

Published in final edited form as:

Nat Ecol Evol. 2018 February ; 2(2): 352–358. doi:10.1038/s41559-017-0423-0.

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

Moises Exposito-Alonso¹, François Vasseur^{1,§}, Wei Ding¹, George Wang¹, Hernán A. Burbano^{1,2}, and Detlef Weigel^{1,*}

¹Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

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

Abstract

Because earth is currently experiencing dramatic climate change, it is of critical interest to understand how species will respond to it. The chance of a species to withstand climate change will likely depend on the diversity within the species and, particularly, whether there are subpopulations that are already adapted to extreme environments. However, most predictive studies ignore that species comprise genetically diverse individuals. We have identified genetic variants in *Arabidopsis thaliana* that are associated with survival of an extreme drought event, a major consequence of global warming. Subsequently, we determined how these variants are distributed across the native range of the species. Genetic alleles conferring higher drought survival showed signatures of polygenic adaptation, and were more frequently found in Mediterranean and Scandinavian regions. Using geo-environmental models, we predicted that Central European, but not Mediterranean, populations might lag behind in adaptation by the end of the 21st century. Further analyses showed that a population decline could nevertheless be compensated by natural selection acting efficiently over standing variation or by migration of adapted individuals from populations at the margins of the species' distribution. These findings highlight the importance of within-species genetic heterogeneity in facilitating an evolutionary response to a changing climate.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding author. 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).

Author contributions

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

The authors declare no competing financial interest.

Keywords

climate change; polygenic adaptation; GWA; environmental niche models; random forest; drought; *Arabidopsis thaliana*; image processing

Ongoing climate change has already shifted latitudinal and altitudinal distributions of many plant species¹. Future changes in distributions by local extinctions and migrations are most commonly inferred from niche models that are based on current climate across species ranges^{2,3}. Such approaches, however, ignore that an adaptive response can occur also *in situ* if there is sufficient variation in genes responsible for local adaptation^{4–6}. The plant *Arabidopsis thaliana* is found under a wide range of contrasting environments, making it distinctively suited for studying evolutionary adaptation to a changing climate^{7–9}. For the next 50 to 100 years, extreme drought events, potentially one of the strongest climate change-related selective pressures¹⁰, are predicted to become pervasive across the Eurasian range of *A. thaliana*^{2,11}. An attractive hypothesis is that populations from the Southern edge of the species' range¹² provide a reservoir of genetic variants that can make individuals resistant to future, more extreme, climate conditions^{12,13}. To investigate the potential of *A. thaliana* to adapt to extreme drought events, we first linked genetic variation to survival under an experimental extreme-drought treatment. By combining genome-wide association (GWA) techniques that capture signals of local and/or polygenic adaptation¹⁴ with environmental niche models^{8,15}, we then predicted genetic changes of populations under future climate change scenarios. An unexpected result of our predictions is that populations at both the Northern and Southern margins of the species' range will likely more easily adapt to increased extreme drought events, due to these populations carrying a greater spectrum of drought survival alleles.

Results and Discussion

Differential survival to an extreme drought event

We began by exposing a high-quality subset of 211 geo-referenced natural inbred *A. thaliana* accessions¹⁶ to an experimental extreme drought event during the vegetative phase, which killed the plants before they could reproduce (Table S1). After two weeks of normal growth, plants were challenged by a terminal severe drought for over six weeks and imaged every 2–4 days (Fig. 1A) (see Supplementary Methods section 2). To quantify the rate of leaf senescence, a polynomial linear mixed model was fit to the time series of green pixels per pot (Fig. 1B–D, Video S1). The average genotype deviations from the mean quadratic-term in the model provided the best estimate of this survivorship trait in late stages of drought (Supplementary Fig. 3, see details in Supplementary Methods), ranging from -5 to $+5 \times 10^{-4}$ green pixels/day². The most sensitive genotypes survived only about 32 days, while the most resilient plants survived about 15 days longer. Genotype-dependent survival probably reflects both constitutive as well as induced drought responses, i.e., both environment-dependent and -independent behaviors of the tested accessions. Additional environments need to be examined in order to disentangle these two types of responses.

The amount of water available during our drought experiment translates to only about 30-40 mm of monthly rainfall, and as expected, accessions with higher survival come from regions with low precipitation during the warmest season (correlation with climate variable bio18 [www.worldclim.org, ref. 17]: Pearson correlation, $r=-0.19$, $p=0.005$), and specifically with low precipitation during May and June ($r=-0.19$, $p=0.005$) (see Fig. 2A). To further exploit current climatic data, we used 19 bioclimatic variables and random forest models¹⁸ for environmental niche modeling (ENM) to predict the geographic distribution of the drought-survival index across Europe (Fig. 1C). Surprisingly, we found that individuals with higher drought survival were not only likely to be present around the Mediterranean, but also at the opposite end of the species' range in Sweden¹⁹ (Fig. 1C, ENM cross-validation accuracy=89%, Table S10). In contrast to the warm-dry Mediterranean climate, Scandinavian dry periods occur on average at freezing temperatures (Supplementary Fig. 12). Consequently, precipitation might occur as snow and soil water content is frozen, thus water is not accessible to plants, producing a physiological drought response²⁰.

Survival across geographically structured population lineages

We then studied whether the different genetic lineages of *A. thaliana* are locally adapted⁶ to low precipitation regimes via increased drought-survival. Using an extended panel of 762 *A. thaliana* accessions (Table S1) we carried out genetic clustering²¹ and studied population size trajectories²² (Fig. 2). This corroborated the existence of a so-called Mediterranean 'relict' group¹² and ten other derived groups of relictual (e.g. Spanish groups) or other (e.g. Central Europe) origin, as an apparent result of complex migration and admixture processes²³. A generalized linear model indicated that genetic group membership explained a significant amount of drought-survival variance (GLM: $R^2=12.8\%$; $p=4 \times 10^{-5}$), with the North (N) Swedish and Northeastern (NE) Spanish groups each having on average higher survival than the other groups (t -test $p=0.01$). A population graph estimated by Treemix²⁴ suggested a gene flow edge between the Mediterranean and Scandinavian drought-resistant genetic groups, potentially indicative of historical sharing of drought survival alleles (Fig. 2D). Finally, an ENM of the genetic group membership with climatic variables from the accession's geographic origin confirmed that the most important predictive variable of genetic structure was precipitation during the warmest quarter (bio18), followed by mean temperature of the driest quarter (bio9), and minimum temperature of the coldest month (bio6) (ENM accuracy > 95%. Supplementary Fig. 8 and Table S10). As our results indicate that the deepest genetic split parallels contrasts in local precipitation regimes and ability to survive drought, we expect that decline in rainfall could lead to a future loss of certain genetic groups and/or to turnover of genetic diversity¹¹ (see Fig. 12 Supplementary Fig. 8).

The genomic basis of survival

Because the potential of populations to adapt to drought will ultimately depend on specific genetic variants and the selected trait architecture, we identified drought-associated loci with EMMAX²⁵, a genome-wide association (GWA) method. Although genotype-associated variance²⁵ h^2 was relatively high, 50%, no individual SNP was significantly associated with drought survival (minimum $p \sim 10^{-7}$, after FDR or Bonferroni corrections $p > 0.05$) (Supplementary Fig. 5, Table S3). Significant associations in multiple phenotypes have been detected in similarly powered *A. thaliana* experiments²⁶. While multiple testing adjustment

can over-correct p-values and obscure true associations, the absence of significant associations may also be due to (i) polygenic trait architecture, with many small-effect loci²⁷, and/or (ii) confounding by strong population structure, consistent with the association of drought survival with genetic group membership.

Polygenic signal of adaptation

To test for polygenic adaptation, we repeated the GWA analyses with a model that specifically handles both oligo- and polygenic architectures, BSLMM²⁸. BSLMM estimates, among other parameters, the probability that each SNP comes from a group of major-effect loci. Around half of the top non-significant EMMAX SNPs were found to have over 99% probability of belonging to such a major-effect group (Fisher's exact test of overlap, $p=3 \times 10^{-7}$; see Supplementary Methods 3.3). We further tested the polygenic hypothesis using the population genetic approach of Berg & Coop¹⁴. The test is based on the principle that if populations diverge in a specific trait such as drought-survival that is due to many loci, there should be an orchestrated shift in their allele frequencies. After testing some 60 groups of EMMAX SNP hits of variable size and at different ranks, we detected the most significant signal of polygenic adaptation with the group that included the 151 top SNPs (Table S9). The signal was lost for ranks below the top 300-400 EMMAX SNPs (Table S9). We then compared summary statistics of the top 151 SNPs with background SNPs matched in frequency to avoid GWA discovery biases. The top 151 SNPs showed high F_{st} values, consistent with allele frequency differentiation between populations (Supplementary Fig. 5). Tajima's D values were positive (U Mann-Whitney $p < 0.05$), indicating intermediate allele frequencies at the GWA loci (Supplementary Fig. 5), which could be a result of selection favoring alternative alleles in different ecological niches of the species²⁹. The genomic regions containing the top SNPs did not show any evidence for precipitous reductions of haplotypic diversity, as would be expected for hard selective sweeps³⁰ (Supplementary Fig. 5). Together these patterns fit the expectations of local adaptation from a polygenic trait controlled by some hundred loci³¹ — a scenario that should enable a fast response to new environmental shifts.

Ancestry associations suggest a Mediterranean origin of survival alleles

During local adaptation, the relevant loci diverge due to natural selection across populations, which generates a statistical correlation with population groups³². In this situation, the default correction of population structure applied in GWA might obscure some of the true associations. There are cases where F_{st} scans can be useful to identify overly divergent loci that could be involved in local adaptation. However, in cases of strong population structure, the mean genome-wide F_{st} is high³², complicating outlier detection (Supplementary Fig. 4). One can recover relevant variants that are deeply divergent across populations and therefore invisible to conventional GWA by first assigning ancestry to each SNP. Using ChromoPainter³³, which relies on linkage disequilibrium information, we segmented each genome in question into its different population ancestries (here 11 groups). The first outcome of this analysis was that individuals from NW and NE Spain and, to a lesser extent, the Southern Mediterranean (Fig. 2A), have inherited many DNA segments from relictual individuals (Supplementary Fig. 7). In a generalized linear model framework, we then tested whether the ancestries of individuals at a SNP coincided with the observed phenotypic

differences in drought-survival. Performing this “ancestry” genome-wide association (aGWA) and using a permutation correction of p-values (see Supplementary Methods 3.6), we detected 8 distinct peaks ($p < 0.001$, Fig. 3A) including over 1,000 significant SNPs (70 SNPs after linkage disequilibrium pruning) (Table S4). The most prominent peak was located on chromosome 5 and explained over 20% of the variance in drought survival (Table S4). There was no overlap in top SNPs between GWA and aGWA because they search for different association signals. Our aGWA resembles other admixture mapping techniques³⁴, and might be most useful for associations in scenarios of adaptive introgression and local adaptation. Although we do not know yet whether our observations can be generalized, our work demonstrates the power of using alternative GWA approaches in situations where adaptive variation is expected to be tightly linked to population history and structure.

To understand the origin of aGWA-identified SNPs, we constructed trees for all concatenated aGWA SNPs and for genome-wide background SNPs. Although the individuals from both the warm (Iberia and relicts) and cold (Scandinavia) edges of the species distribution are far apart in genome-wide SNPs, they are closely related in drought-associated SNPs (Fig. 3B). Overall, this is consistent with a common Mediterranean origin of drought-adaptive genetic variants of both Northern and Southern individuals (Fig. 2D, Fig. 3B), and highlights the relevance of populations at the latitudinal extremes of the species range as a possible genetic reservoir for future climate change adaptation¹².

Drought survival is a resilience trait independent on phenology

Drought adaptation can be accomplished by diverse mechanisms, with cross-stress resistance being pervasive³⁵. An annual life history enables drought survival through an escape strategy based on the acceleration of the life cycle from germination to flowering and seed production. An alternative strategy, the avoidance strategy, is employed by many xeric perennials with increased water efficiency³⁶. Previous drought experiments with *A. thaliana* have shown that both strategies exist, although early flowering, which is associated with an escape strategy, was more favourable under water-limiting conditions^{37,38}. In our experiment, drought-survival was not negatively correlated with flowering time in unstressed conditions³⁹ (Pearson correlation, $r = 0.07$, $p = 0.12$). Although a correlation was not significant at the individual ecotype level, the GWA effect sizes of drought-survival for the top 151 SNPs were positively correlated with the ability of the same SNPs to delay flowering (Pearson correlation, $r = 0.51$, $p = 1 \times 10^{-11}$, see Supplementary Methods 3.4). Given the described trade-off between escape by flowering and water use efficiency in *A. thaliana*^{37,40,41}, our drought-survival index might be related to the avoidance strategy, although this needs to be tested with specific physiological experiments (Supplementary Fig. 11, Table S6). Gene enrichment analysis revealed a weak signal for membrane transport (see Supplementary Methods 3.7). Adjustment of osmotic balance through cell membrane transport is a drought avoidance mechanism⁴² that might also confer cross-tolerance to other abiotic stresses⁴³. Therefore, it might be of relevance for Scandinavian *A. thaliana* accessions or other populations in extreme environments (Supplementary Fig. 12)¹⁹.

Forecast of genetic changes to global warming reveals regional differences in evolutionary potential

It is expected that populations with increased survival to severe abiotic stresses should have an evolutionary advantage in face of the predicted increase in drought frequency and intensity both around the Mediterranean and in Europe, which will constitute a critical hazard for many plants^{2,11}, including *A. thaliana*. Surprisingly, environmental niche models (ENM) of species distributions, which have been used to predict future changes of species' ranges^{2,3}, do not usually include information of within-species diversity that can lead to adaptation from standing variation^{44–46}. This could in turn lead to overestimates of extinction rates^{47–49}. By fitting ENMs of current climate with SNP data, using a similar rationale as for the “climate GWA” of Hancock and colleagues⁷, we attempted to forecast the most likely genetic makeup under current and future climate conditions. We trained one ENM for each of the 151 GWA and 70 aGWA drought-associated SNPs to predict which allele, either the high or the low survival one, is more likely, given a set of environmental variables (all ENM 5CV accuracy >92%; Table S3-4, Supplementary Fig. 13-16). Consequently, from each model, we geographically mapped the potential distribution of the high survival allele using available environmental datasets (www.worldclim.org, ref. 17). Finally, concatenating the resulting 221 maps, we inferred the most likely individual genotype at each location. At present, individuals from both northern and southern edges of the species' Eurasian and N. African range are predicted to harbor more drought-survival alleles than those located in between (Fig. 3C, Supplementary Fig. 15-16, with the quadratic term in a regression of allele count on latitude being positive at $p=10^{-3}$), corroborating our previous observations. Using the trained ENM, we also forecast the distribution of the 221 drought-survival alleles in 2070 (rpc 8.5, IPCC, www.ipcc.ch, ref. 17). While it was expected that populations in the Mediterranean Basin need to become more drought resistant¹¹, our predictions anticipate a greater increase in the total number of drought-survival alleles for Central Europe (Fig. 3, Supplementary Fig. 14-15). This is because by 2070 rainfall in Central Europe will likely become more similar to that in the Mediterranean^{2,11} (Supplementary Fig. 12).

Because some drought-survival alleles are currently not present in Central Europe, we speculated that gene migration might be necessary to facilitate adaptation to future conditions⁵⁰. An underlying assumption of the ENM is that alleles will be present wherever required by the environment, but this assumption of “universal migration” may not be realistic for future predictions if the presence of alleles is currently geographically restricted. We therefore included two geographic boundary conditions in the ENM to generate alternative models that were either more or less “migration-limited” (see Supplementary Methods 4.2). After fitting all possible models and predicting allele distributions with future climate, we calculated the difference of predicted allele presence per map grid cell between the naïve, free migration ENM and the two geographically constrained ones (Fig. 3D-E). If an allele has currently a narrow distribution or is specific to a certain genetic background, its future presence in an area might not be predicted by the constrained models, even though the climate variables coincide with the SNP's environmental range. Such a scenario seems to apply to Central Europe, as the deficit in drought-survival alleles predicted by the free over the constrained models was 8-30% (18-66 out of 221) (Fig. 3E; with the quadratic term in a

regression of the allele count difference on latitude being negative at $p < 10^{-10}$). Central European populations may therefore be under threat of lagging adaptation by the end of the 21st century.

In the end, for a population to persist, not only must drought-survival alleles be present locally, but they also need to increase in frequency⁵¹. The chance of this occurring will depend on current local allele frequencies and the strength of natural selection favouring the drought-survival alleles. Therefore, we studied current allele frequencies at three representative locations with the highest sampling density in our dataset (40 samples within a 50 kilometer area): Madrid (Spain), Tübingen (Germany) and Malmö (Sweden), which are near the southern edge, center and northern edge of the Eurasian and N. African range, respectively. Based on ENM predictions, we calculated allele frequency changes from present to 2070. Frequencies are predicted to increase significantly only in the Tübingen population (Student's t test, $p < 10^{-16}$, Table S11), but not in Madrid and Malmö, indicating that these two populations might already be adapted to the future local climate. Although not all drought-associated alleles are found in Tübingen (32 of 70 aGWA SNPs and 136 of 151 GWA SNPs), increasing the number of the alleles in single genotypes should be feasible, since there are already single genotypes that have 24 (aGWA) and 123 (GWA) of these alleles (see Supplementary Methods 4.2). Running 50-generations simulations starting at the present Tübingen frequency of each of the drought-survival alleles and assuming a range of selection coefficients, we estimated that a 1-3% of fitness advantage on average would be necessary to increase frequencies to match those of the adapted Madrid and Malmö populations (Supplementary Fig. 17, see Supplementary Methods 4.2). Such selection could take place efficiently when populations are large, as is typical for highly-proliferative weeds^{51,52}.

Conclusion

Leveraging the genetic resources available for *A. thaliana*, we have begun to address the question of how climate change will affect biodiversity. We provide evidence for the possibility of adaptive genetic variation to extreme drought events from standing variation. Specifically, we found that drought survival in *A. thaliana* has a polygenic basis and that favorable alleles are more abundant toward the edges of the species' distribution range. Extreme adaptation at range edges might thus be critical for a species' persistence under climate change. Although many aspects of future adaptation are not considered here, namely non-drought related or seasonal climate change⁵¹, biotic interactions, phenotypic plasticity, or novel adaptive mutations, our spatially explicit analyses emphasize the potential of adaptive evolution from standing variation to mitigate climate change's detrimental effects.

Methods

Study populations

211 natural inbred lines from the 1001 Genomes project¹⁶ were grown in a terminal drought experiment, and 762 lines were analyzed for genetic structure and genome-environment models. These two subsets were selected based on sequence quality and homogeneity of geographic distribution (see Supplementary Methods 1.1). We retrieved the genomes

corresponding to the above natural lines from <http://1001genomes.org/data/GMI-MPI/releases/v3.1/> and extracted the biallelic SNPs with >95% calling rate. This resulted in keeping ~4M SNP.

Genetic structure

To understand the genetic structure of *Arabidopsis thaliana* we ran, on the 762 samples, the software ADMIXTURE v1.2 (ref. 21) assuming two to 20 groups and using a 5-fold cross-validation procedure. The number of groups with the smallest cross-validation error was 11 (Fig. 2, Fig. Supplementary Table 1, Supplementary Fig. 8). We computed a genomic PCA using PLINK v1.9 (ref. 53). The three first PC axes explained 33.5% of the genomic variance (see Supplementary Methods 3).

We used genomes with probability >0.9 of assignment to one of the 11 ADMIXTURE groups to run MSMC v.3 (ref. 22). This was done in quartets of genomes, i.e. four genomes for within-population coalescent mode, and two genomes of each of two populations for the cross-coalescent mode (Fig. 2, Supplementary Fig. 5). Using the 11 genetic groups as population lineages, we run Treemix assuming zero to five migration edges²⁴ (Fig. 2, Supplementary Fig. 5)

Terminal drought experiment

Stratified seeds from the selected 211 natural lines were sown in greenhouse pots and abundantly watered every three days during two weeks. Thereafter watering only occurred every three weeks, which dramatically reduced soil water content (Fig. 1, Supplementary Methods 1.2). Top-view photographs of the potting trays were done at 20 timepoints during the whole experiment with a high resolution Panasonic DMC-TZ61 digital camera mounted in a closed black box setting to ensure image consistency (Supplementary Methods 2). Using customized Python scripts and the module Open Computer Vision, we segmented the green plant-leave pixels from the brown soil background to monitor plant area over time (Supplementary Video). Starting from the day with the largest rosettes areas, until the end of the experiment, we modeled the decay of green area (i.e. # pixels) using a polynomial generalized linear mixed model with Poisson link as described in the MCMCglmm R package v.2.25 (see Supplementary Methods 2). The random genotype effects captured the average deviation of each genotype from a general intercept, slope and quadratic curvature. After calculating the heritability of each of the three coefficient deviations and their correlation with the genotype's climate variables of origin, we understood that it was the quadratic curvature that was the most suitable to use as index of survival (Supplementary Methods 2).

Genome-Wide Association (GWA)

Using the index of survival per genotype as the trait and the SNPs with a minimum allele frequency > 5% as predictors (n=879,654 SNPs), we carried out associations using the linear mixed model implemented in EMMAX software²⁵ to find SNPs that excessively contributed to the prediction of survival of genotypes (Supplementary Table 3) (see Supplementary Methods 3.3). To corroborate the identified top SNPs we also performed a Bayesian Sparse Linear Mixed Model (BSLMM) with GEMMA software²⁸. Both approaches, EMMAX and

GEMMA, fit a model as: $Y = X_i\beta + Zu + \epsilon$, where Y is the vector of trait values, X is the alternative allele dosage at SNP i and β the allelic effect of SNP i on the trait. Population structure is corrected with a random genotype term (of 211 levels) represented by u , which follows a Multivariate Normal distribution $\mathcal{N}(0, A\sigma_G^2)$, where A is the relationship matrix between all individual genotypes built from SNP information and σ_G^2 is the genotype-associated variance. Different from EMMAX, the BSLMM model considers that the β coefficients follow a mixture of two distributions, one that expects many small effects and another that generates few strong effects.

To determine whether the top SNPs identified in the GWA might have been subject to polygenic adaptation, we used the method from Berg & Coop¹⁴. We did this for several groupings of top SNPs and reported the group that yielded the strongest signal (see all results in Supplementary Table 9).

Using painted chromosomes generated using ChromoPainter v. 2.0.7 (ref. 33), we carried out another set of associations between the survival trait and the local ancestry category (11 groups) of a chunk of the genome. We used a linear model, $Y = \mu + X\beta + \epsilon$, and reported the positions in the genome with the least mean square error (i.e. highest R^2) (Supplementary Table 4). To compute p-values, we took an empirical p-value distribution approach based on 1,000 random permutation runs (see Supplementary Methods 3.6). To understand the ancestry of the associated genomic positions, we concatenated the SNP genotypes of the top-associated positions, computed genetic distances between natural lines and generated a Neighbour Joining tree. This tree was compared with a tree built from an equal number of randomly-picked background SNPs.

Genome-wide diversity and selection summary statistics

We calculated genome-wide F_{st} among the ADMIXTURE-defined groups and Tajima's D with PLINK v1.9 (ref. 53) and likelihood of a selective sweep with SweeD (ref. 30). We investigated the enrichment of the top SNPs in the upper tail of the distributions of those statistics by calculating a right-tailed t-test in contrast with genome-background SNPs with the same frequency values (Supplementary Fig. 4, Supplementary Table 3, rank columns).

Environmental Niche Models

We used classification and regression Random Forest models implemented in the *randomForest* R package, available environmental databases www.worldclim.com v.1.4 (ref. 17, 19 bioclimatic variables at 2.5 arc-minutes resolution), and geographic locations of GWA-identified alleles, to fit environmental niche models (ENM). To evaluate model's predictive ability for each allele, we used a 5-fold cross-validation procedure in which $\frac{4}{5}$ parts of the data were used to train the model and $\frac{1}{5}$ was used to test it. This enabled us to assign a percentage of successful assignment of an allele given the environmental variables at a location (Supplementary Tables 3-4). The fitted Random Forest model was used to generate potential geographic distributions of survival-associated alleles which, all overlapped, provided a geographic map of density of survival alleles. Using existing predictions of the same 19 bioclimatic variables to 2050 and 2070 under both low (2.6 rcp) and high (8.5 rcp) CO₂ accumulation scenarios, we re-predicted the distribution of alleles to

the different future scenarios using the previously fitted Random Forest models. Because of the implicit assumption of free movement of alleles, we generated two additional models per SNP: (1) ENM including the latitude and longitude variables in the Random Forest models and (2) ENM including the three first PC axes geographically modeled with present day climate (see below). By repeating predictions with future climate data, but keeping the latitude, longitude and PC components constant, some alleles would not be predicted in areas where the appropriate environment exists but which are outside of the current geographic distribution (1) or current local genomic background (2) (see Supplementary Methods 4, Supplementary Fig. 13-16).

Apart from the potential distribution of putatively adaptive alleles, we also modeled the geographic distribution of continuous traits, namely the aforementioned PCA components of population structure or the index of survival under drought itself. In those cases the Random Forest was of the regression type and the predictive ability was computed for the test data calculating the squared Pearson's correlation coefficient between predicted and true values (see Supplementary Methods 4).

To complement observations of presence and absence of alleles from ENM predictions, we carried out Wright-Fisher simulations of single biallelic SNPs (for details see Supplementary Methods 4.2.4). We ran simulations for 50 discrete generations. The population size was assumed of 300,000 plants, as inferred from diversity data, and was constant over time. Fitness was only determined by the selection coefficient of the drought alleles, which varied from 0 to 20% in an array of simulation runs. The starting frequency of the allele was set equal to the present day frequency of all natural lines sampled in a given geographic area (e.g., Tübingen). These simulations could be extended in the future to incorporate joint fitness effects from multiple adaptive mutations and complex environment-driven demographic processes (Supplementary Methods 4.2.4).

Code availability

Code for the image analysis pipeline available at <http://github.com/MoisesExpositoAlonso/hippo> with DOI: <https://doi.org/10.5281/zenodo.1039888>, code for ancestryGWA is available at <https://github.com/MoisesExpositoAlonso/aGWA> with DOI: <https://doi.org/10.5281/zenodo.1039882>, code for Wright-Fisher population simulations at <http://github.com/MoisesExpositoAlonso/popgensim> with DOI: <https://doi.org/10.5281/zenodo.1039886>.

Data availability

Phenotypic datasets available in the Supplementary Dataset. Processed genome matrices are available at <http://1001genomes.org/data/GMI-MPI/releases/v3.1/>. Raw reads are stored in the www.ncbi.nlm.nih.gov/sra archive under the ID number: SRP056687.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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

References

1. Parmesan C, Yohe G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*. 2003; 421:37–42. [PubMed: 12511946]
2. Thuiller W, Lavorel S, Araújo MB, Sykes MT, Prentice IC. Climate change threats to plant diversity in Europe. *Proc Natl Acad Sci U S A*. 2005; 102:8245–8250. [PubMed: 15919825]
3. Jezkova T, Wiens JJ. Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. *Proc R Soc B*. 2016; 283:20162104.
4. Barrett RDH, Schluter D. Adaptation from standing genetic variation. *Trends Ecol Evol*. 2008; 23:38–44. [PubMed: 18006185]
5. Hereford J. A quantitative survey of local adaptation and fitness trade-offs. *Am Nat*. 2009; 173:579–588. [PubMed: 19272016]
6. Turesson G. The species and the variety as ecological units. *Hereditas*. 1922; 3:100–113.
7. Hancock AM, et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*. 2011; 334:83–86. [PubMed: 21980108]
8. Fournier-Level A, et al. A map of local adaptation in *Arabidopsis thaliana*. *Science*. 2011; 334:86–89. [PubMed: 21980109]
9. Lasky JR, et al. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol Ecol*. 2012; 21:5512–5529. [PubMed: 22857709]
10. Siepielski AM, et al. Precipitation drives global variation in natural selection. *Science*. 2017; 355:959–962. [PubMed: 28254943]
11. Dai A. Increasing drought under global warming in observations and models. *Nat Clim Chang*. 2012; 3:52–58.
12. Hampe A, Petit RJ. Conserving biodiversity under climate change: the rear edge matters. *Ecol Lett*. 2005; 8:461–467. [PubMed: 21352449]
13. Lee-Yaw JA, et al. A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. *Ecol Lett*. 2016; doi: 10.1111/ele.12604
14. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *PLoS Genet*. 2014; 10:e1004412–e1004412. [PubMed: 25102153]
15. Dormann CF, et al. Correlation and process in species distribution models: bridging a dichotomy. *J Biogeogr*. 2012; 39:2119–2131.
16. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell*. 2016; 166:481–491. [PubMed: 27293186]
17. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 2005; 25:1965–1978.
18. Breiman L. Random Forests. *Mach Learn*. 2001; 45:5–32.
19. Mojica JP, et al. Genetics of water use physiology in locally adapted *Arabidopsis thaliana*. *Plant Sci*. 2016; 251:12–22. [PubMed: 27593459]
20. Ingram J, Bartels D. The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Physiol Plant Mol Biol*. 1996; 47:377–403. [PubMed: 15012294]
21. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009; 19:1655–1664. [PubMed: 19648217]
22. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014; 46:919–925. [PubMed: 24952747]

23. Lee C-R, et al. On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat Commun.* 2017; 8:14458. [PubMed: 28181519]
24. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012; 8:e1002967. [PubMed: 23166502]
25. Kang HM, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet.* 2010; 42:348–354. [PubMed: 20208533]
26. Atwell S, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature.* 2010; 465:627–631. [PubMed: 20336072]
27. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet.* 2011; 13:135–145.
28. Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 2013; 9:e1003264. [PubMed: 23408905]
29. Hedrick PW. Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annu Rev Ecol Evol Syst.* 2006; 37:67–93.
30. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol.* 2013; 30:2224–2234. [PubMed: 23777627]
31. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 2010; 20:R208–15. [PubMed: 20178769]
32. Josephs EB, Stinchcombe JR, Wright SI. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol.* 2017; doi: 10.1111/nph.14410
33. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet.* 2012; 8:e1002453. [PubMed: 22291602]
34. Shriner D, Adeyemo A, Ramos E, Chen G, Rotimi CN. Mapping of disease-associated variants in admixed populations. *Genome Biol.* 2011; 12:223. [PubMed: 21635713]
35. Tardieu F. Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. *J Exp Bot.* 2012; 63:25–31. [PubMed: 21963615]
36. Ludlow, MM. Strategies of response to water stress. Structural and functional responses to environmental stress. Kreeb, KH, Richter, H., Minckley, TM., editors. The Hague, the Netherlands: SPB Academic; 1989. p. 269-281.
37. Kenney AM, McKay JK, Richards JH, Juenger TE. Direct and indirect selection on flowering time, water-use efficiency (WUE, $\delta 13C$), and WUE plasticity to drought in *Arabidopsis thaliana*. *Ecol Evol.* 2014; 4:4505–4521. [PubMed: 25512847]
38. Bac-Molenaar JA, Granier C, Keurentjes JJB, Vreugdenhil D. Genome-wide association mapping of time-dependent growth responses to moderate drought stress in *Arabidopsis*. *Plant Cell Environ.* 2016; 39:88–102. [PubMed: 26138664]
39. Vasseur F, Wang G, Bresson J, Schwab R, Weigel D. Image-based methods for phenotyping growth dynamics and fitness in large plant populations. *bioRxiv.* 2017; 208512. doi: 10.1101/208512
40. Juenger TE, et al. Identification and characterization of QTL underlying whole-plant physiology in *Arabidopsis thaliana*: $\delta 13C$, stomatal conductance and transpiration efficiency. *Plant Cell Environ.* 2005; 28:697–708.
41. McKay JK, Richards JH, Mitchell-Olds T. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol Ecol.* 2003; 12:1137–1151. [PubMed: 12694278]
42. Jarzyniak KM, Jasiński M. Membrane transporters and drought resistance - a complex issue. *Front Plant Sci.* 2014; 5:687. [PubMed: 25538721]
43. Swindell WR. The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics.* 2006; 174:1811–1824. [PubMed: 17028338]
44. Pauls SU, Nowak C, Bálint M, Pfenninger M. The impact of global climate change on genetic diversity within populations and species. *Mol Ecol.* 2013; 22:925–946. [PubMed: 23279006]
45. Brown JL, et al. Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *Am J Bot.* 2016; 103:153–163. [PubMed: 26747843]

46. Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett.* 2015; 18:1–16. [PubMed: 25270536]
47. Catullo RA, Ferrier S, Hoffmann AA. Extending spatial modelling of climate change responses beyond the realized niche: estimating, and accommodating, physiological limits and adaptive evolution. *Glob Ecol Biogeogr.* 2015; 24:1192–1202.
48. Moritz C, Agudo R. The future of species under climate change: resilience or decline? *Science.* 2013; 341:504–508. [PubMed: 23908228]
49. Hoffmann AA, Sgrò CM. Climate change and evolutionary adaptation. *Nature.* 2011; 470:479–485. [PubMed: 21350480]
50. Aitken SN, Whitlock MC. Assisted gene flow to facilitate local adaptation to climate change. *Annu Rev Ecol Evol Syst.* 2013; 44:367–388.
51. Fournier-Level A, et al. Predicting the evolutionary dynamics of seasonal adaptation to novel climates in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 2016; 113:E2812–21. [PubMed: 27140640]
52. Roux F, Giancola S, Durand S, Reboud X. Building of an experimental cline with *Arabidopsis thaliana* to estimate herbicide fitness cost. *Genetics.* 2006; 173:1023–1031. [PubMed: 16582450]
53. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]

One-sentence summary

“Future genetic changes in *A. thaliana* populations could be forecast by combining climate change models with genomic predictions based on experimental phenotypic data.”

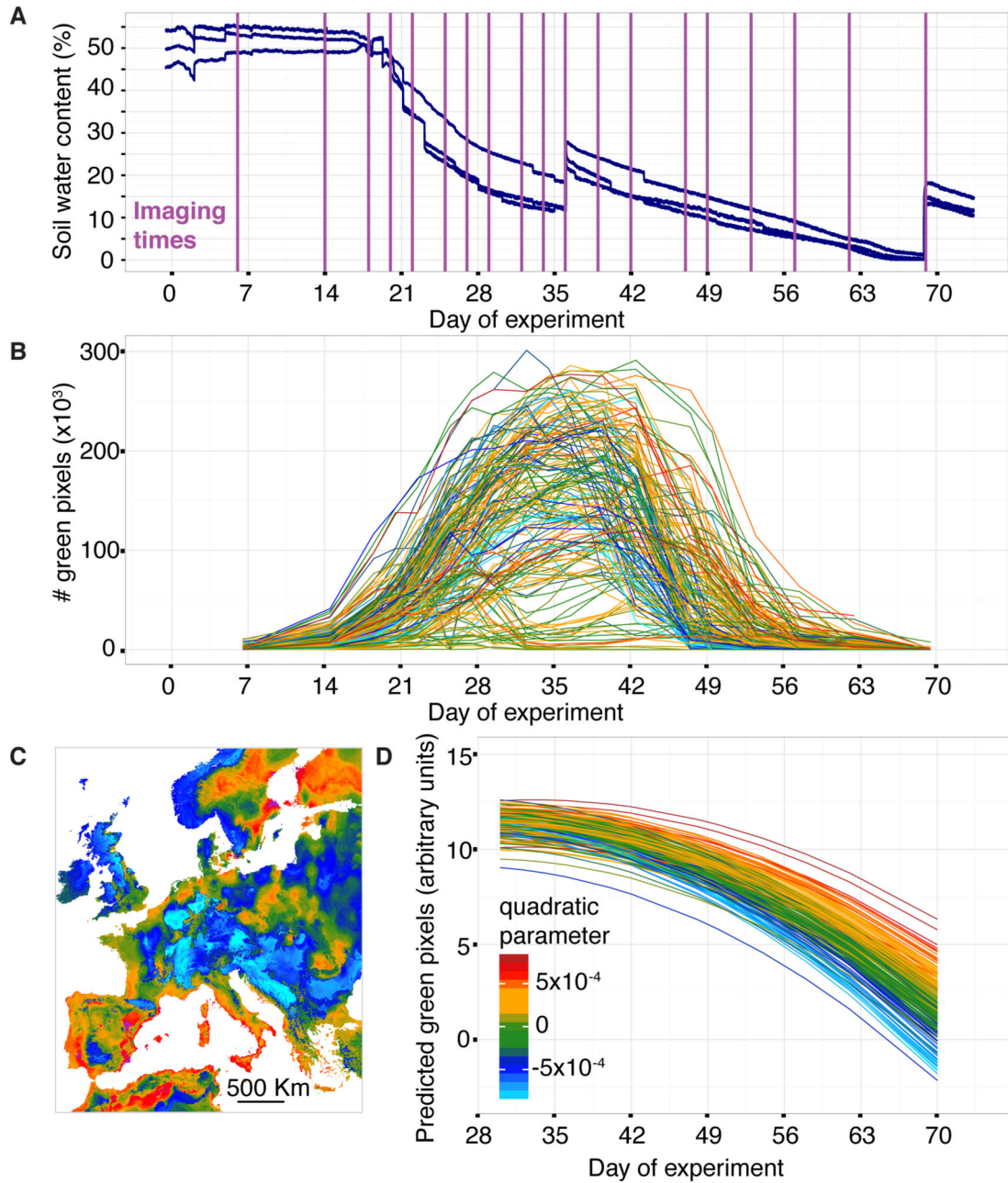


Figure 1. Terminal drought treatment and phenotyping of 211 accessions.

(A) Soil water content as measured by sensors in three well spaced experimental trays. Vertical lines indicate dates of image acquisition. (B) Trajectories of total rosette area of 200 randomly chosen pots (see Video S1). Color index according to quadratic parameter in (D). (C) Map projection of the environmental niche model prediction of the quadratic parameter (the drought-survival index) in (D). (D) Decay trajectory modeled with a polynomial regression, with genotypes as random factors, from the day of maximum number of green pixels until the end of the experiment. Each line corresponds to one genotype.

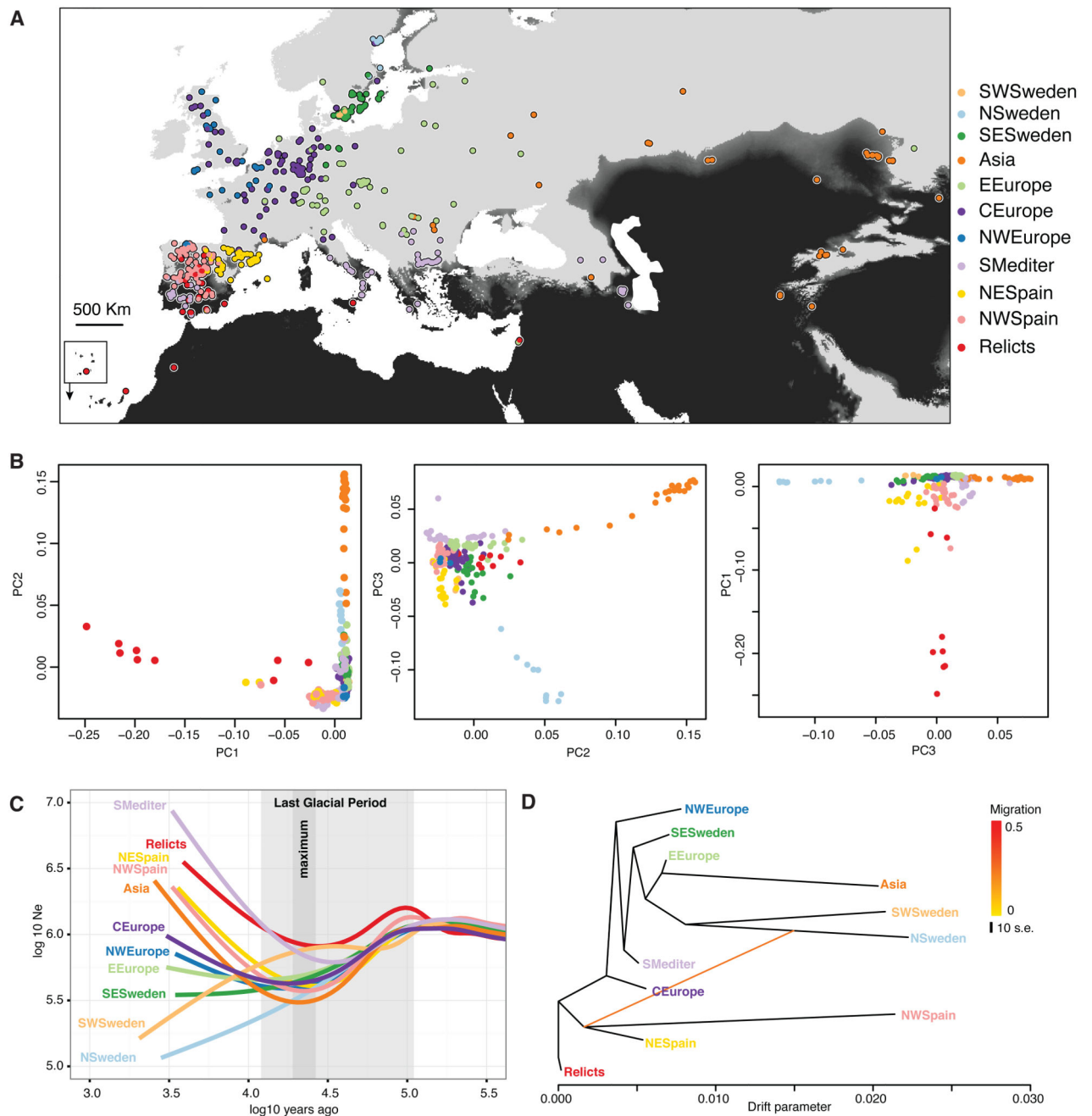


Figure 2. Population structure and history of 762 high-quality genomes.

(A) Geographic locations and 11 genetic clusters estimated by ADMIXTURE ($k=11$ having the lowest cross-validation error). Black indicates less than 40 mm of June rainfall (1960 to 1990 average), which corresponds to the amount of water provided in our drought experiment (Fig. 1). Note areas of very low June rainfall in the Mediterranean basin and along the coast in Scandinavia (partially obscured by colored circles). Cape Verde Islands are shown as inset. (B) Principal Component Analysis of genome-wide SNPs. (C) Effective

population sizes in time estimated from MSMC. **(D)** Population ancestral graph and the first migration trajectory from Treemix.

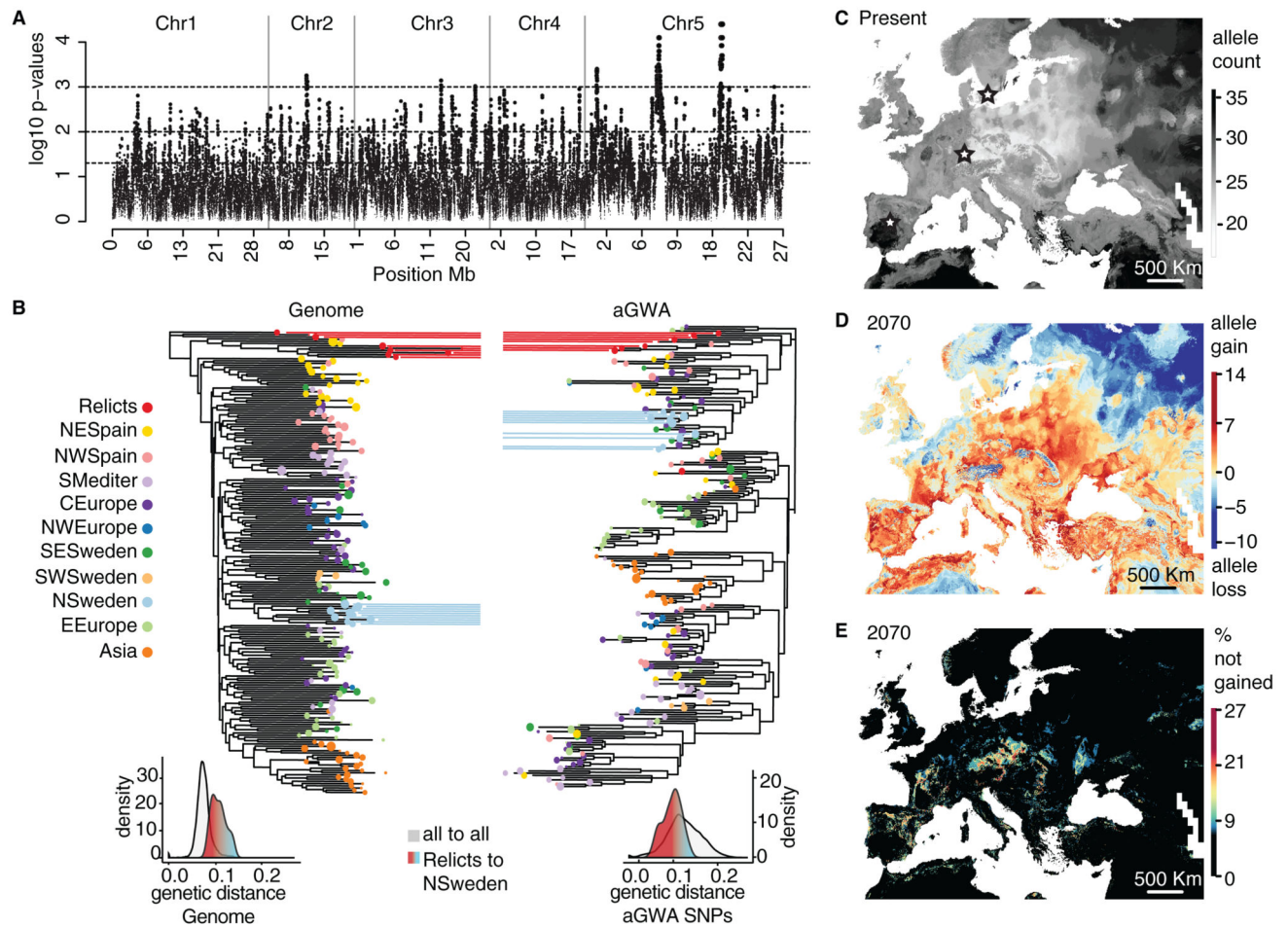


Figure 3. Ancestry GWA of drought survival and environmental predictions.

(A) Manhattan plot of SNPs from ancestry GWA (aGWA) after permutation correction of p-values. Dashed lines indicate significant thresholds at $p < 0.05$, 0.01, and 0.001. (B) Top, Neighbour-joining phylogeny of 1,000 concatenated genome-wide SNPs compared with a phylogeny of all significant aGWA SNPs (ca. 1,000). Colors indicate population clusters (Fig. 2). Relicts and N. Swedish groups are highlighted. Bottom, genetic distances for genome-background SNPs or aGWA SNPs. (C) Environmental niche models of 70 top aGWA SNPs (after LD pruning), trained with climate averages from 1960-1990, and then (D) used to forecast gain or loss of alleles in 2070 under free migration. (E) Discrepancy of alleles that can be gained by 2070 between the geographically constrained (PCA control) model and the free migration model.