ARTICLES
https://doi.org/10.1038/s41559-017-0423-0
nature ecology & evolution

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

Moises Exposito-Alonso[1], François Vasseur[1,3], Wei Ding[1], George Wang[1,4], Hernán A. Burbano[1,2] and Detlef Weigel[1*]

As 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 withstanding climate change is likely to depend on the diversity within the species and, particularly, whether there are sub-populations 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 twenty-first 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.

Ongoing climate change has already shifted latitudinal and altitudinal distributions of many plant species[1]. Future changes in distributions by local extinctions and migrations are most commonly inferred from niche models that are based on the current climate across species ranges[2,3]. Such approaches, however, ignore that an adaptive response can also occur in situ if there is sufficient variation in the 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 50–100 years, extreme drought events, potentially one of the strongest climate change-related selective pressures[10], 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' range[12] 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 the signals of local and/or polygenic adaptation[14] with environmental niche models (ENMs)[8,15], we predicted the 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 responses to an extreme drought event. We began by exposing a high-quality subset of 211 geo-referenced natural inbred *A. thaliana* accessions[16] to an experimental extreme drought event during the vegetative phase, which killed the plants before they could reproduce (Supplementary Table 1). After two weeks of normal growth, the plants were challenged by a terminal severe drought for over six weeks and imaged every two to four days (Fig. 1a; see Supplementary Methods 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 and Supplementary Video). The average genotype deviations from the mean quadratic term in the model provided the best estimate of this survivorship trait in the late stages of drought (Supplementary Fig. 3; see details in Supplementary Methods), ranging from -5 to $+5\times10^{-4}$ green pixels day^{-2} (ref. [2]). 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 and induced drought responses; that is, both environment-dependent and -independent behaviours of the tested accessions. Additional environments need to be examined 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's coefficient of correlation (r) $=-0.19$, $P=0.005$) and specifically with

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany. [2]Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany. Present address: [3]Centre National de la Recherche Scientifique, Unités Mixtes de Recherche 5175, Centre d'Ecologie Fonctionnelle et Evolutive, Montpellier, France. [4]Computomics, Davis, CA, USA. *e-mail: weigel@weigelworld.org

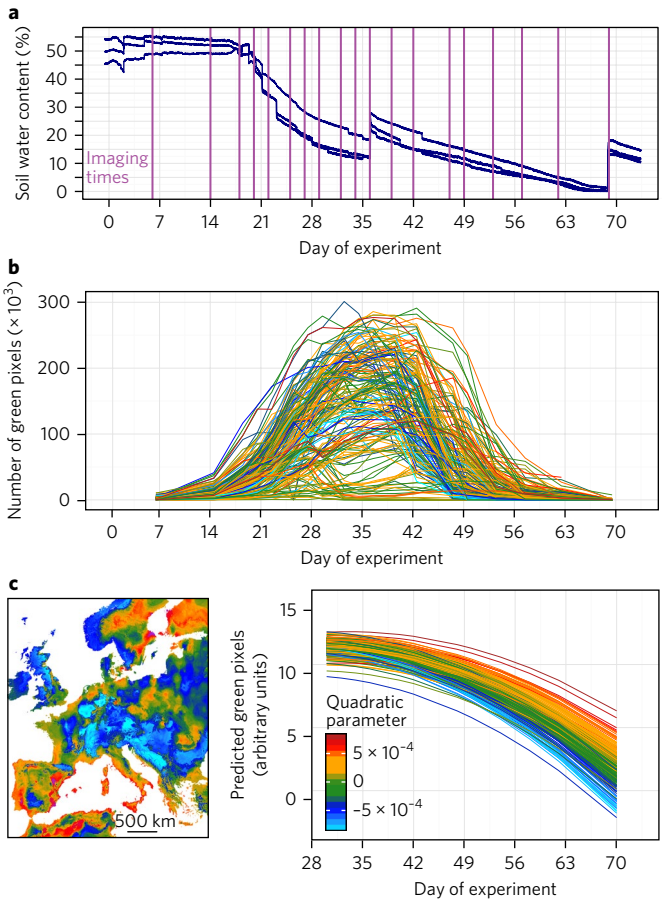

Fig. 1 | Terminal drought treatment and phenotyping of 211 accessions. **a**, Soil water content as measured by sensors in three well-spaced experimental trays. The vertical lines indicate the dates of image acquisition. **b**, Trajectories of the total rosette area of 200 randomly chosen pots (see Supplementary Video). The colour index is according to the quadratic parameter in **d**. **c**, Map projection of the ENM prediction of the quadratic parameter (the drought-survival index) in **d**. **d**, Decay trajectory modelled with a polynomial regression, with genotypes as random factors, from the day of the maximum number of green pixels until the end of the experiment. Each line corresponds to one genotype.

low precipitation during May and June ($r \leq -0.19$, $P \leq 0.005$) (see Fig. 2a). To further exploit current climatic data, we used 19 bioclimatic variables and random forest models[18] for environmental niche modelling 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 Sweden[19] (Fig. 1c, ENM cross-validation accuracy = 89%; Supplementary Table 10). In contrast with 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 response[20].

Survival across geographically structured population lineages. We then studied whether the different genetic lineages of *A. thaliana* are locally adapted[6] to low precipitation regimes via increased drought survival. Using an extended panel of 762 *A. thaliana* accessions (Supplementary Table 1), we carried out genetic clustering[21] and studied population size trajectories[22] (Fig. 2). This corroborated the existence of a so-called Mediterranean 'relict' group[12] and ten other derived groups of relictual (for example, Spanish groups) or other (for example, Central Europe) origin, as an apparent result of complex migration and admixture processes[23]. A linear model indicated that genetic group membership explained a significant amount of drought-survival variance (adjusted $R^2 = 12.8\%$; $P = 4 \times 10^{-5}$), with the north Swedish and northeastern Spanish groups each having, on average, higher survival than the other groups (*t*-test: $P \leq 0.01$). A population graph estimated by Treemix[24] 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 the mean temperature of the driest quarter (bio9) and the minimum temperature of the coldest month (bio6) (ENM accuracy > 95%; Supplementary Fig. 8 and Supplementary Table 10). As our results indicate that the deepest genetic split parallels contrast in local precipitation regimes and ability to survive drought, we expect that a decline in rainfall could lead to future loss of certain genetic groups and/or turnover of genetic diversity[11] (see Supplementary Figs. 8 and 12).

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 EMMAX[25], a GWA method. Although genotype-associated variance[25], h^2, was relatively high (50%), no individual single nucleotide polymorphism (SNP) was significantly associated with drought survival (minimum $P \sim 10^{-7}$; after false discovery rate or Bonferroni corrections: $P > 0.05$) (Supplementary Fig. 5 and Supplementary Table 3). Significant associations in multiple phenotypes have been detected in similarly powered *A. thaliana* experiments[26]. While multiple testing adjustment can over-correct *P* values and obscure true associations, the absence of significant associations may also be due to (1) the polygenic trait architecture, with many small-effect loci[27] and/or (2) 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—the Bayesian sparse linear mixed model (BSLMM)[28]. The 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 ref. [14]. This test is based on the principle that if populations diverge in a specific trait such as drought survival 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 (Supplementary Table 9). The signal was lost for ranks below the top 300–400 EMMAX SNPs (Supplementary Table 9). 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 fixation index (F_{ST}) values, consistent with allele frequency differentiation between populations (Supplementary Fig. 5). Tajima's D values were positive (Mann–Whitney *U*-test, $P < 0.05$), indicating intermediate allele frequencies at the GWA loci (Supplementary Fig. 5), which could be a result of selection favouring alternative alleles in

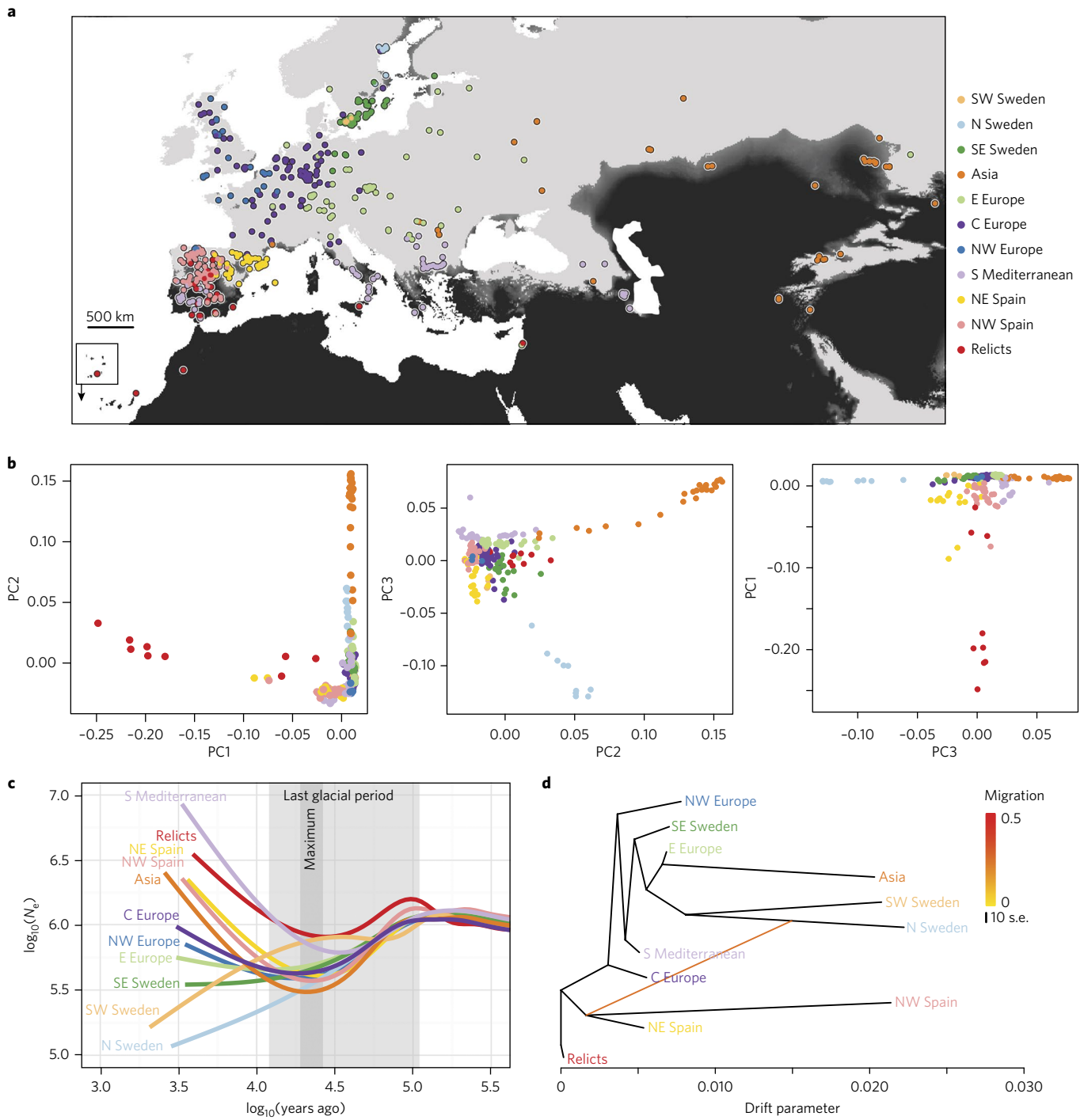

Fig. 2 | Population structure and history of 762 high-quality genomes. a, Geographic locations of 11 genetic clusters estimated by ADMIXTURE ($k = 11$ having the lowest cross-validation error). Black indicates less than 40 mm of June rainfall (1960–1990 average), which corresponds to the amount of water provided in our drought experiment (Fig. 1). Note the areas of very low June rainfall in the Mediterranean basin and along the coast in Scandinavia (partially obscured by coloured circles). The Cape Verde Islands are shown as an inset. **b**, Principal component (PC) 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. C, central; E, east; N, north; NE, northeast; NW, northwest; S, southern; SE, southeast; SW, southwest.

different ecological niches of the species[29]. 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 sweeps[30] (Supplementary Fig. 5). Together, these patterns fit the expectations of local adaptation from a polygenic trait controlled by some hundred loci[31]—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 groups[32]. 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

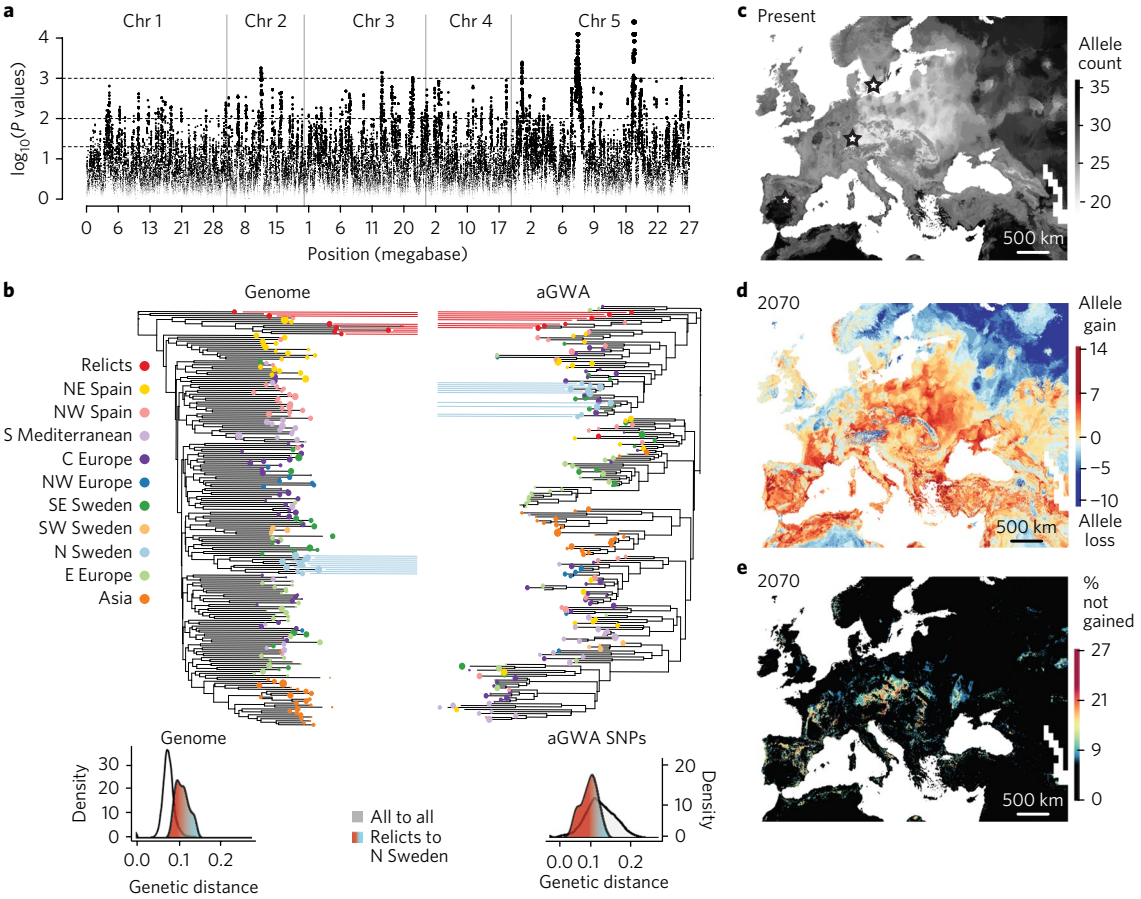

Fig. 3 | aGWA of drought survival and environmental predictions. a, Manhattan plot of SNPs from aGWA after permutation correction of *P* values. The dashed lines indicate significant thresholds at $P < 0.05$, $P < 0.01$ and $P < 0.001$. Chr, chromosome. **b**, Top: neighbour-joining phylogeny of 1,000 concatenated genome-wide SNPs compared with a phylogeny of all significant aGWA SNPs (around 1,000). The colours indicate population clusters (Fig. 2). Relicts and north Swedish groups are highlighted. Bottom: genetic distances for genome background SNPs or aGWA SNPs. C, central; E, east; N, north; NE, northeast; NW, northwest; S, southern; SE, southeast; SW, southwest. **c**,**d**, ENMs of the 70 top aGWA SNPs (after linkage disequilibrium pruning), trained with climate averages from 1960 to 1990 (**c**) and then used to forecast the gain or loss of alleles in 2070 under free migration (**d**). **e**, Discrepancy of alleles that can be gained by 2070 between the geographically constrained (principal component analysis control) model and the free migration model.

local adaptation. However, in cases of strong population structure, the mean genome-wide F_{ST} is high[32], complicating outlier detection (Supplementary Fig. 4). One can recover relevant variants that are deeply divergent across populations and therefore invisible to conventional GWA by studying the association between the ancestry of each SNP and an adaptive phenotypic trait. Using ChromoPainter[33], 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 northwest and southeast 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' GWA (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) (Supplementary Table 4). The most prominent peak was located on chromosome 5 and explained over 20% of the variance in drought survival (Supplementary Table 4). There was no overlap in the top SNPs between GWA and aGWA because they search for different association signals. Our aGWA resembles other admixture mapping techniques[34] 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 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 the genome-wide SNPs, they are closely related in the 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 (Figs. 2d and 3b) and highlights the relevance of populations at the latitudinal extremes of the species range as a possible genetic reservoir for future climate change adaptation[12].

Drought survival is a resilience trait independent of phenology. Drought adaptation can be accomplished by diverse mechanisms, with cross-stress resistance being pervasive[35]. 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 efficiency[36]. 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 conditions[39] (Pearson's 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's correlation, $r=0.51$, $P=1\times10^{-11}$; 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 and Supplementary Table 6). Gene enrichment analysis revealed a weak signal for membrane transport (see Supplementary Methods 3.7). Adjustment of the osmotic balance through cell membrane transport is a drought avoidance mechanism[42] that might also confer cross-tolerance to other abiotic stresses[43]. 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 responses to severe abiotic stresses should have an evolutionary advantage in the 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, ENMs of species distributions, which have been used to predict future changes in species' ranges[2,3], do not usually include information on within-species diversity that can lead to adaptation from standing variation[44–46]. This could, in turn, lead to over-estimates of extinction rates[47–49]. By fitting ENMs of current climate with SNP data, using a similar rationale as for the 'climate GWA' of ref. [7], 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; that is, either the high or the low survival one, is more likely, given a set of environmental variables (all ENM fivefold cross-validation accuracy > 92%; Supplementary Tables 3 and 4 and Supplementary Figs. 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 North African range are predicted to harbour more drought-survival alleles than those located in between (Fig. 3c and Supplementary Figs. 15 and 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 (representative concentration pathway 8.5, Intergovernmental Panel on Climate Change, www.ipcc.ch; ref. [17]). While it was expected that populations in the Mediterranean Basin need to become more drought resistant[11], our predictions anticipate a greater increase in the total number of drought-survival alleles for Central Europe (Fig. 3 and Supplementary Figs. 14 and 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 conditions[50]. 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 currently has 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 models over the constrained ones 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 twenty-first century.

Ultimately, for a population to persist, not only must drought-survival alleles be present locally, but they also need to increase in frequency[51]. 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 km area): Madrid (Spain), Tübingen (Germany) and Malmö (Sweden), which are near the southern edge, centre and northern edge of the Eurasian and North African range, respectively. Based on ENM predictions, we calculated allele frequency changes from the present to 2070. Frequencies are predicted to increase significantly only in the Tübingen population (Student's t-test, $P<10^{-16}$; Supplementary Table 11), 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 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-generation 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 favourable alleles are more abundant towards 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 change[51], biotic interactions, phenotypic plasticity or novel adaptive mutations, our spatially explicit analyses emphasize the potential of adaptive evolution from standing variation to mitigate the detrimental effects of climate change.

Methods

Study populations. Some 211 natural inbred lines from the 1001 Genomes project[16] were grown in a terminal drought experiment and 762 lines were analysed 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 a >95% calling rate.This resulted in keeping ~4 million SNPs.

Genetic structure. To understand the genetic structure of *A. thaliana*, we ran the software ADMIXTURE version 1.2 (ref. [21]) on the 762 samples, assuming 2–20 groups and using a fivefold cross-validation procedure. The number of groups with the smallest cross-validation error was 11 (Fig. 2, Supplementary Table 1, 2 and Supplementary Fig. 8). We computed a genomic principal component analysis using PLINK version 1.9 (ref. [53]). The three first principal component axes explained 33.5% of the genomic variance (see Supplementary Methods 3).

We used genomes with a probability of assignment to one of the 11 ADMIXTURE groups of >0.9 to run Multiple Sequential Markovian Coalescent version 3 (ref. [22]). This was done in quartets of genomes (that is, four genomes for the within-population coalescent mode and two genomes of each of two populations for the cross-coalescent mode; Fig. 2 and Supplementary Fig. 5). Using the 11 genetic groups as population lineages, we ran Treemix assuming zero to five migration edges[24] (Fig. 2 and 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 for two weeks. Thereafter, watering only occurred every three weeks, which dramatically reduced the soil water content (Fig. 1 and Supplementary Methods 1.2). Top-view photographs of the potting trays were taken at 20 time points 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-leaf pixels from the brown soil background to monitor plant area over time (Supplementary Video). From the day with the largest rosette areas until the end of the experiment, we modelled the decay of the green area (that is, the number of pixels) using a polynomial generalized linear mixed model with Poisson link as described in the MCMCglmm R package version 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 the index of survival (Supplementary Methods 2).

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 the EMMAX software[25] 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 BSLMM with GEMMA software[28]. EMMAX fits a model as: $Y = \mu + X_i\beta + Zu + \varepsilon$, where Y is the vector of trait values, μ is the mean value, X is the alternative allele dosage at SNP i and β is 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 of GEMMA fits a multilocus model such as: $Y = \mu + X\beta + \varepsilon$, where all SNPs are fitted at once but there is a strong prior distribution of the β coefficients. These follow a mixture of two distributions—one that expects many small effects and another that generates few strong effects. Because all SNPs are included in the model, the population structure is implicitly accounted for.

To determine whether the top SNPs identified in the GWA might have been subject to polygenic adaptation, we used the method from ref. [14]. 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 version 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, $Y = \mu + X_i\beta + \varepsilon$, and reported the positions in the genome with the least mean square error (that is, the 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 the genome-wide F_{ST} among the ADMIXTURE-defined groups, as well as Tajima's D using PLINK version 1.9 (ref. [53]) and the likelihood of a selective sweep using 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 and Supplementary Table 3, rank columns).

ENMs. We used classification and regression random forest models implemented in the *randomForest* R version 1.4 package to build ENMs using the available climatic databases at www.worldclim.com (refs [17,19]; bioclimatic variables at 2.5 arc-minutes resolution) and the geographic locations of GWA-identified alleles. To evaluate each model's predictive ability for each allele, we used a fivefold cross-validation procedure in which four-fifths of the data were used to train the model and one-fifth 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 and 4).The fitted random forest model was used to generate potential geographic distributions of survival-associated alleles, which, when overlapped, provided a geographic map of the density of survival alleles. Using existing predictions of the same 19 bioclimatic variables in 2050 and 2070 under both low (representative concentration pathway 2.6) and high (representative concentration pathway 8.5) CO_2 accumulation scenarios, we re-predicted the distribution of alleles in 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) an ENM including the latitude and longitude variables in the random forest models and (2) an ENM including the three first principal component analysis (PCA) axes geographically modelled with present day climate (see below). By repeating the predictions with future climate data, but keeping the latitude, and longitude or principal component axes constant, some alleles would not be predicted in areas where the appropriate environment exists, but which are outside the current geographic distribution (1) or current local genomic background (2) (see Supplementary Methods 4 and Supplementary Figs. 13–16).

Apart from the potential distribution of putatively adaptive alleles, we also modelled the geographic distribution of continuous traits; namely, the aforementioned principal component analysis axes of population structure or the index of survival under drought itself. In these 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 the predicted and true values (see Supplementary Methods 4).

To complement observations of the presence and absence of alleles from the 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 to be 300,000 plants, as inferred from the 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 the natural lines sampled in a given geographic area (for example, 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).

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. Code for the image analysis pipeline is available at https://doi.org/10.5281/zenodo.1039888. Code for the ancestry GWA is available at https://doi.org/10.5281/zenodo.1039882. Code for the Wright–Fisher population simulations is available at https://doi.org/10.5281/zenodo.1039886.

Data availability. Phenotypic datasets are available in the Supplementary Dataset. Processed genome matrices are available from the 1001 Genomes Data Center, https://1001genomes.org/data/GMI-MPI/releases/v3.1/. Raw reads are available in the Sequence Read Archive with the identifier SRP056687, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA273563.

Received: 20 April 2017; Accepted: 21 November 2017;
Published online: 18 December 2017

References

1. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
2. Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. Climate change threats to plant diversity in Europe. *Proc. Natl Acad. Sci. USA* **102**, 8245–8250 (2005).
3. Jezkova, T. & Wiens, J. J. Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. *Proc. R. Soc. B Biol. Sci.* **283**, 20162104 (2016).

4. Barrett, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
5. Hereford, J. A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173**, 579–588 (2009).
6. Turesson, G. The species and the variety as ecological units. *Hereditas* **3**, 100–113 (1922).
7. Hancock, A. M. et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83–86 (2011).
8. Fournier-Level, A. et al. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86–89 (2011).
9. Lasky, J. R. et al. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol. Ecol.* **21**, 5512–5529 (2012).
10. Siepielski, A. M. et al. Precipitation drives global variation in natural selection. *Science* **355**, 959–962 (2017).
11. Dai, A. Increasing drought under global warming in observations and models. *Nat. Clim. Change* **3**, 52–58 (2012).
12. Hampe, A. & Petit, R. J. Conserving biodiversity under climate change: the rear edge matters. *Ecol. Lett.* **8**, 461–467 (2005).
13. Lee-Yaw, J. A. et al. A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. *Ecol. Lett.* **19**, 710–722 (2016).
14. Berg, J. J. & Coop, G. A population genetic signal of polygenic adaptation. *PLoS Genet.* **10**, e1004412 (2014).
15. Dormann, C. F. et al. Correlation and process in species distribution models: bridging a dichotomy. *J. Biogeogr.* **39**, 2119–2131 (2012).
16. 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
17. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
18. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
19. Mojica, J. P. et al. Genetics of water use physiology in locally adapted *Arabidopsis thaliana*. *Plant Sci.* **251**, 12–22 (2016).
20. Ingram, J. & Bartels, D. The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 377–403 (1996).
21. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
22. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
23. Lee, C.-R. et al. On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat. Commun.* **8**, 14458 (2017).
24. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
25. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
26. Atwell, S. et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
27. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2011).
28. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
29. Hedrick, P. W. Genetic polymorphism in heterogeneous environments: the age of genomics. *Annu. Rev. Ecol. Evol. Syst.* **37**, 67–93 (2006).
30. Pavlidis, P., Živkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
31. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
32. Josephs, E. B., Stinchcombe, J. R. & Wright, S. I. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol.* **214**, 21–33 (2017).
33. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
34. Shriner, D., Adeyemo, A., Ramos, E., Chen, G. & Rotimi, C. N. Mapping of disease-associated variants in admixed populations. *Genome Biol.* **12**, 223 (2011).
35. Tardieu, F. Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. *J. Exp. Bot.* **63**, 25–31 (2012).
36. Ludlow, M. M. in *Structural and Functional Responses to Environmental Stress* (eds Kreeb, K. H, Richter, H. & Minckley, T. M.) 269–281 (SPB Academic, The Hague, 1989).
37. Kenney, A. M., McKay, J. K., Richards, J. H. & Juenger, T. E. Direct and indirect selection on flowering time, water-use efficiency (WUE, $\delta^{13}C$), and WUE plasticity to drought in *Arabidopsis thaliana*. *Ecol. Evol.* **4**, 4505–4521 (2014).
38. Bac-Molenaar, J. A., Granier, C., Keurentjes, J. J. B. & Vreugdenhil, D. Genome-wide association mapping of time-dependent growth responses to moderate drought stress in *Arabidopsis*. *Plant Cell Environ.* **39**, 88–102 (2016).
39. Vasseur, F., Wang, G., Bresson, J., Schwab, R. & Weigel, D. Image-based methods for phenotyping growth dynamics and fitness in large plant populations. Preprint at https://www.biorxiv.org/content/early/2017/10/25/208512 (2017).
40. Juenger, T. E. et al. Identification and characterization of QTL underlying whole-plant physiology in *Arabidopsis thaliana*: $\delta^{13}C$, stomatal conductance and transpiration efficiency. *Plant Cell Environ.* **28**, 697–708 (2005).
41. McKay, J. K., Richards, J. H. & Mitchell-Olds, T. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* **12**, 1137–1151 (2003).
42. Jarzyniak, K. M. & Jasiński, M. Membrane transporters and drought resistance—a complex issue. *Front. Plant Sci.* **5**, 687 (2014).
43. Swindell, W. R. The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics* **174**, 1811–1824 (2006).
44. Pauls, S. U., Nowak, C., Bálint, M. & Pfenninger, M. The impact of global climate change on genetic diversity within populations and species. *Mol. Ecol.* **22**, 925–946 (2013).
45. Brown, J. L. 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.* **103**, 153–163 (2016).
46. Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1–16 (2015).
47. Catullo, R. A., Ferrier, S. & Hoffmann, A. A. Extending spatial modelling of climate change responses beyond the realized niche: estimating, and accommodating, physiological limits and adaptive evolution. *Glob. Ecol. Biogeogr.* **24**, 1192–1202 (2015).
48. Moritz, C. & Agudo, R. The future of species under climate change: resilience or decline? *Science* **341**, 504–508 (2013).
49. Hoffmann, A. A. & Sgrò, C. M. Climate change and evolutionary adaptation. *Nature* **470**, 479–485 (2011).
50. Aitken, S. N. & Whitlock, M. C. Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).
51. Fournier-Level, A. et al. Predicting the evolutionary dynamics of seasonal adaptation to novel climates in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **113**, E2812–E2821 (2016).
52. Roux, F., Giancola, S., Durand, S. & Reboud, X. Building of an experimental cline with *Arabidopsis thaliana* to estimate herbicide fitness cost. *Genetics* **173**, 1023–1031 (2006).
53. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

Acknowledgements

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

Author contributions

M.E.-A. conceived and designed the project. G.W. and F.V. helped with and advised on image phenotyping and F.V. provided additional phenotypes. M.E.-A. and W.D. performed chromosome painter analyses. M.E.-A. performed the drought experiment, processed the image data, and designed and carried out the statistical analyses. D.W. and H.A.B. advised and oversaw the project. M.E.-A. wrote the first draft and, together with H.A.B. and D.W., wrote the final manuscript with input from all authors.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-017-0423-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Detlef Weigel

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

We used a subset of 211 natural lines of the 1001 Genomes Project selected based on genome quality and geographic location. Details can be found in Supplementary Methods 1.1

2. Data exclusions

Describe any data exclusions.

See Supplementary Methods 1.1

3. Replication

Describe whether the experimental findings were reliably reproduced.

A number of 8 replicates per natural line were used in the greenhouse experiment and any inference was done based on the mean trends. To corroborate our findings, we replicated our experiment in a field experiment setup with the same natural lines and found a good correspondence between both experiments. See Supplementary Methods 1.3.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

We used a randomized blocked design to account for heterogeneity inside of the greenhouse facility and variation in the pot position within trays.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Data collection was done automatically using image analysis and thus all measurements were taken using the same standardized framework.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	All customized software used has been put available in Github with associated DOIs. See Data Availability section.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Seeds of all natural lines are available through the Arabidopsis Biological Resource Center with ID: CS78942

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	NA

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	NA
b. Describe the method of cell line authentication used.	NA
c. Report whether the cell lines were tested for mycoplasma contamination.	NA
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	NA

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	NA

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	NA