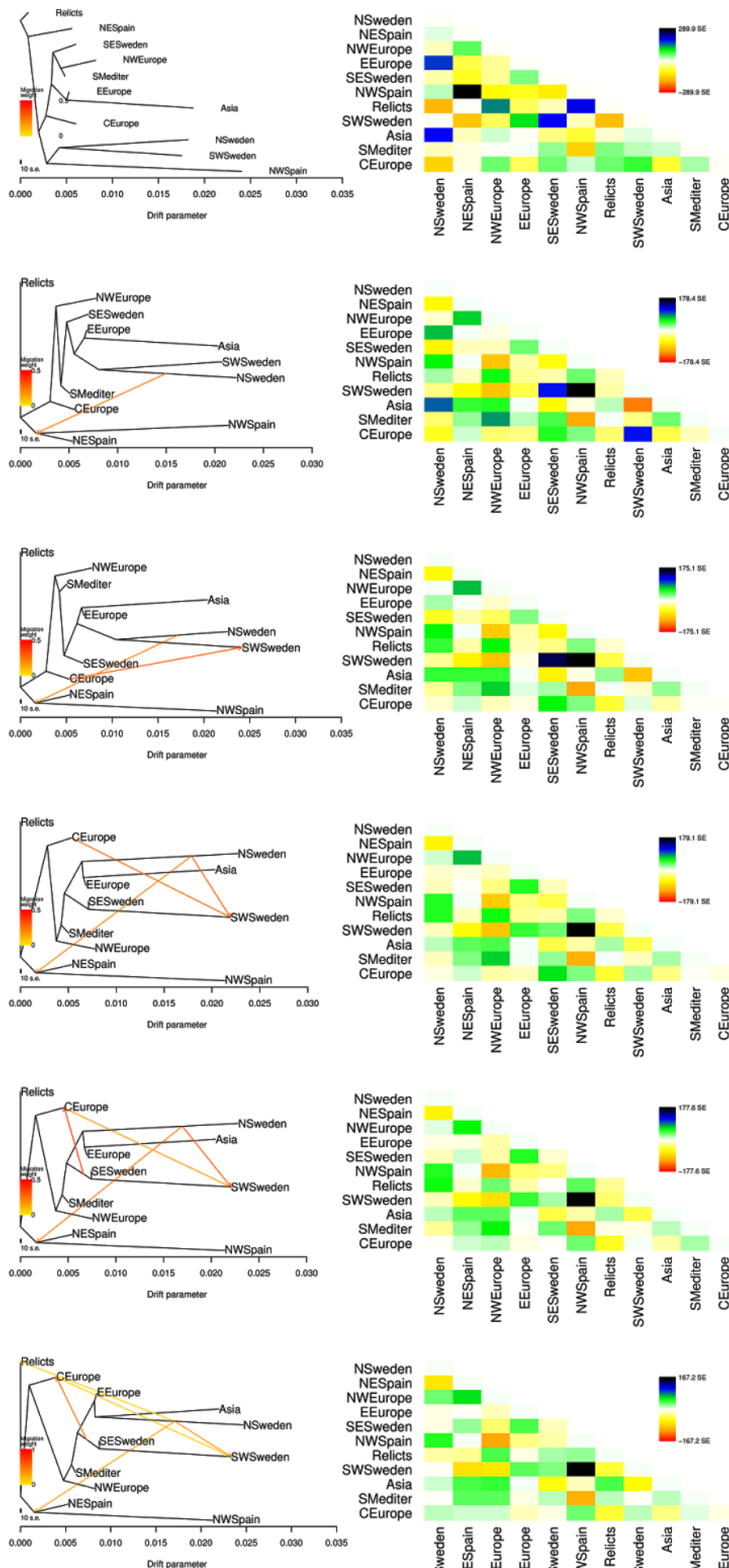
840

relicts.

841

842

843



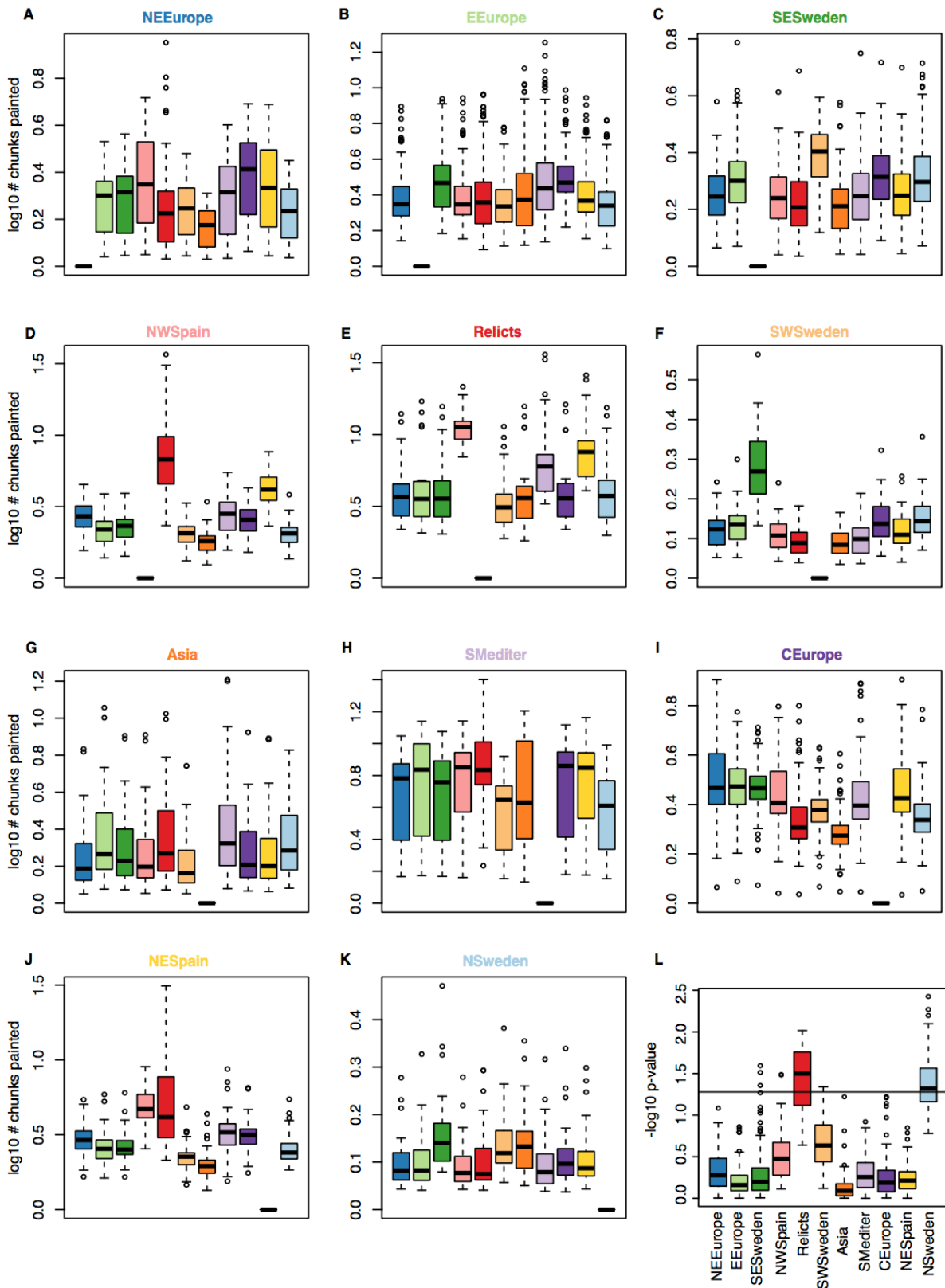
844 **Figure S6. Treemix with different migration rates**

845 Maximum likelihood (ML) population trees from Treemix (left). Analyses with zero to five  
846 migration edges are presented. Heatmaps with the residual fit of the ML trees are shown on the  
847 right. Note that the unexpected closeness of NW Spain and Sweden without migration is resolved  
848 with one migration edge. With this, a more parsimonious tree that adheres to geographic locations  
849 is uncovered.

850

851

852  
853



854 **Figure S7. Genomic ChromoPainter chunks per population**  
855 (A-K) Summary of the number of ChromoPainter chunks inherited from other genomes that had

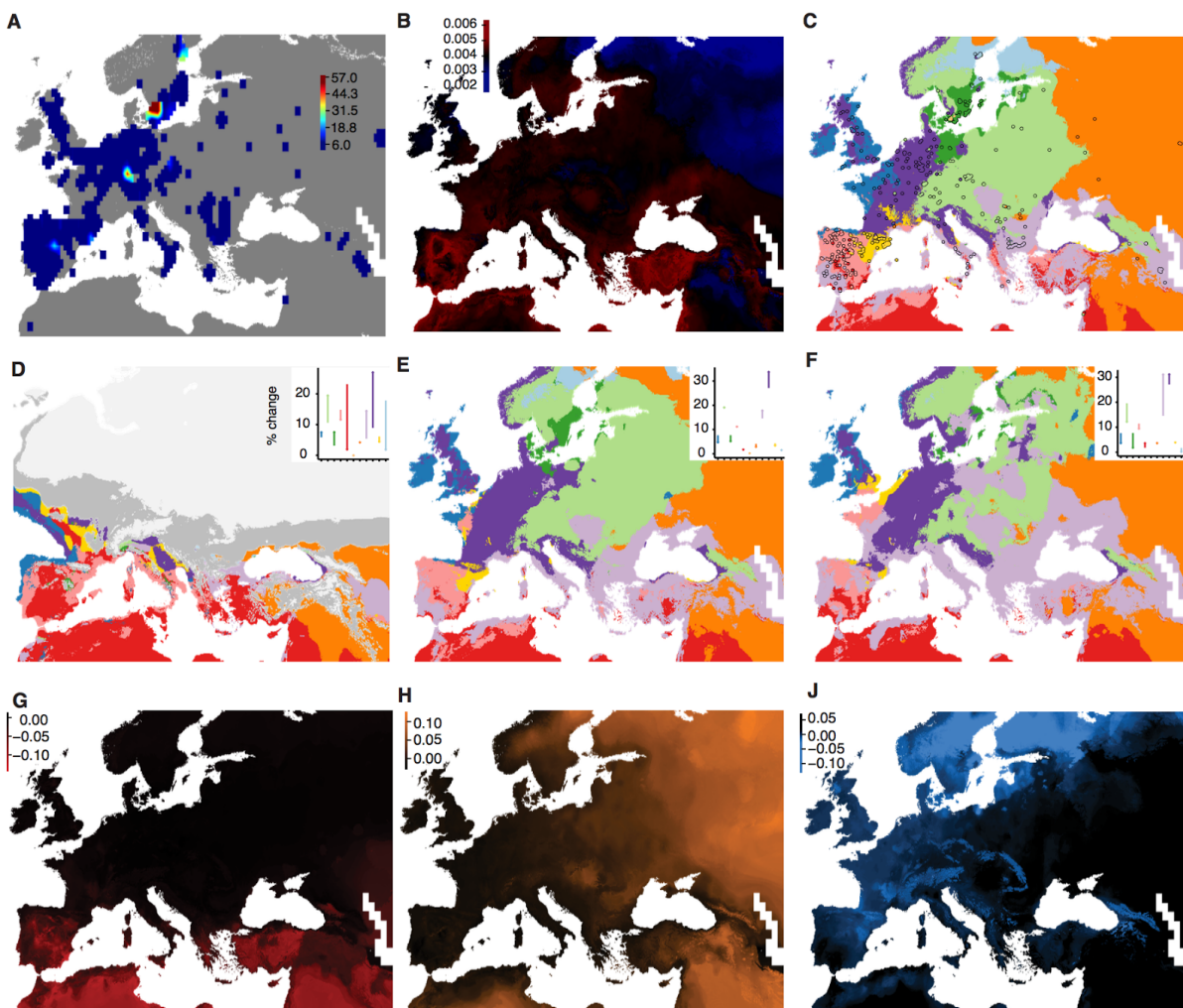
856 been assigned to ADMIXTURE groups. Each graph summarises the information of all the genomes  
857 from an admixture group. (L) The p-value of the Pearson correlation test between an accession's  
858 drought survival index and the number of chunks received from another genome. The p-value  
859 distributions of genomes from the same ADMIXTURE group are grouped in a box plot. Intuitively  
860 this can be interpreted as how well the number of chunks inherited from a specific donor predicts  
861 the drought survival of the receiver. The black line indicates the 5% significance threshold, which is  
862 passed by most relict and N. Sweden groups. Therefore, chunks that have N. Swedish and/or relict  
863 ancestry explain the drought survival of other individuals well.

864

865

866

867



868 **Figure S8. Environmental niche model of genetic diversity and population structure**

869 (A) Distribution of 762 accessions from the 1001 Genomes project used for environmental niche  
870 modeling of genetic diversity and analysis of population structure. Colors indicate number of  
871 accessions within a 1°x1° latitude x longitude grid. (B) Random forest environment niche models  
872 using estimates of pairwise nucleotide diversity ( $\pi$ ) diversity of each accession with its closest  
873 10 geographic neighbours. The trained model was used to predict diversity based on  
874 environmental data. (C) Random forest environment model of the 11 genetic groups (see Fig. 1).  
875 Locations with accessions are shown as points filled with the actual genetic group assigned, and  
876 are used for model training as in (B). The trained model was used to predict a raster of  
877 environmental variables and is shown in the background. When the circle is filled with the same  
878 color as the background, the model succeeds in the prediction. The trained model was also used  
879 to predict presence of different genetic groups at the Last Glacial Maxima (D) and for 2070 under  
880 low (E) and high (F) CO<sub>2</sub> concentration scenarios. (see Fig. 1 for color keys). (G-J) The first three  
881 genome-wide principal components from Fig. 2 were modeled based on environmental variables.

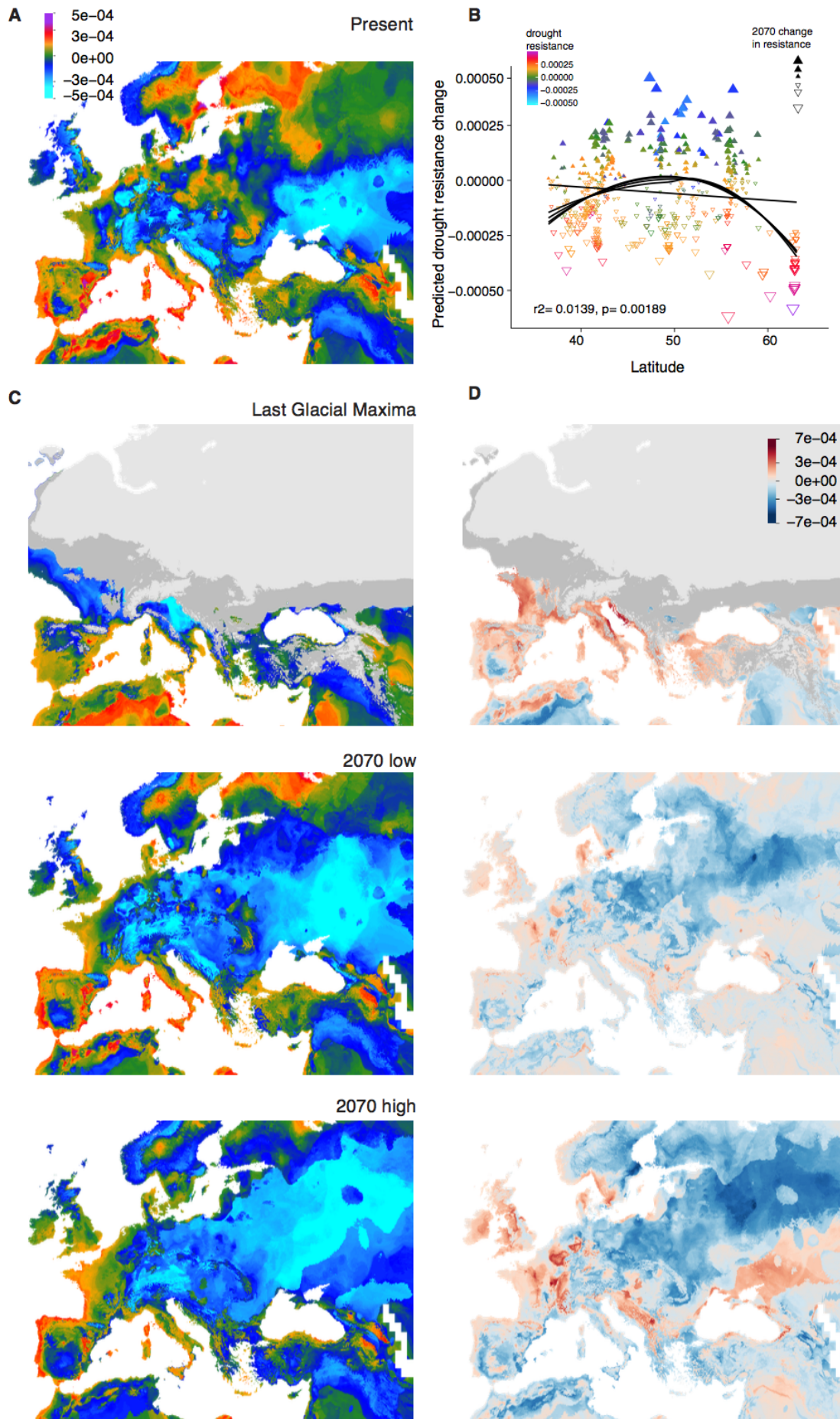


882 Later, these were used as covariates of Genome-Environment Models (Fig. S13-14).

883

884

885



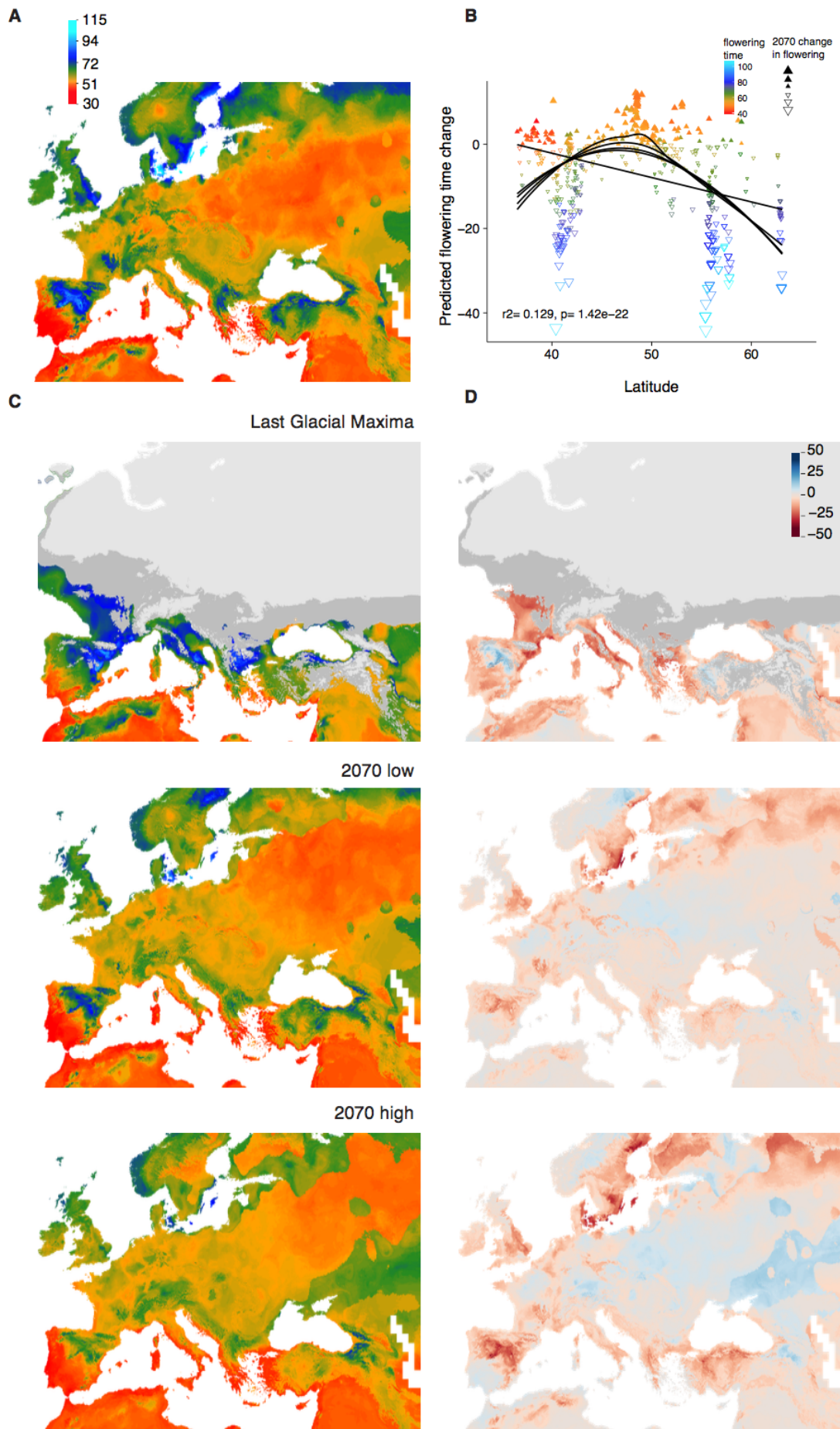
886 **Figure S9. Environmental niche model of drought index**

887 (A) Present geographic prediction of drought survival index from a random forest environment  
888 niche model trained on experimentally determined phenotypes for 211 accessions. Note that the  
889 highest drought survival values are inferred for the Mediterranean as well as N. Sweden. (B)  
890 Correlation of phenotypic change in 2070 under a high CO<sub>2</sub> scenario with latitude; colors indicate  
891 present drought survival values. (C) The trained model is also used to predict drought survival  
892 index under the Last Glacial Maximum, and for two 2070 scenarios of low and high CO<sub>2</sub>  
893 concentrations. (D) For the three scenarios, the change is shown relative to the current date  
894 prediction for easier comparison.

895

896

897



898

899 **Figure S10. Environmental niche model of flowering time**

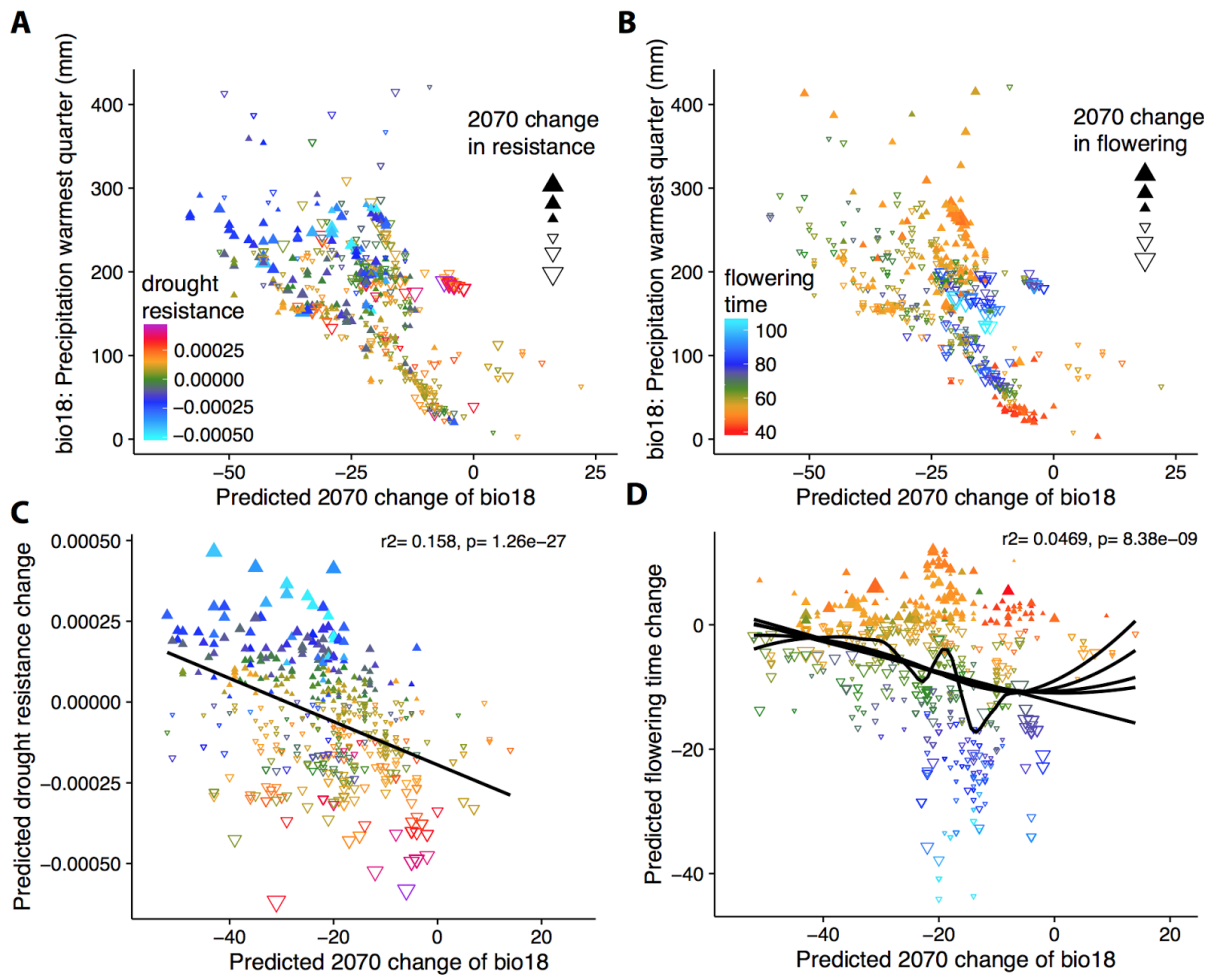
900 Same models as in Fig. S9, but for flowering time.

901



902

903



904

### Figure S11. Profile of phenotypic change under climate change

905

(A, B) Correlation of precipitation during the warmest quarter today and in 2070 under a high CO<sub>2</sub> scenario. Colors indicate current drought survival (A) or flowering time (B), and shapes indicate

906

increase or decrease in trait values for 2070. (C, D) Regression of the predicted change in drought-survival (C) and flowering time (D) on the predicted change in precipitation in 2070.

907

Note that areas with already low precipitation will not have large decreases in precipitation in 2070 (A-B). Note also the linear relationship between decreased precipitation in 2070 and

908

predicted increase of drought-survival in (C). Flowering will be on average faster in 2070 (D), but the relationship between precipitation reduction and flowering time change is not linear, which

909

suggests that areas with a moderate reduction in precipitation will have accelerated flowering (rather than increased drought survival).

910

911

912

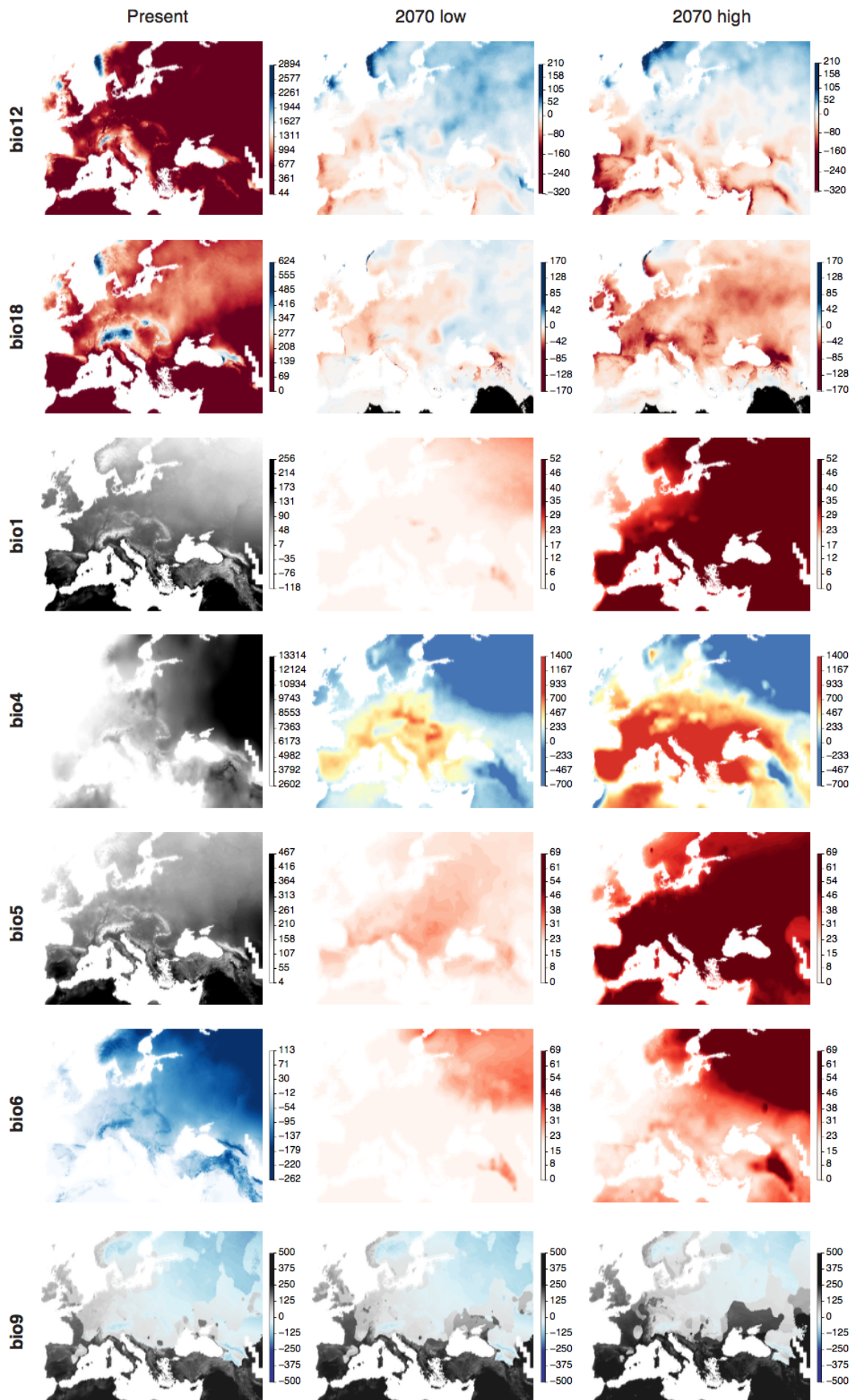
913

914

915



916



917 **Figure S12. Maps of the most important climatic variables**

918 The bioclimatic variables ([www.worldclim.org](http://www.worldclim.org)) that typically had more importance in phenotypic  
919 and genome environmental models are shown as an aid for interpretation of the results from our  
920 study. bioclim variable shown are annual precipitation (bio12), precipitation of the warmest quarter  
921 (bio18), annual mean temperature (bio1), temperature seasonality (bio4), maximum temperature  
922 of the warmest month (bio5), minimum temperature of the coldest month (bio6), and mean  
923 temperature of the driest quarter (bio9). The columns show distributions at present, in 2070  
924 under a scenario of low CO<sub>2</sub> concentration, and in 2070 under a scenario of high CO<sub>2</sub> scenario.  
925 Except for bio9, the values for future scenarios were expressed as future-present difference to  
926 highlight geographic areas that will change the most. Note the bimodality of bio9: areas in black  
927 are summer drought (Mediterranean climate) areas, whereas blue areas are winter-drought. Also  
928 note that bio18 is predicted to change mostly along the transition from the Mediterranean to  
929 non-Mediterranean climate. In bio18, areas that will be under lower precipitation than any current  
930 location of *A. thaliana* are shown in black, to highlight that most areas will remain within the range  
931 of current precipitation across the species range.

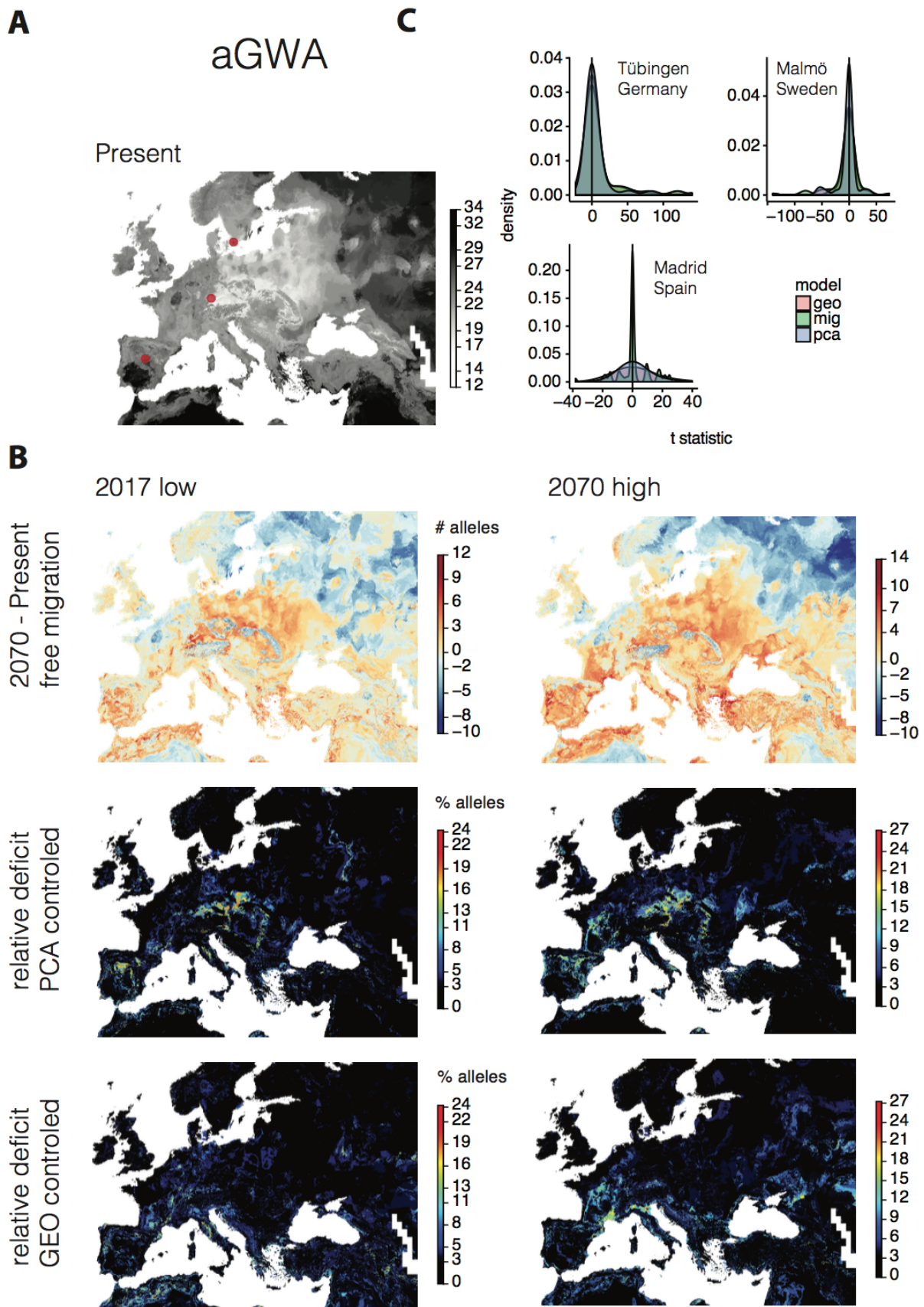
932

933

934

935

936



937 **Figure S13. aGWA Genome Environment Models (GEM)**

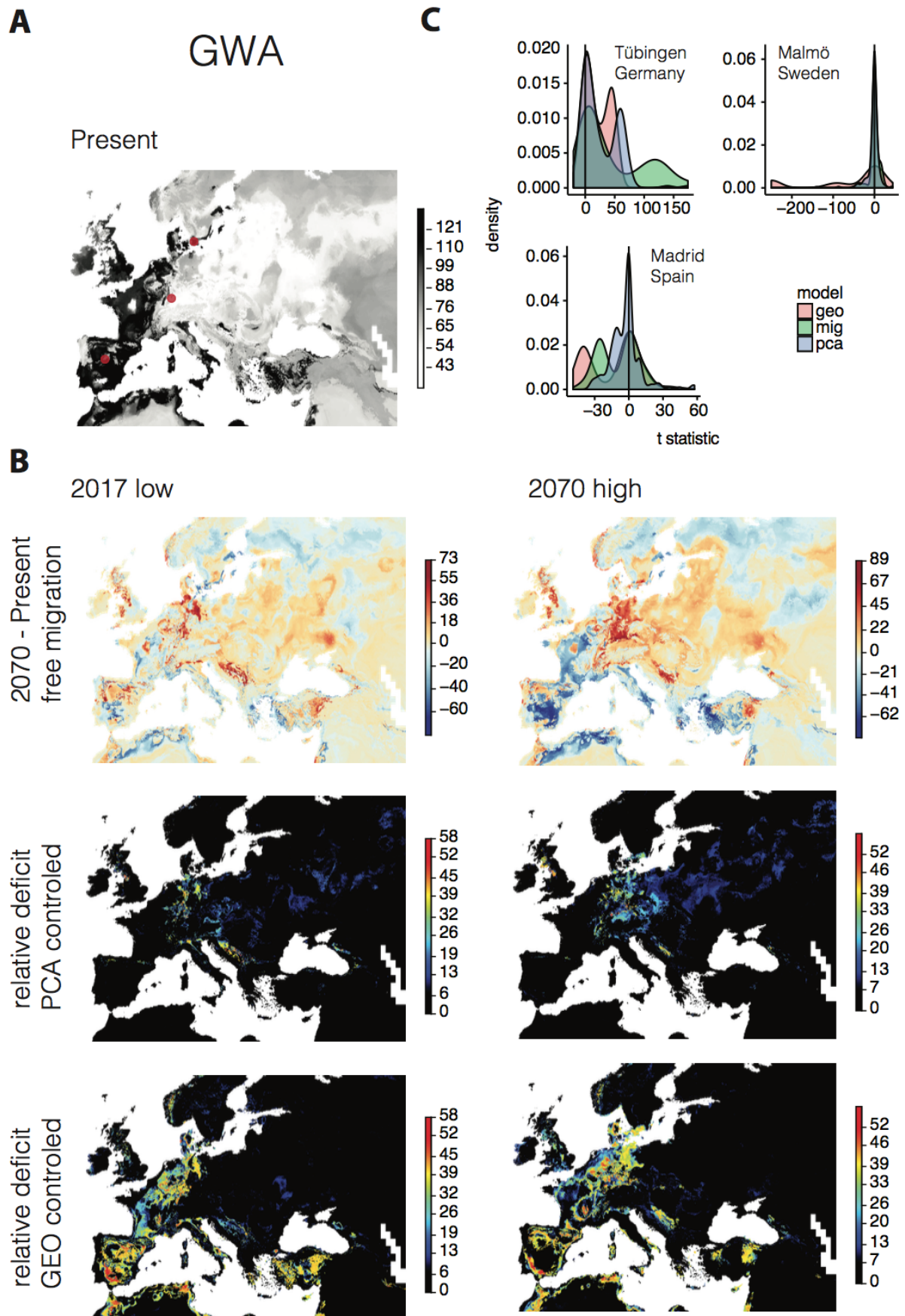
938 (A) We ran GEMs to describe the geographic distribution of alleles at the 70 aGWA top loci.  
939 Concatenating all maps, we produced a map of the count of all drought-survival alleles that a  
940 genotype is expected to have in a given location today. (B) The trained model from (A) was used  
941 to predict distribution of drought alleles in the future. The difference to numbers inferred for today  
942 (A) corresponds to the alleles that will have been gained or lost in 2070 in a given location. Two  
943 additional models were trained which included a genome background (PCA) correction and  
944 latitudinal and longitudinal (GEO) correction of the allele distributions. The percentage of gained  
945 alleles from the “free” model that were not present in the corrected models is shown as a deficit in  
946 percentage. (C) For three highly sampled locations, Madrid (Spain), Tübingen (Germany) and  
947 Malmö (Sweden), we calculated allele frequency differences between today and 2070 (under  
948 high CO<sub>2</sub>) and calculated a t-statistic to describe the effect size of the change. A skew towards the  
949 right (increase) is observed for Tübingen only.

950



951

952



953

954 **Figure S14. GWA Genome Environment Models (GEM)**

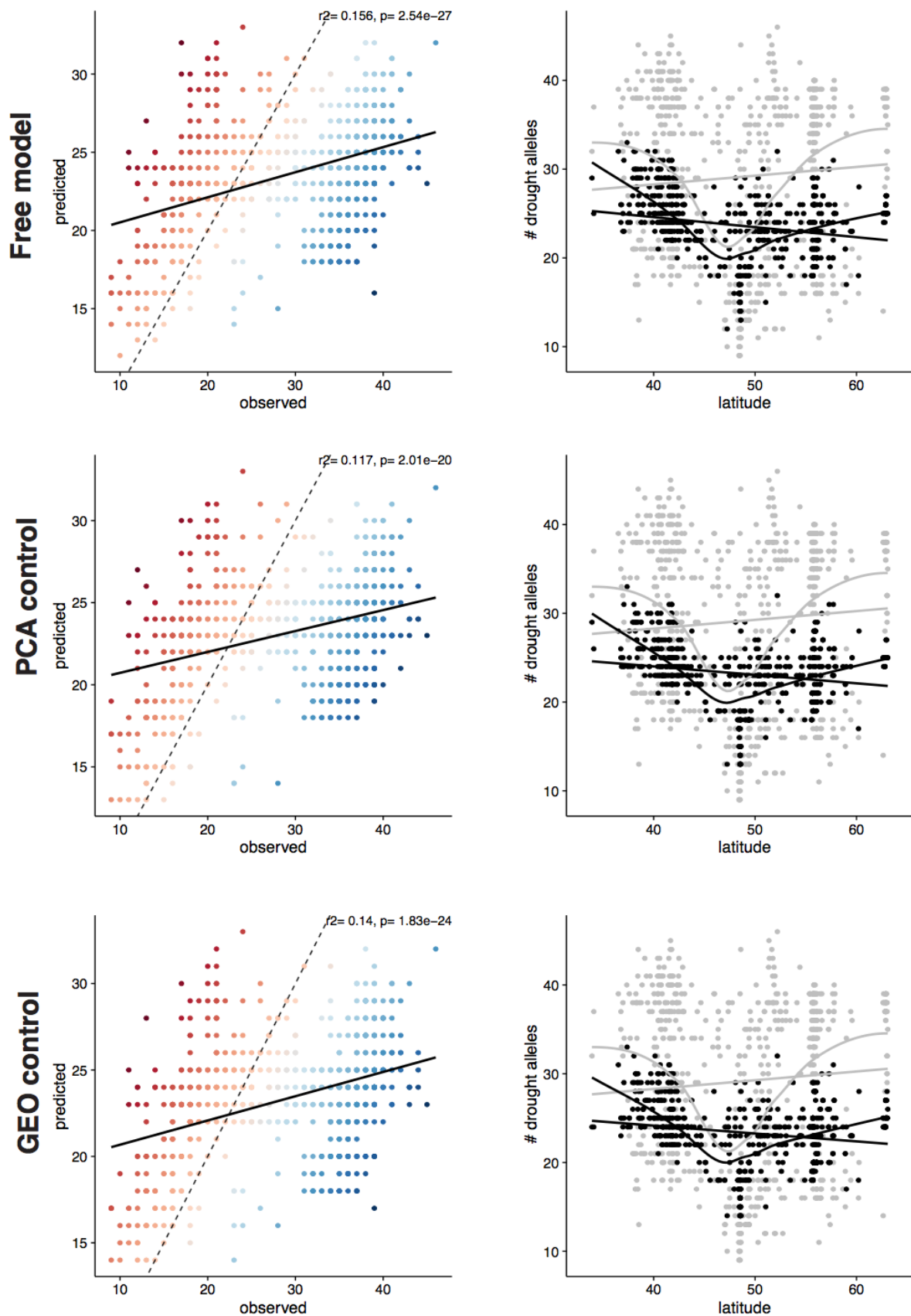
955 See Fig. S13 for legend.

956



957

### aGWA



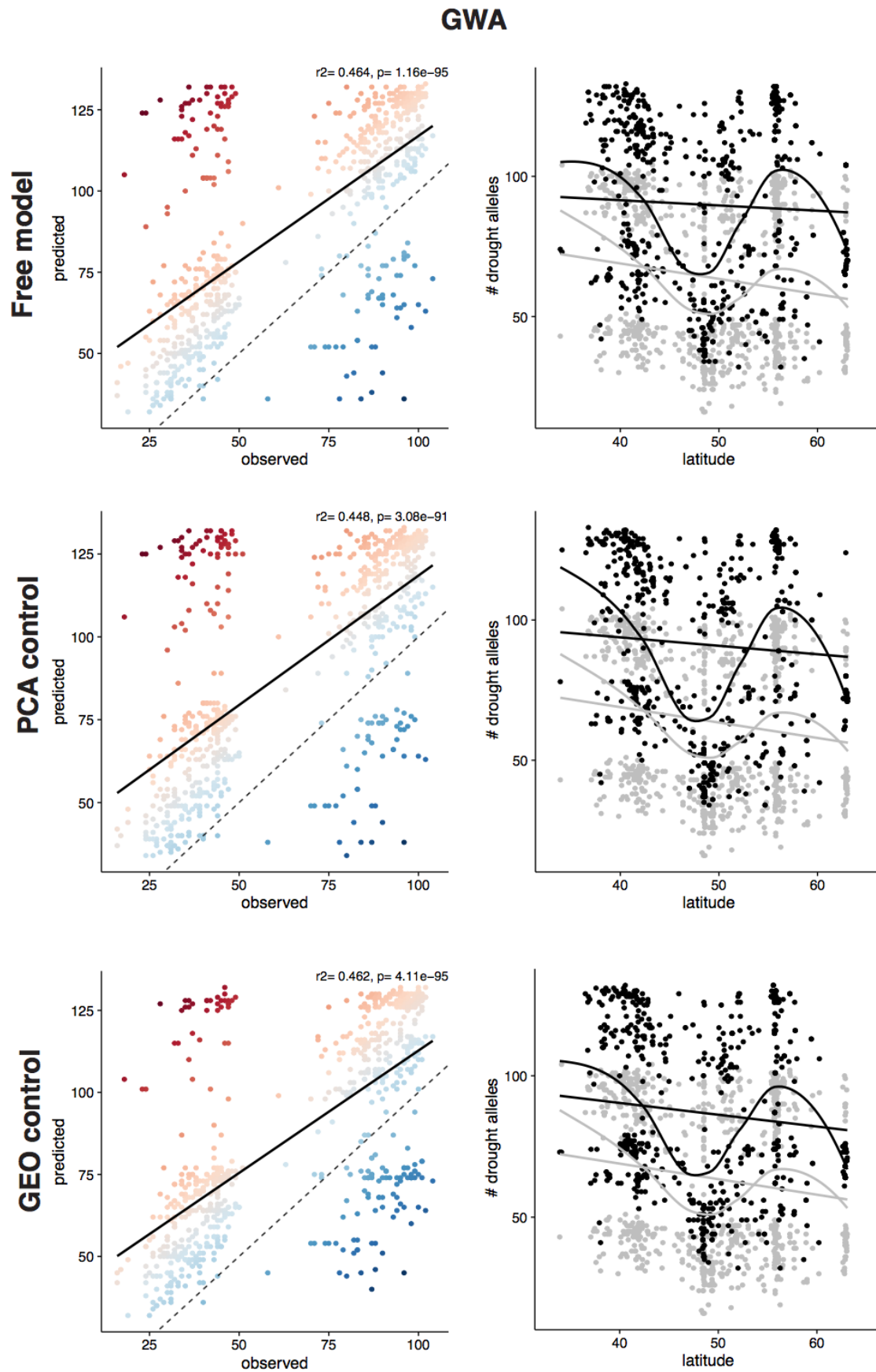
958 **Figure S15. aGWA GEM residuals**

959 For each GEM, we plot the predicted against the observed (empirical) number of  
960 drought-associated alleles at each sampled locations. Red color indicates overestimation and blue  
961 underestimation. Latitudinal trends of predicted (grey) and observed (black) are shown (right).  
962 Note that the variance of predictions is larger than the empirical observations, probably due to the  
963 discrete nature of random forests.

964

965

966



967 **Figure S16. GWA GEM residuals**

968 See Fig. S15 for legend.

969

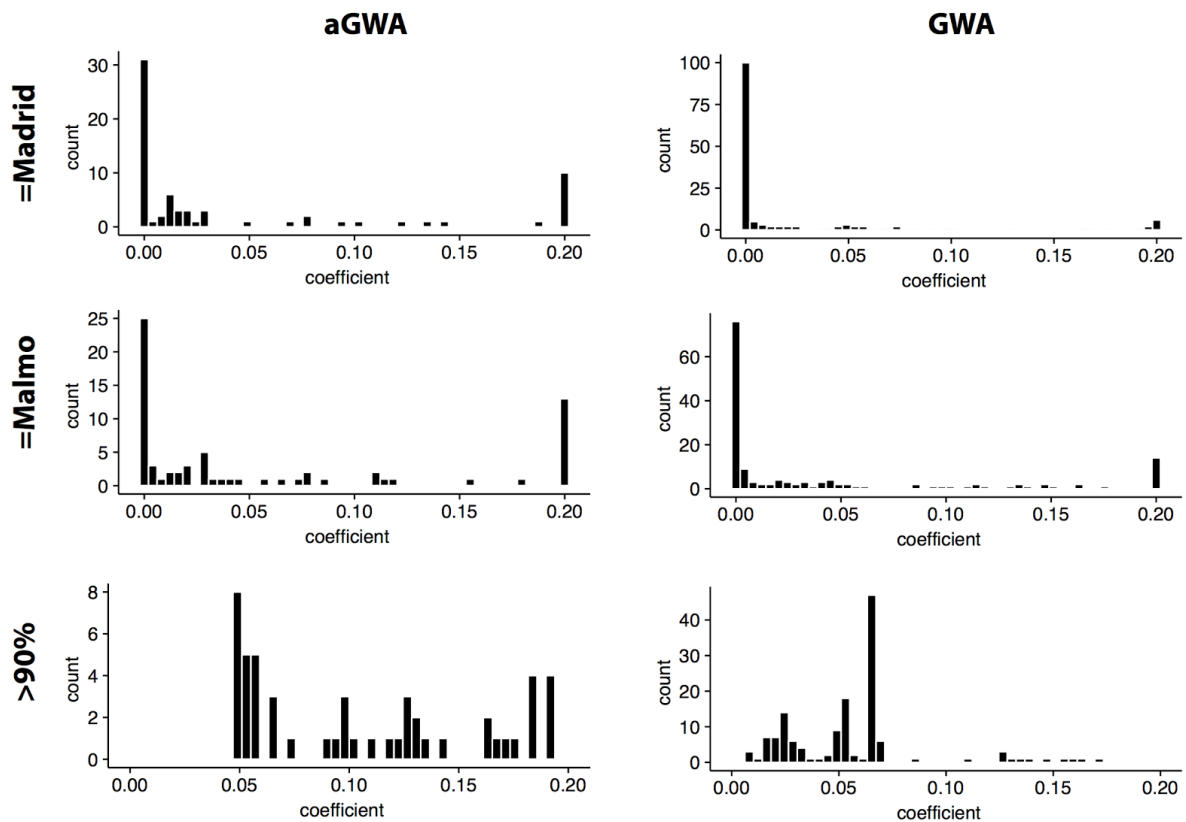
970

971

972

973

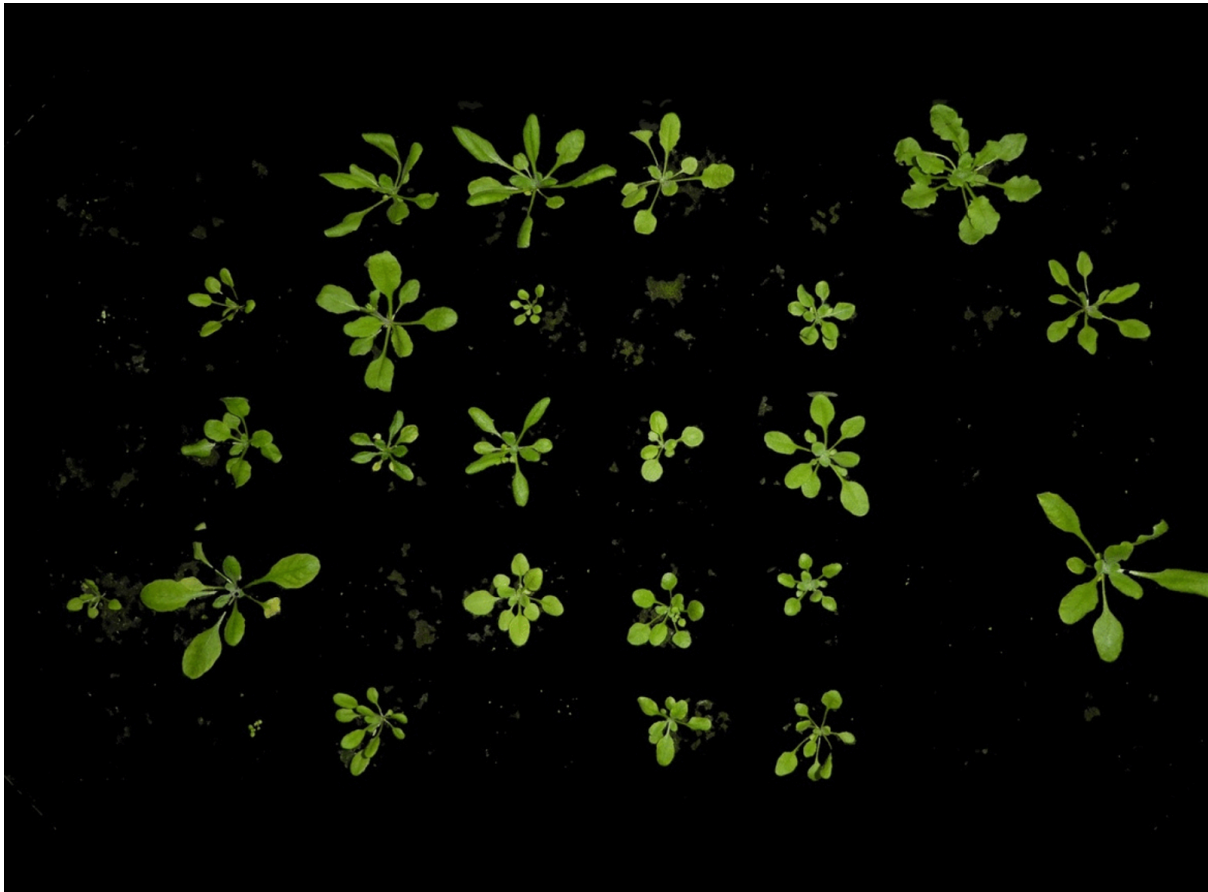
974



975 **Figure S17. Population genetics simulations**

976 We ran Wright-Fisher population simulations of 70 (aGWA) or 151 (GWA) independent loci for 50  
977 generations of evolution under mutation-selection balance, starting with the current allele  
978 frequencies in the Tübingen population., and repeating each simulation with an array of selection  
979 coefficients from 0.0001 to 0.2 (relative fitness advantage) for each locus. The distributions  
980 shown correspond to the positive selection coefficients that are required for the drought -alleles  
981 to rise to the frequency at which they are currently found in Malmö (top) or Madrid (center), or to  
982 at least 90% (bottom), which is close to fixation.

983



984 **Video S1. Example of segmentation**

985 19-frames time series of green-segmented images for one exemplary tray. See the online file  
986 Video\_S1.gif (or [click here](#)).

987

988

989



990

## 991 SUPPLEMENTARY TABLES & DATASETS

992 The combined tables can be downloaded as .xlsx file from:  
993 [http://github.com/MoisesExpositoAlonso/Exposito-Alonso\\_2017\\_drought\\_Supplementary\\_Table](http://github.com/MoisesExpositoAlonso/Exposito-Alonso_2017_drought_Supplementary_Table_s)  
994 [s](#) [here link to the journal's web upon publication]

995

### 996 **Table S1. Accession information.**

997 1001 Genomes IDs, common names, countries of origin, and geographical and environmental  
998 information.

999

### 1000 **Table S2. ADMIXTURE cross-validation for all possible groups from 2 to 20.**

1001

### 1002 **Table S3. Selection signatures and annotation of top SNP hits from GWA.**

1003

### 1004 **Table S4. Annotation of top SNP hits from aGWA.**

1005

### 1006 **Table S5. Information on phenotypic and climate traits.**

1007

### 1008 **Table S6. Correlations between climate and phenotype variables per accession.**

1009 Pearson product-moment correlation coefficients between all phenotype and climate variables of  
1010 Table S5. Lower triangle shows p-values, upper triangle correlation coefficients. The drought index  
1011 parameter of choice (m1d\_polqua) negatively correlates with the precipitation in the driest month  
1012 and quarter, bio14 and bio18, respectively.

1013

### 1014 **Table S7. Correlations between different GWA effects of the 150 polygenic SNPs.**

1015 Pearson product-moment correlation coefficients between SNP effects estimated from GWA of a  
1016 large subset of all phenotype and climate variables of Table S5.

1017

### 1018 **Table S8. Canonical Correlation Analysis.**

1019 CCA between GWA effects on different phenotypes and the SNP associations with climate  
1020 variables.

1021

### 1022 **Table S9. Polygenic model at different top SNPs groups.**

1023 We applied the Berg & Coop model (66) of polygenic adaptation to different groups of top SNPs  
1024 and report the value of  $Q_x$  statistics.

1025

1026 **Table S10. Importance of variables in Random Forest analyses.**

1027 For each random forest model, the importance of bioclimatic variables is reported. For  
1028 classification random forest, importance is reported as the mean decreased accuracy (MDA) and  
1029 for regression random forest, importance is reported as the mean square error (MSE). MDA is the  
1030 number of misclassified observations when removing a variable and MSE is the increase of mean  
1031 square error produced by removing a variable.

1032

1033 **Table S11. Allele frequency change**

1034 Student's t-test results of allele frequency changes in the locations of Madrid, Tübingen and  
1035 Malmö under the three forecasting Genome Environment Models: free migration, principal  
1036 components control, and geography control.

1037

1038