Multiple Sources of Introduction of North American 

*Arabidopsis thaliana* From Across Eurasia

Gautam Shirsekar¹, Jane Devos¹, Sergio M. Latorre¹†, Andreas Blaha¹, Maique Queiroz Dias††, Alba González Hernando¹, Derek S. Lundberg¹, Hernán A. Burbano¹², Charles B. Fenster³, and Detlef Weigel¹*

¹Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
²Centre for Life’s Origin and Evolution, University College London, London WC1E 6BT, UK
³Oak Lake Field Station, South Dakota State University, Brookings, SD 57007, USA

*To whom correspondence should be addressed. Email: weigel@weigelworld.org

¹Current addresses: Centre for Life’s Origin and Evolution, University College London, London WC1E 6BT, UK (S.M.L.); Federal University of Viçosa, 36570-900 Viçosa - MG, Brazil (M.Q.D.)

**Keywords:** *Arabidopsis thaliana*, population genetics, admixture, non-native species, migration

**Abstract**

Large-scale movement of organisms across their habitable range, or migration, is an important evolutionary process that can shape genetic diversity and influence the adaptive spread of alleles. While human migrations have been studied in great detail with modern and ancient genomes, recent anthropogenic influence on reducing the biogeographical constraints on the migration of non-native species has presented opportunities in several study systems to ask the questions about how repeated introductions shape genetic diversity in the introduced range. We present an extensive overview of population structure of North American *Arabidopsis thaliana* by studying a set of 500 whole-genome sequenced and over 2,800 RAD-seq genotyped individuals in the context of global diversity represented by Afro-Eurasian genomes. We use methods based on haplotype and rare-allele sharing as well as phylogenetic modeling to identify likely sources of introductions of extant N. American *A. thaliana* from the native range in Africa and Eurasia. We find evidence of admixture among the introduced lineages having increased haplotype diversity and reduced mutational load. We also detect signals of selection in immune-system related genes that may impart qualitative disease resistance to pathogens of bacterial and oomycete origin. We conclude that multiple introductions to a non-native range can rapidly enhance the adaptive potential of a colonizing species by increasing haplotypic diversity through admixture. Our results lay the foundation for further investigations into the functional significance of admixture.
Introduction

*Arabidopsis thaliana* is predominantly a human commensal that is native to Africa and Eurasia. Its demographic history is filled with episodes of range expansions, bottlenecks, migrations and admixture. Current models of *A. thaliana*’s population history highlight the recurrent theme of lineage migration and admixture with locally adapted genotypes in the native range of the species (Durvasula et al. 2017; Lee et al. 2017; Zou et al. 2017; Fulgione et al. 2018; Fulgione and Hancock 2018; Hsu et al. 2019). A further opportunity to learn about the impact of demographic processes and selection in *A. thaliana* arises from its relatively recent colonization of N. America.

When a species is introduced outside its native range, where its long-term eco-evolutionary history has been established, different factors determine how well the introduced population adapts to the new environment. These factors include, but are not limited to history of introduction, founder effects, and strength of natural selection (Colautti and Lau 2015; Estoup et al. 2016). In the post-Columbian era, *A. thaliana* has benefited from mostly unidirectional cross-continental species movement facilitated by human migrations to North America (La Sorte et al. 2007; Winter et al. 2010). Thus, the N. American *A. thaliana* metapopulation presents a unique natural experiment for studying the role of history in explaining extant diversity and understanding how the colonizers have thrived despite population bottlenecks and seemingly low genetic diversity, also known as genetic paradox of invasion (Allendorf and Lundquist 2003). Important questions that can be addressed using this study system are: How much of the native diversity was introduced to N. America? How much new diversity has been generated in situ through mixing of lineages that originated from distant parts in the native range? How much of the observed diversity is due to selection?

*A. thaliana* has become established across much of N. America. Coarse-scale population structure analysis of N. American individuals with 149 single nucleotide polymorphism (SNP) markers has revealed the presence of a dominant lineage “Haplogroup1” (Hpg1) (Platt et al. 2010). Patterns of mutation accumulation in the genomes of pure Hpg1 individuals have supported an arrival in N. America about 400 years ago, soon after Europeans started to arrive en masse on the continent. A parsimonious explanation of the ubiquitous nature of this lineage could be that it was the earliest to be introduced to N. America (Exposito-Alonso, Becker, et al. 2018). So far, little consideration has been given to the supposedly subsequent arrival of other lineages, their origins in the native range, their fate as migrations continued during the past centuries, and how the genomes of the current N. American population have been shaped by processes such as admixture and adaptation.

We present the fine-scale population structure of the N. American *A. thaliana* population as viewed through the lens of range-wide genetic diversity of the species. Using genomes of *A. thaliana* individuals collected from the Midwest, the Eastern Seaboard and the North-East of the current United States of America, we infer possible sources of ancestry based on haplotype-sharing, phylogenetic tree-based modeling and rare allele
sharing with the worldwide dataset. We also describe how admixture in this predominantly selfing species is generating new haplotype diversity and how admixture affects the fate of deleterious mutations and allows selection on immunity related loci. The work presented here shows that increased global connectivity through the past two centuries has made species invasions from across the species range common and could have accelerated invasion of N. American habitat by avoiding the genetic paradox of invasion. Further, our work highlights that processes such as admixture and selection that determine success of lineages in the native range play a similarly important role in the non-native range.

Results

An overview of population structure and genetic variation from RAD-seq

We collected *A. thaliana* samples across an area of about 1,200 by 900 km in the Eastern United States of America during the spring seasons (mid-March to early June) of 2014, 2015 and 2016 (Fig. 1; Table S1a,b). We genotyped these samples using a RAD-seq implementation of reduced representation sequencing (Miller et al. 2007). After filtering for sequencing output and quality, we retained 2,861 individuals, which shared 4,907 polymorphic SNPs. In order to compare the population structure and genetic diversity in our North American to the global Afro-Eurasian collection (AEA) we used data from the 1001 Genomes project (1001 Genomes Consortium 2016) in addition to whole genome sequences from 13 Irish (this work), 10 African (Durvasula et al. 2017) and 5 Yangtze River basin accessions (Zou et al. 2017). From these AEA individuals, information on the 4,907 polymorphic SNPs found in our N. American individuals (average depth ~36X) were extracted and merged with the N. American dataset for further analysis. Although pairwise similarity using “identity-by-state” (IBS) and “identity-by-descent” (IBD) across the genome is greater in N. American than in AEA individuals, genetic variation relative to AEA individuals could nevertheless be observed in N. American individuals with principal component analysis (PCA) (Fig. S1A). N. American individuals in our collection were genetically much more diverse than the N. American individuals previously sequenced as part of the 1001 Genomes Project (Fig. S1B).

Diversity of N. American haplogroups

We first used RAD-seq to rapidly genotype thousands of individuals, but because of its inherent biases (low density of markers, strand-bias, underestimation of genetic diversity), these data are not well suited for fine-scale, quantitative population genomic analyses (Arnold et al. 2013; Cariou et al. 2016; Lowry et al. 2017). We therefore selected and sequenced a subset of distantly related individuals with whole-genome sequencing (WGS) approach, at an average of ~ 8x coverage (0.63-43.91x, median= 5.81x). A PCA of 500 N. American individuals, including a subset of previously analyzed herbarium individuals (Exposito-Alonso, Becker, et al. 2018) (Fig. S1C), resulted in an arrangement in which most individuals were found along distinct clines. We
decided to explore this population structure in detail using different complementary population genetic methods.

Finer-scale population structure can be revealed by explicitly modeling the effects of linkage disequilibrium (LD) and clustering individuals based on their shared ancestry that emerges after accounting for LD (Busby et al. 2015; Leslie et al. 2015; Montinaro et al. 2015). Therefore, we hierarchically partitioned the N. American individuals into 58 clusters (from here on called groups) using a co-ancestry matrix derived using CHROMOPAINTER v2 and MCMC-based clustering in fineSTRUCTURE (Lawson et al. 2012) (Fig. S3). Haplogroup1 (Hpg1) is the most frequently observed group across the sampled populations (Fig. 2), consistent with previous observations (Platt et al. 2010; Exposito-Alonso, Becker, et al. 2018). OHML (Ohio) and NJSC (New Jersey) had the highest within-population haplotype diversity, with 11 and 12 groups. Several groups, such as OhioNewJersey2, IndianaNewJersey1 and NewJerMich1, were found in populations from geographically distinct regions (Fig. 2).

Figure 1. Locations and number of sampled individuals
Abbreviations of the locations sampled are shown along with the number of RAD-sequenced samples (in black) and the number of whole-genome sequenced (WGS) samples (in red). Left inset: bar plot of total number of samples sequenced. Right inset: sampling area in the context of N. America.

We further analyzed the genetic relationships among these groups using several complementary approaches. Treemix (Pickrell and Pritchard 2012), without considering migration edges, reconstructed relationships among the groups (Fig. S4A), similar to the topology inferred by fineSTRUCTURE clustering (Fig. S3B). Notably, residuals from the fitted model with high positive values indicated that the fit could be improved by including admixture edges among the groups (Fig. S4B). High positive residual values between Hpg1, which is the omni-present group with high frequency and other groups, suggested possible gene flow between them.

Stochastic changes in allele frequency, as a result of the neutral process of drift, hold information about shared ancestry. We therefore estimated values for the $f_3$-outgroup statistic (Raghavan et al. 2014) to understand the shared drift among groups relative to an outgroup (individuals of relict ancestry). Relicts comprise highly diverged individuals from ice age refugia (Lee et al. 2017) and therefore were chosen as an outgroup. Indeed, some of the N. American groups (OhioMich1, SouthIndiana4 and Ohio7) along with Hpg1 shared excess drift with other groups (Fig. S5). As the $f_3$-outgroup test identifies the closest relative population and does not itself point to the admixture, we applied $f_s$ statistic to explore the possibility of admixture among these groups.
We calculated values for the $f_3$ statistic in all trios (groupA, groupB: groupTest) of N. American groups to detect whether groupTest was admixed between groupA and groupB. There were several groupTest examples with negative $f_3$ scores and $Z$-scores below -3 in several trios (Fig. S6A, Table S2). In several cases, Hpg1 emerged as a putative source (as either groupA or groupB) (Table S2). To investigate this in more detail, we calculated the shared drift of Hpg1 relative to the other N. American groups. We found more groups with a gradient of shared drift with Hpg1; Massachusetts1 was one of the groups with least shared drift (Fig. S6B). Therefore, we calculated the ABBA-BABA statistic (D-statistic) in the form of (Massachusetts1, Test : Haplogroup1, Relicts) to learn the extent of gene flow between Hpg1 and other N. American groups (Fig. S6C). Many groups showed significantly more ABBA sites ($Z$-score < -3) than BABA sites, confirming the contribution of Hpg1 ancestry to the genetic makeup of these groups.

**Figure 2. Identification of North American groups based on haplotype sharing and their distribution in different populations.**

Collapsed fineSTRUCTURE tree generated by merging North American individuals into groups (herbarium individuals are denoted by JKxxx) based on their co-ancestry (derived using CHROMOPAINTERv2, co-ancestry matrix in FigS3). Last row of numbers represents the total count of groups present in the population, and last column of numbers represents the number of populations in which a specific group is present (here a group present in a single population is not counted, count=1).
Contribution of distinct sources of ancestry to N. American diversity

North American groups vary in terms of their drift relative to the earliest arrival, Hpg1, which suggests that there have been multiple introductions of *A. thaliana* to N. America. It is also unclear whether the observed haplogroups already existed in Eurasia, or whether they only formed by intercrossing in N. America. We therefore wanted to learn whether N. American extant haplogroups include ancestry from different geographic regions in Eurasia. We first excluded lineages that showed evidence of recent admixture (*groups* with significantly negative $f_3$-scores), and we then applied statistical procedures based on shared haplotype chunks (fineSTRUCTURE), shared drift (*f$_3$*-outgroup, D-statistic and qpWave) and enrichment of rare alleles with respect to the AEA haplotype diversity to identify sources of ancestry in Eurasia based on whole genome sequences from AEA individuals (n=928) (1001 Genomes Consortium 2016). We traversed the genomes of N. American individuals to assign local ancestry along each chromosome. To this end, we performed haplotype based inference in three steps: (i) Paint each AEA individual against the others (excluding itself) with CHROMOPAINTER v2, (ii) Based on haplotype sharing, cluster individuals into *sub-clusters* using fineSTRUCTURE. These *sub-clusters* were then grouped into *clusters*, and *clusters* were further grouped into *regions* (Fig. S7; details of these hierarchical partitions for each AEA individual are given in Table S3). (iii) We chose 15 representative individuals per AEA *region* and estimated an ancestry profile for individual N. American recipients.

Fig. 3A shows these inferred ancestry profiles for the N. American individuals. It can be seen that although the majority of groups are enriched for Upper/EastFranceBritishIsles ancestry, other British Isles *regions* (BritishIsles1 and BritishIsles2) also feature significantly across several groups (Fig. 3C,D and Fig. S8). Apart from these, some N. American *groups* such as MichiganManistee1, OhioOSU and SouthIndiana1 had substantially higher contributions from East European *regions* such as RussiaAsia, CentralEurope/Baltic and Italy/BalkanPeninsula. NorthGermany and SouthGermany *regions* have contributed to the ancestry of OhioMich1, RhodeIsland1 and Mid-Western1 *groups* (Fig. 3C,D and Fig.S8). These results also highlight how geo-genetically distant AEA ancestries could be found within the same population (INRC) or within the same regions (Midwest) in N. America.

We explored these haplotype sharing patterns further by measuring shared drift between a test N. American *group* and 158 *sub-clusters* of AEA individuals using $f_3$ - outgroup statistic of the form *test, sub-cluster; relictsFs12-3* (We chose relictsFs12-3 as an outgroup as it is a highly diverged *sub-cluster* comprising relict population individuals) and building a maximum likelihood (ML) tree by fitting Treemix(Pickrell and Pritchard 2012) model without any admixture edges. At a coarser scale, the results agree with the haplotype-based inferences. Shared allelic drift measured with $f_3$ -outgroup statistic and captured in the ML tree showed that the current N. American *groups* are related to the AEA *sub-clusters* that belonged to either western, central or eastern Europe (Fig. S9, S10). We also observed these patterns of relatedness qualitatively in a PCA plot where we projected N.American individuals into PC space occupied.
by AEA individuals (Fig. S11 A). Even finer details became apparent with uniform manifold approximation and projection (UMAP) embeddings (McInnes et al. 2018) (Fig. S11B) derived from the first 50 PC components of all the individuals (without projection).

![Diagram of chromosome painting of N. American groups with Afro-Eurasian (AEA) regions as donors. A. Copying profile of the N. American individuals inferred with CHROMOPAINTERv2 using a reference panel of individual haplotypes belonging to different AEA regions, each dot represents an individual (cumulative genomic segment length copied is in centiMorgans). B. Geographic locations of the AEA individuals used in the reference panel (colored by region). C,D. Two major contributions from AEA regions to N. American groups found on Eastern Seaboard-Northeast and Midwest. Arrows point from the mean geographic position of the AEA regions to that of the N. American groups (colors of the contributing region are the same as in panel A. Credit for original design of A. thaliana rosettes: Frédéric Bouché)](https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab268/6367783)

The coarse patterns of shared ancestry emerging from $f_3$-outgroup statistic, PCA projection, and UMAP embeddings were tested in a more systematic way by evaluating the “treeness” of different topological configurations. We first calculated $D$-statistic (Green et al. 2010) for all the N. American non-admixed groups.
(X) in the form of NorthGermany, RelictsFs12; X, *A. lyrata* (*Arabidopsis lyrata* as a closest relative to *A. thaliana* (Schmickl et al. 2010) was chosen as an outgroup). NorthGermany region was chosen because of its central geographic location among the possible sources of N. American *A. thaliana*. We then calculated D-statistics for the N. American groups by replacing NorthGermany with 4 regions: BritishIsles1, Upper/EastFranceBritishIsles, CentralEurope/Baltic and RussiaAsia. We then plotted D-statistics with replacement regions to D-statistics obtained using NorthGermany separately to understand region specific drift (Ebenesersdóttir et al. 2018). These biplots (Fig. 4C) clearly differentiate western and eastern European sources of ancestry in N. American *A. thaliana*. OhioOSU, Ohio2, SouthIndiana1 and MichiganManistee1 clearly showed the relative eastern European ancestry component. The analysis also revealed that Col-0, the reference genome accession for *A. thaliana* research, shares significant ancestry with individuals from NorthGermany, confirming the origin of Col-0 in or near Germany (Rédei 1992). Despite constrained “tree” topologies explored, since we used sub-clusters to estimate the D-statistic, it allowed us to capture variation in the shared drift (horizontal and vertical bars in Fig. 4) experienced by a target N. American group with a given AEA region.

**Fig. 4. Multiple sources of origin of N. American haplogroups.** Biplot of mean D-statistics of N. American haplogroups (X) with sub-clusters comprising NorthGermany region (NorthGermany, RelictsFs12; X, *A. lyrata*), against mean D-statistics of sub-clusters comprising different regions in an eastward direction (testRegion, RelictsFs12; X, *A. lyrata*). Vertical and horizontal bars represent the spread of D-statistics from member sub-clusters of each region.

We extended this analysis using qpWave (Reich et al. 2012) to test whether any two N. American groups would be symmetrically related to a set of outgroups (AEA regions). Specifically, we tested whether a set of f<sub>r</sub>-statistics comprising two N. American groups across a set of layer1 outgroups (AzerbaijanGeorgia, Barcelona, NorthSweden, NorthWestEngland, Relicts Fs13, SouthTyrol, WestScania and West/NorthCentralFrance) makes a matrix of rank 0 (same wave of ancestry) (Table S4). We then tested whether addition of an extra outgroup region (consisting of putative sources of ancestry) to the layer 1 outgroup set affected the symmetry of shared ancestry. If the two test N. American groups are differentially related to
the extra outgroup region, then it would increase the rank of the original matrix of \( f^*_r \) statistics (rejection of rank 0), indicating distinct streams of the ancestry among the test groups. We added an extra outgroup from additional regions of BritishIsles2, ItalyBalkanPeninsula, NorthGermany, RussiaAsia and Upper/EastFranceBritishIsles one-by-one. Adding these putative source regions affected the symmetrical relationships observed with our original outgroup set. Except in the case of SouthIndiana4 and Ohio7, all the N. American group combinations showed asymmetric relationships (rejection of rank 0) with these extra outgroups (Table S4). These results validated the findings from qualitative observations made with PCA projection, and UMAP embeddings. It further confirmed results obtained from \( f_3 \)-outgroup statistic, ML tree, and \( D \)-statistics analysis, that the N. American \( A. \) thaliana groups have ancestral components from western Europe (mainly British Isles), central Europe and eastern Europe.

More subtle patterns of ancestry can be inferred by finding rare variants (Schiffels et al. 2016) from AEA that have risen to higher frequency in N. American individuals. Because we had moderate- to high-coverage whole genomes of the AEA and N. American individuals, we could use such rare variants to independently ascertain the results obtained from the haplotype-based ancestry inference and shared ancestry based inference, mostly on moderate to high frequency alleles. We identified variants from AEA individuals with frequency of 1% or lower and tracked their enrichment in the N. American groups. We found that N. American groups have accumulated rare alleles from different AEA sub-clusters (Fig. S12). Whereas several N. American groups have inherited rare alleles from British Isles sub-clusters, groups RhodeIsland1, MichiganManistee1, OhioOSU, SouthIndian1 and OhioMich1 have accumulated rare alleles from central/eastern European sub-clusters, while Hpg1 has accumulated a significant number of rare alleles from sub-clusters from the Upper/EastFrance/BritishIsles region. Taken together, this analysis confirmed that N. America was colonized by \( A. \) thaliana in multiple waves with distinct sources of ancestry.

**Environmental conditions at source and success of colonizing lineages**

As we had inferred the shared ancestry of the colonizing lineages with different complementary methods, we hypothesized that besides human-assisted migration, environmental similarity between putative source sub-clusters and colonizing lineages contributed to successful colonization of the lineages. To test this hypothesis, we fit a regression model to predict shared ancestry with AEA sub-clusters (as measured by \( f_3 \)-outgroup statistics of the form test N. American group, AEA sub-cluster: RelictsFs12_3 (outgroup), value of the statistic is proportional to the shared ancestry between the populations relative to the outgroup), using linear combinations of four environmental variables: average temperature (tavg), precipitation (prec), solar radiation (srad) and water vapor pressure (vapr). \( A. \) thaliana shows significant local adaptation to climate (Exposito-Alonso, Vasseur, et al. 2018; Fulgione and Hancock 2018), thus the choice of these four variables should provide a general climatic niche. We used Bayesian multilevel modeling (bMLM) framework (Gelman 2006)
to understand each N. American group’s environmental association with its putative source AEA sub-clusters without ignoring the environmental association to the entire cohort of N. American groups.

Population-scale coefficients for the environmental variables precipitation (mm) and water vapor pressure (kPa) revealed that environmental dissimilarity calculated by Euclidean distance between each N. American group and AEA sub-cluster is negatively correlated with the $f_3$-outgroup statistics (Table 1). Although average temperature dissimilarity is slightly negatively correlated with $f_3$-outgroup statistics, the compatibility interval with the model is large, with slightly positive correlation in posterior distribution. Upon closer examination of the coefficients estimated for individual N. American groups, it can be seen that precipitation and water vapor pressure dissimilarity is negatively correlated with the $f_3$-outgroup statistic for all groups but MichiganManistee1 (Fig. S13). Overall the general trend of negative correlation of the linear combination of the dissimilarity of the variables (average temperature, precipitation, solar radiation, and vapor pressure) to the $f_3$-outgroup statistic can be captured with the individual estimates sampled from the posterior distribution (Fig. S14).

The negative correlation between environmental dissimilarity and shared ancestry led us to hypothesize that in reduced dimensional space of environmental variables (average temperature, precipitation, and vapor pressure), N. American groups should be closer to their source AEA sub-clusters. To test this, we performed UMAP on the standardized values for environmental variables for N. American groups and AEA clusters together, followed by hierarchical clustering on the reduced environmental space (see details in the Supplementary Materials). We observed that the N. American groups and their putative source clusters, as inferred by population genomic approaches (specifically sub-clusters from Upper/EastFranceBritishIsles, NorthGermany, SouthGermany, BritishIsles1, BritishIsles2 and CentralEurope/Baltic regions) occupied similar space in the UMAP embeddings (Fig. S16) and were in the same major clades (Fig. S17), thus confirming that overall environmental similarity between source populations and N. America might be an important contributor to the success of colonization.

### Table 1. Posterior summary of the regression coefficients for environmental variables

| Parameter | mean  | sd    | hdi.3% | hdi.97% | $\hat{R}$ |
|-----------|-------|-------|--------|---------|-----------|
| $T_{avg}$ (°C) | 0.1390 | 0.0920 | -0.0280 | 0.3190 | 1.0000 |
| Precipitation (mm) | -0.0640 | 0.0960 | -0.2420 | 0.1170 | 1.0000 |
| Solar radiation (kJ m$^{-2}$ day$^{-1}$) | -0.1800 | 0.0910 | -0.3530 | -0.0080 | 1.0000 |
| Water vapor pressure (kPa) | 0.0010 | 0.0910 | -0.1800 | 0.1600 | 1.0000 |
| $\sigma$ | 0.3950 | 0.0110 | 0.3730 | 0.4160 | 1.0000 |
Effect of admixture on deleterious mutations

Evolutionary theory predicts that during range expansions and new colonizations deleterious mutations accumulate gradually and steadily, resulting in increased mutational load that can be reduced again by outcrossing (Peischl et al. 2013). We hypothesized that the levels of mutational load in N. American individuals would be related to rates of historic outcrossing as inferred from admixture. To test this, we chose individuals from populations INRC, MISJ, NJSC, and OHPR because of: 1. Presence of earliest colonizing lineage Hpg1, and 2. two-way admixture events between Hpg1 and non-Hpg1 groups (Fig. S6 and Table S2) in these populations. First, we inferred admixture proportion in the admixed groups by averaging local ancestry inference (LAI) obtained with Loter (Dias-Alves et al. 2018). We divided derived mutations observed in N. American groups into three different categories according to SIFT4G (Kumar et al. 2009) predictions (see details in the methods) and calculated the frequency of these mutations in each group including the source groups (Hpg1 and non-Hpg1s). Derived alleles with frequency of 1 were considered as fixed, and based on the count we calculated fixed-to-total derived alleles ratios (Φ) for each group. Considering admixture tracts of the diverged source lineages will result in slightly increased effective population sizes of the admixed groups, admixture is expected to reduce estimates of Φ regardless of mutational categories. As genome-wide synonymous derived variation is expected to be effectively neutral with respect to fitness in small populations (Walsh and Lynch 2018), we considered Φ_syn as a baseline for the reduction in Φ due to admixture and divergence, and therefore scaled Φ_non-syn-deleterious and Φ_non-syn-tolerated with Φ_syn (details in the methods). The scaled Φ_non-syn-deleterious (Φ) is significantly lower (Welch’s t-test p-value=6.35e-12) than for non-synonymous tolerated mutations (Fig. 5, Fig. S18C and Table S17). Further, admixture had a strong effect on reducing Φ_non-syn-deleterious, whereas Φ_non-synonymous-tolerated was not affected by it (Fig. S18B). The change from, Φ_non-syn-tolerated to Φ_non-syn-deleterious was significantly different between admixed and source groups (Fig. S18D). Similar patterns were observed when Φ was estimated using the total genome-wide Φ for scaling across mutational categories (Fig. S18E). This strongly suggests that admixture helps to eliminate derived mutations with potential deleterious impacts and it efficiently reduces non-synonymous mutational load.

![Fig. 5. Scaled fixed-to-total derived alleles ratios (Φ) for N. American groups across non-synonymous mutation categories.](image_url)

Scaled fixed-to-total derived alleles ratio (Φ) for each group in populations OHPR, INRC, MISJ and NJSC for non-synonymous mutation categories was calculated, scaled by the fixed-to-total derived alleles ratio for synonymous mutations. Points represent different groups in the focal populations and colors of the individual points represent its non-Hpg1 ancestry proportion, as indicated on the scale at the right. Fixed derived alleles are alleles with frequency = 1. p-value is from Welch’s t-test.
Ongoing selection at several immunity loci in N. America

Apart from admixture potentially reducing mutational load, it can also be a source of beneficial alleles. If such alleles are strongly selected, they will create signatures of a selective sweep (Smith and Haigh 1974; Stephan 2019; Moest et al. 2020). To look for such a scenario, we focused on large populations comprising several groups that apparently arose as a result of admixture between lineages that diverged before their introduction to N. America (Fig. S3). These populations were INRC (Indiana), NJSC (New Jersey), MISJ (Michigan), OHML and OHPR (both Ohio).

Methods that track the decay of haplotype homozygosity in a population (Vatsiou et al. 2016) can be used to detect such sweeps. We scanned whole genomes for signals of natural selection using haplotype homozygosity based tests \( iHS \) (integrated haplotype homozygosity score) (Voight et al. 2006) and \( nS_L \) (number of segregating sites-by-length) (Ferrer-Admetlla et al. 2014) for individual populations (Table S5-S9) and \( xp-EHH \) (cross population extended haplotype homozygosity) (Sabeti et al. 2007) for comparisons between population pairs. For individual populations, we focused on variants with \( |iHS| \) p-values for <0.001 and \( |nS_L| \) values >2 (Table S5-S9). GO-term analysis of the 82 genes tagged by these variants revealed an enrichment of genes in the categories ‘response to stress’ and ‘response to stimulus’ (Fisher’s exact test with Bonferroni correction, \( p \)-value <0.001 and FDR <0.05) are plotted with gene names or gene IDs.

Fig. 6. Genomewide haplotype-based selection statistics in five N. American populations A. Genome-wide \( p \)-values of \( |iHS| \) scores (based on empirical distribution). Dashed horizontal lines correspond to a \( p \)-value significance threshold of 0.001. Selection candidates, which also had \( |nS_L| \) scores of >2, from the enriched GO categories of response to stress and response to stimulus (Fisher’s exact test with Bonferroni correction, \( p \)-value <0.001 and FDR <0.05) are plotted with gene names or gene IDs.
0.001 and FDR < 0.05) (Fig. 6A, and Table S10) We further confirmed this enrichment of GO-terms using a permutation method based approach implemented in Gowinda (Kofler and Schlötterer 2012) (Table S14). Consistent with the GO category enrichment, we noticed several NLR genes, a family that includes many known disease resistance genes (Van de Weyer et al. 2019). These included RPP13 and BAR1, which confer resistance to the oomycete Hyaloperonospora arabidopsidis (Har) (Bittner-Eddy et al. 2000) and bacteria of the genus Pseudomonas (Laflamme et al. 2020). We measured the frequency of alternative haplotypes around these loci to determine the nature of the selective sweep (Garud et al. 2015). The most frequent haplotype is designated as H₁ and the second most as H₂, from which a modified product of haplotype frequency (H₁₂) and the H₂/H₁ ratio are calculated. At RPP13 and BAR1, we observed relatively low values for these two metrics and the presence of the selected alleles on multiple backgrounds, which together suggests soft sweeps at these loci (Fig. S19A,B). On the other hand, a pronounced hard sweep was observed in and around another putatively selected NLR, BURNOUT1, in the population OHML (Fig. S19A,B), with the selected allele found on a single haplotype. Similar to the $|iHS|$ and $|nS_L|$ results, genes with high $xp$-EHH scores included several genes known to be involved in biotic and abiotic stress responses (Table S11).

In order to obtain direct evidence of whether the positively selected alleles are entering the population through admixture, we used data collected for presence of disease symptoms of downy mildew caused by Har (details in Supplementary Material) on the samples. We chose the MISJ population for this analysis because the rate of infection in this population was comparatively higher than others with many RAD-seq genotyped individuals (36 diseased and 115 healthy individuals based on visual observation at the time of collection, Table S15). We observed that the genetic differentiation measured by $F_{ST}$ was higher on chromosome 4 between groups of diseased and healthy individuals (Fig. S20A). MISJ has admixed individuals from Hpg1 and OhioMich1 source groups (Fig. 2, Fig. S6 and Table S2), therefore we calculated $F_{ST}$ of the two groups with Hpg1 separately. We found that the healthy individuals were significantly differentiated to Hpg1 in the genomic region of 9-12Mbp of chromosome 4 compared to the diseased individuals (Fig. S20B), whereas diseased individuals showed lower non-Hpg1 ancestry than the healthy individuals on chromosome 4 in general (Fig. S20C) and significantly lower non-Hpg1 local ancestry in the genomic region of 9-12Mb (Fig. S20D,E). This genomic region is characterized by the presence of a well-known Har disease resistance gene cluster of RPP4/5 (Noël et al. 1999; Holub 2001; van der Biezen et al. 2002). Thus, this result strongly suggests that RPP4/5 genomic region from non-Hpg1 group OhioMich1 is preferentially kept in the MISJ population as it confers disease resistance.

**Discussion**

How newly introduced, non-indigenous species adapt to new environments is a topic of long-standing interest in eco-evolutionary biology of invasive species (Baker, H. G. & Stebbins, G. L. 1965; Bock et al. 2015). There are two potential challenges facing invasive species: First, the niches in the new environment might be different
from the ones in the native range and/or already filled by other species. Second, introductions typically begin with few individuals and therefore potentially a narrow genetic basis. The initial lack of genetic diversity can be overcome by new mutations or through the generation of new genetic combinations, either by crosses among the introduced population or with close relatives that are present in and already adapted to the new environment.

We have used *A. thaliana* to address these questions.

*Arabidopsis thaliana* is native to Europe, Asia and Africa, where it is found mostly as a human commensal (Hoffmann 2002; 1001 Genomes Consortium 2016; Durvasula et al. 2017; Lee et al. 2017; Zou et al. 2017). The human-assisted expansion of this species to N. America presents an excellent system to study processes associated with colonization of a new environment because it occurred recently and because the genetic diversity in the native range is so well documented for *A. thaliana*. Previous work has laid the groundwork for our study, but was limited by a paucity of genetic markers (Platt et al. 2010) or a focus on a single dominant lineage (Exposito-Alonso, Becker, et al. 2018). We have investigated multiple individuals from several N. American populations at the whole-genome level, allowing us to describe fine-scale haplotype sharing within N. America and between N. America and individuals from the native range, either sequenced as a part of the 1001 Genomes project (1001 Genomes Consortium 2016) or subsequent efforts focused on Africa (Durvasula et al. 2017), China (Zou et al. 2017) and Ireland (this work).

**Multiple independent introductions**

The extant diversity among *A. thaliana* individuals in N. America can be traced back to multiple, almost certainly independent introductions of lineages of diverged ancestries from three distinct geographic regions of Western Europe (British Isles/Ireland, Upper and Eastern France), central Europe (Germany, Czechia, and Austria) and Eastern Europe (the Baltic region and Russia). We detected these introductions using methods based on haplotype sharing (Lawson et al. 2012), allele frequencies (Patterson et al. 2012), and rare-allele sharing (Schifflers et al. 2016; Flegontov et al. 2019), lending considerable confidence to our findings and illuminating the extant diversity from several different angles. Significantly, even though we confirm that North-Western Europe and specifically the British isles are a major source of multiple introductions, the predominant lineage Hpg1, which has been estimated to have been introduced ~400 years ago (Exposito-Alonso, Becker, et al. 2018), has more ancestry from Upper and Eastern France than from the British Isles. Its spread in N. America could be attributed to rapid expansion of French colonists from current Canada (then Acadia) along the Mississippi valley during the early period of post-Columbian colonization (Hamilton 1902).

Our approach of haplotype-based clustering of individuals at different hierarchical levels using fineSTRUCTURE (Lawson et al. 2012) has allowed us to pinpoint several Western European sources of N. American *A. thaliana*. While the sparse representation of individuals from Eastern Europe and Asia has limited our ability to more precisely identify the source of introductions from these regions, it is clear that Eastern Europe has contributed to extant N. American *A. thaliana* ancestry. Historical patterns of human migration
indicate that northern and western Europeans arrived in significant numbers from the 1840s to 1880s, followed by waves of southern and eastern Europeans from the 1880s to 1910s (Passel and Fix 1994), which are reflected in the genetic make-up of present-day humans in N. America (Bryc et al. 2015; Dai et al. 2020). In the regions where we collected *A. thaliana* in N. America, humans have more British, Irish, central and eastern European ancestry than western, southern and northern European ancestry (Bryc et al. 2015), consistent with the *A. thaliana* ancestry patterns. Thus, local anthropogenic introduction of *A. thaliana* can be proposed as a parsimonious explanation for the presence of diverged lineages in the regions of N. America that we sampled in our study.

**Wide-spread admixture**

Perhaps our most significant finding is how multiple introductions have led to present-day N. American *A. thaliana* being surprisingly genetically diverse, different from many other colonizing or invading species (Dlugosch and Parker 2008). This highlights how between-population variation in the native range has translated into within-population variation in N. America (Rius and Darling 2014). In organisms with low out-crossing rates such as *A. thaliana*, benefits of local adaptation in the native range hinder admixture from other populations, even in the face of inbreeding depression. It has been argued that during invasion of new territory, there is a temporary loss of local adaptation that not only lifts the maladaptive burden of admixture but even favors admixture (Verhoeven et al. 2011; Rius and Darling 2014). We indirectly observe this in AEA regions’ geographically restricted haplotype sharing patterns (Fig. S24C, quantified as Bray-Curtis distance in Fig. S25) whereas due to multiple introductions, N. American *A. thaliana* has a mixture of diverged ancestries (Fig. 3A) and among some individuals levels of increased compositional dissimilarity (Fig. S25) similar to that seen in individuals from AEA regions. Further, compared to some AEA regions, nucleotide diversity (pi) and total derived allele count are elevated in a few N. American populations (Fig. S26). Patterns that are similar to the ones we have reported here for *A. thaliana* have been suggested for other systems, albeit mostly based on limited genetic information and without the benefit of being able to infer ancestry along each chromosome (Kolbe et al. 2004; Lavergne and Molofsky 2007; Facon et al. 2008; Smith et al. 2020).

Based on the observed lower selfing rates in N. America compared to Europe, it has been suggested that under slightly increased outcrossing, mixing of haplotypes should be expected (Platt et al. 2010). In line with this hypothesis, we observed that most N. American *A. thaliana* populations have individuals with admixture from the dominant Hpg1 group. Being apparently already well-adapted to the N. American ecological context upon its introduction, today Hpg1 is a wide-spread lineage in N. America (Platt et al. 2010; Exposito-Alonso, Becker, et al. 2018). Admixture with Hpg1, followed by selection, might have benefited and accelerated the spread of new incoming lineages. A case in support of this can be made for groups that are found in Indiana, where the human settlers in the mid-19th century came predominantly from North Carolina, Virginia and Kentucky (Lynch 1915). Our demographic reconstructions using herbarium samples of SouthIndiana4 group
estimated divergence times of 121 years-before present (126-119 HPD; 95%, Fig. S21) coincidental to the human migration and its admixture with Hpg1 has resulted in groups that are extant in North Carolina (our collection), Kentucky (1001 Genomes collection) and Georgia (herbarium collection). Alternative explanations such as short-term fitness benefits through heterosis (Facon et al. 2005; Keller and Taylor 2010) can currently not be ruled out, but could be tested with common garden experiments across N. American field sites. We also note that while some populations show richness in terms of haplotype diversity inferred with WGS (Fig. 2) as a direct result of admixture events, our RAD-seq genotyping of over 2000 individuals and coarse-scale haplotype ancestry and diversity estimates (Fig. S22 and Fig. S23) suggest that populations CTDE, CTPN, MAUR and OHOS might have yet unexplored haplotype diversity.

**Purging of deleterious mutations**

An important aspect of colonization is the severe genetic bottleneck due to founder effects and subsequent accumulation of deleterious mutations (Kirkpatrick and Jarne 2000; Verhoeven et al. 2011; Willi 2013; Schrieber and Lachmuth 2017), further exacerbated by predominant self-fertilization (Noël et al. 2017). One of the ways out of this invasion paradox (Estoup et al. 2016) might be admixture between colonizing lineages, which can both remove deleterious mutations (Heller and Maynard Smith 1978) and generate new genetic combinations that are only adaptive in the new environment (Dlugosch and Parker 2008; Rius and Darling 2014). Consistent with the expectation under admixture alone, we observed that the admixed N. American *A. thaliana* haplogroups have fewer fixed derived deleterious alleles. When background levels of reduction in fixed derived alleles using synonymous mutations were accounted for, we observed that compared to source groups, non-synonymous tolerated mutations are removed at a lower rate than non-synonymous deleterious mutations in the admixed groups (Fig. S18D). This demonstrates that admixture has been successful in removing some of the potential non-synonymous mutational load carried by the founder lineages. A caveat is that the deleteriousness of variants is based on presumed reduction or loss of molecular function (Kono et al. 2018), even though gene inactivation can be adaptive as well (Olson 1999). A more direct approach to determining the extent of purging of mutational load in N. American colonizing lineages could come from direct estimates of local adaptation deficits and selection coefficients, by comparing the fitness of N. American individuals at their site of collection against a global sample of *A. thaliana* accessions (Exposito-Alonso et al. 2019) or by quantifying the amount of genetic rescue or F1 heterosis in crosses between populations (Koski et al. 2019).

**Resistance genes as loci under selection**

An indication of selection having potentially shaped the geographic distribution of genetic diversity in N. American *A. thaliana* is the observation of environmental dissimilarity between N. American haplogroups and their source lineages from the native range being negatively correlated with shared ancestry between them.
Given that \textit{A. thaliana} is a human commensal in its native range, it is not hard to envision that anthropogenically induced adaptation to invade (AIAI) (Hufbauer et al. 2012) might play a significant role in having accelerated \textit{A. thaliana}’s adaptation to the N. American environment.

If a species is far from an adaptive peak, large-effect mutations are particularly likely to affect the speed of adaptation (Fisher 1930). While the relative importance of abiotic and biotic factors for adaptation is still debated (Morris et al. 2020), some of the most drastic effects arise from disease resistance genes, where single genes have outsized effects on fitness and survival on plants in the presence of pathogens. In \textit{Capsella}, it has been shown that dramatic losses of genetic diversity after extreme genetic bottlenecks can be tolerated at most genes in the genome, except for immunity loci (Koenig et al. 2019). Our selection scans with \textit{A. thaliana} individuals from five different N. American populations have revealed that genes related to biotic stress are enriched among selection candidates. These include genes known to have alleles that confer resistance to two of the most prominent pathogens of \textit{A. thaliana}, \textit{H. arabidopsidis} and \textit{Pseudomonas} (Holub and Beynon 1997; Karasov et al. 2014; Karasov et al. 2018). One of the loci we found to be under selection is \textit{RPP13} (Rose et al. 2004), whose product recognizes the co-evolved, highly polymorphic effector ATR13 from \textit{H. arabidopsidis} (Allen et al. 2004). Another one is \textit{BAR1}, whose product recognizes members from the conserved HopB effector family from \textit{Pseudomonas} (Lafllame et al. 2020). While \textit{RPP13} is under balancing selection in at least part of the native range (Allen et al. 2004), we observe that a specific \textit{RPP13} allele is found on different haplotypes (Fig. S28), has a comparable nucleotide diversity to AEA ancestral source regions (Fig. S29) and has undergone a selective sweep in North American \textit{A. thaliana} populations. Given that \textit{H. arabidopsidis} appears to be an \textit{A. thaliana} specialist (Slusarenko and Schlaich 2003), it must have been introduced with its \textit{A. thaliana} host, and its genetic diversity in the introduced range might be as low or even lower than that of its host, potentially providing an explanation for the apparent selective sweep at \textit{RPP13}. Apart from haplotype homozygosity based scans, we observed high differentiation in the \textit{RPP4/5} region of the genome between individuals in the MISJ population that were or were not visibly infected with \textit{H. arabidopsidis} when we collected them. Local ancestry estimates confirm that this region of the genome has been entering the population through admixture and is of non-Hpg1 origin. As \textit{RPP4/5} is known to harbor high levels of polymorphism and is known to be involved in frequency-dependent selection for resistance to \textit{H. arabidopsidis} (Noël et al. 1999), this introgression event could be driven by positive selection against local strains of this pathogen.

**Conclusions**

Altogether, our analysis using whole-genome sequences from extant N. American \textit{A. thaliana} has established a scenario of multiple introductions from sources of previously diverged Eurasian lineages. We provide evidence that new haplotype diversity has been generated through wide-spread admixture among introduced lineages, relieving mutational load and providing raw material for selection to act upon. Our findings are thus
consistent with earlier proposals that hybridization can lead to the introduction of adaptive variation via introgression or admixture (Anderson 1948; Anderson 1949; Stebbins 1959; Grant 1981). The advent of molecular analyses has confirmed the relevance of hybridization for adaptation and speciation (Arnold 1996; Rieseberg 1997; Arnold 2004) and our observations are consistent with admixture being important for invasive success. Admixture can facilitate successful colonization when individuals from divergent populations have been recurrently introduced to a new range (Rius and Darling 2014; Dlugosch et al. 2015; Estoup et al. 2016). North American A. thaliana therefore may not have suffered from the genetic paradox of invasion (Allendorf and Lundquist 2003; Estoup et al. 2016). Finally, because A. thaliana has also colonized other continents, including S. America and Australia (Alonso-Blanco and Koornneef 2000; Kasulin et al. 2017), it will be interesting to determine both how genetic diversity of A. thaliana in these other places compares with N. America, and how genetic diversity of A. thaliana compares with that of other plants that have been inadvertently introduced to N. America by humans (Neuffer and Hurka 1999; Durka et al. 2005; La Sorte et al. 2007).

Methods

Sample collection and sequencing

Some samples were collected dried by pressing in acid-free paper with a wooden press for 8-12 weeks to produce herbarium samples. For other field samples, 2-3 well expanded leaves were collected in a microcentrifuge tube and immediately placed on dry ice and kept at -80°C until further processing. Seeds of Irish accessions were grown in the lab from seeds. Details of DNA extraction using different protocols and sequencing can be found in Supplementary Material.

Mapping and variant calling

Reads were mapped using bwa-mem (bwa-0.7.15) (Heng Li 2009) to the TAIR10 reference genome (https://www.arabidopsis.org/download_files/Genes/TAIR10_genome_release) and sorted using samtools v1.3 (Li et al. 2009). Reads from herbarium samples were additionally trimmed with skewer (v. 0.1.127) (Jiang et al. 2014) using default parameters and merged with flash (v. 1.2.11) (Magoč and Salzberg 2011) with a maximum overlapping value of 150 bp, prior to mapping. SNP calling was performed with the Genome Analysis Tool Kit (GATK) best practices with modifications for single-end reads (DePristo et al. 2011; Van der Auwerda et al. 2013). GATK tools used are described in supplementary methods and detailed parameters can be found in the script provided in the accompanying repository. Strategy and procedure to include SNPs from remaining A. thaliana global diversity dataset can be found in Supplementary Material.

Estimation of recombination rates

Haplotype phasing for estimation of recombination rate was performed with ShapeIt2 (v2.r837) (Delaneau et al. 2013) on samples from this project and a subset of the 1001 Genomes project (1001 Genomes Consortium...
2016). After phasing, the recombination rate variation along the chromosomes was estimated using LDhelmet v1.7 (Chan et al. 2012). Detailed procedure is described in supplementary methods.

**Population genetic analysis**

**PCA, UMAP and IBD:** Principal Component Analysis (PCA) was performed using SmartPCA of EIGENSOFT version 6.0.1 (Patterson et al. 2006) package. We used the first fifty PCs as input for generating two UMAP embeddings using Python package umap v0.4.6 (McInnes et al. 2018). Details of the analyses are in the supplementary methods. Identity-by-descent and identity-by-state analysis was carried out with PLINK v1.90 (Chang et al. 2015).

**Chromosome painting and clustering:** Clustering of individuals based on shared ancestry from haplotype data was performed using fineSTRUCTURE on a coancestry matrix derived with the software CHROMOPAINTER v2 (Lawson et al. 2012), which treats all the individuals (except the individual whose ancestry is being reconstructed) as donor haplotypes and generates a mosaic of shared chunks copied from these donors in a given recipient individual. Similarity in the patterns of shared chunks (copying vectors) is indicative of shared ancestry and is the basis of the model-based clustering approach taken by the fineSTRUCTURE algorithm. Specifically, we performed this analysis in the following hierarchical way:

1. All N. American individual haplotypes were painted as a mosaic of all other N. American individuals’ haplotypes (self-excluding).
2. All non-N. American (Afro-Eur-Asian / AEA) haplotypes were formed as a mosaic of each other. Based on the haplotype sharing these individuals were then clustered and grouped into what we call sub-clusters, clusters (comprising sub-clusters) and regions (comprising clusters representing specific geographical regions).
3. All N. American haplotypes were then formed as a mosaic of AEA haplotype clusters. Detailed description of the analysis is in supplementary methods.

**Treemix analysis:** We determined the phylogenetic relationship among the N. American groups and among these groups and AEA sub-clusters as inferred by fineSTRUCTURE using Treemix v1.13 (Pickrell and Pritchard 2012). Details are in Supplementary Methods.

**f3-outgroup analysis:** To determine the extent of shared drift between the Afro-Eur-Asian (AEA) sub-clusters (smallest fineSTRUCTURE grouping) and N. American haplogroups, we used f3-outgroup tests as described (Patterson et al. 2012). *N. American _j_0, AEA SubCluster _j_: Relicts (Fs12_3)* configuration was used and implementation of the test was carried out using R package “admixr” (Petr et al. 2019).

**qpWave and D-statistic analysis:** To determine the minimum number of ancestry waves from Afro-Eur-Asian(AEA) regions (comprised of different haplogroup sub-clusters defined by fineSTRUCTURE analyses) we used D-statistic and qpWave analysis from ADMIXTOOLS (Reich et al. 2012). Details of the tree configurations and outgroups can be found in Supplementary Materials.
**Rare allele sharing:** 1,039 AEA individuals that formed the fineSTRUCTURE sub-clusters were used as a reference panel to ascertain rare alleles and calculate rare allele sharing (RAS) between AEA sub-clusters and N. American haplogroups. The input files were prepared with the tools from repository at [https://github.com/stschiff/rarecoal-tools](https://github.com/stschiff/rarecoal-tools) and the analysis was performed by the pipeline available at [https://github.com/TCLamnidis/RAStools](https://github.com/TCLamnidis/RAStools). Minimum allele count of 2 and maximum allele count of 20 was used on the SNPs with less than 10% missing data. Alleles were polarized with the *A. lyrata* data.

**Phylogenetic methods:** Bayesian phylogenetic analyses were carried out using BEAST v.2.4.8 (Bouckaert et al. 2014) for groups Hpg1 and SouthIndiana4. Details of the substitution model and prior used are in Supplementary Material.

**Local ancestry inference:** We performed local ancestry inference (LAI) for RAD-seq genotyped samples from population MISJ and WGS samples from MISJ, NJSC, OHML and OHPR with Loter model (Dias-Alves et al. 2018). Exact parameters and individuals used as reference source groups are described in Supplementary Materials.

**Genetic differentiation in MISJ population:** We calculated genetic differentiation \(F_{ST}\) between individuals that showed visible symptoms of infection by *Hyaloperonospora arabidopsidis* (described in (Koch and Slusarenko 1990)) at the time of collection and individuals were visibly healthy. Detailed strategy to determine the significance of the differentiation between the two groups is described in Supplementary Materials.

**Environmental factor analysis**

Historical climate data from 1970-2000 were downloaded from WorldClim2.0 (Fick and Hijmans 2017) at 2.5-minute resolution using Python library latlon-utils 0.0.5 ([https://github.com/Chilipp/latlon-utils](https://github.com/Chilipp/latlon-utils)). Environmental variables average temperature (°C), precipitation (mm), solar radiation (kJ m\(^{-2}\) day\(^{-1}\)) and water vapor pressure (kPa) were used for further analysis. Pairwise Euclidean distances of all the environmental variables were calculated for each N. American haplogroup to AEA sub-clusters (mean Latitude -Longitude of individuals in a given sub-cluster was used) and standardized values were used to model shared drift (measured by \(f_3\)-outgroup statistics) among N. American haplogroups and AEA sub-clusters as a function of the environmental variables using Bayesian multilevel (hierarchical) linear regression. Description of the priors and hyper-priors is in the supplementary methods.

Projection of the N. American haplogroups in reduced dimension formed by standardized average temperature, precipitation and vapor pressure was performed using uniform manifold approximation and projection (UMAP) (McInnes et al. 2018). Two independent runs of UMAP were performed with different random numbers. In both the runs default "Euclidean" distance was used to compute distances in high dimensional space. We further used Ward’s linkage function on UMAP embeddings to determine hierarchical clustering patterns in the dataset based on Euclidean distance. Details of the scripts and notebooks used for the analysis are in the accompanying repository.
Estimation of scaled fixed-to-total derived allele ratio ($\Phi$) for N. American groups across mutational categories

Ancestral state of the positions were determined using pairwise alignments between $A. thaliana$ and outgroups $A. halleri$ and $A. lyrata$ (ftp://ftp.ensemblgenomes.org/pub/plants/release-44/maf/ensembl-compara/pairwise_alignments/). The detailed strategy is described in the Supplementary Methods.

Precomputed SIFT 4G predictions for $A. thaliana$ were obtained from (https://sift.bii.a-star.edu.sg/sift4g/public//Arabidopsis_thaliana/). Using these predictions, positions were divided into three different categories, 1.High-confidence deleterious mutations (score=0 to 0.05), 2.Non-synonymous tolerated mutations (score = 0.05 to 1) and , 3. synonymous mutations. We then calculated the ratio of fixed-to-total derived alleles ($\phi_{\text{category}}$) in every group separately for all categories (frequency of fixed alleles = 1). We also calculated the genome-wide ratio of fixed-to-total derived alleles ($\phi_{\text{genome}}$). $\phi$ is influenced by the increase in effective population size as a result of admixture between the source groups and divergence. To account for this in comparisons among groups, we scaled $\phi$ for high-confidence deleterious and for non-synonymous tolerated mutations by dividing it with $\phi$ calculated for synonymous mutations for individual groups:

$$\phi_{\text{category}} = \frac{\phi_{\text{category}}}{\phi_{\text{syn-tol}}}$$

In addition to the scaling with synonymous variation we separately scaled the ratio with genome-wide fixed-to-total derived alleles ratio (using all the derived variation).

**Genome-wide selection scans**

We performed haplotype homozygosity based selection scans to detect recent and ongoing selection. $iHS$ (integrated haplotype score) (Voight et al. 2006) and XP-EHH (cross-population extended haplotype homozygosity) (Sabeti et al. 2007) were calculated using hapbin (Maclean et al. 2015), details are described in the supplementary methods. Recombination map generated earlier was used in the estimation of both the statistics. $nS_L$ (number of segregating sites by length) (Ferrer-Admetlla et al. 2014), Garud’s H1, H12 and H2/H1 (Garud et al. 2015) (window size = 500, step size= 10), Tajima’s D (window size=50000 and step size=5000) were calculated with scikit-allel (Miles et al. 2020). Nucleotide diversity for the population was calculated using a pipeline described by (Martin et al. 2015).

$|iHS|$ and $|nS_L|$ were used in a complementary manner. As $iHS$ is known to be affected by recombination rate variation (O’Reilly et al. 2008), we used $iHS$ first, and based on empirical distribution of the scores, $p$-values were calculated per SNP. $nS_L$ was then calculated on the same dataset. As $nS_L$ is robust to variation in mutation and recombination rates (Ferrer-Admetlla et al. 2014), overlap of the SNPs that showed $|iHS| p$-value less than 0.001 and|$nS_L|$ higher than 2 was taken as a signal of selection. GO-term analysis of the genes carrying the candidate selected SNPs was performed with AgriGOv2 (Tian et al. 2017) with PlantGo-Slim categories.

For enrichment of GO terms, Fisher’s exact test with Bonferroni correction was used. We also performed GO-
term enrichment analysis using a permutation-based method implemented in Gowinda v1.12 (Kofler and Schlötterer 2012) because it takes into account gene clustering and size.

**Data and Code Availability**

Code and high resolution images from the main text are available from [https://github.com/weigelworld/north_american_A.thaliana](https://github.com/weigelworld/north_american_A.thaliana) repository. Short reads have been deposited in the European Nucleotide Archive under the accession number PRJEB42417.

**Author Contributions**

GS, JD and DW conceived the project. GS and JD organized the collection trips. GS and JD collected the samples. GS, JD, AB, MQD, AGH and DSL processed the samples for sequencing and performed DNA extractions. GS, JD and AB prepared sequencing libraries. GS and JD scripted and ran sequence processing pipelines. GS performed formal analysis. HAB, CF and DW supervised the research. GS wrote the original draft. GS, SML, DSL, CF, HAB and DW reviewed and edited the draft.

**Acknowledgements**

We thank Rebecca Schwab for help with organizing first collection trip; Claudia Friedemann, Cathy Herring (NCSU Central Crops Research Station), Carol Ann McCormick (University of North Carolina), Jon Peter (New York Botanical Garden), Nelson Garcia (Rutgers University), Rosa Raudales and Sanjukta Majumder (University of Connecticut), Robert Capers (George Safford Torey Herbarium), Irina Kadis and Alexei Zinovjev (Arnold Arboretum), Jason Parrish, Jiangbo Fan, Guo-liang Wang, Lynn Jin, David Mackey, Maria Bellizzi and John Freudenstein (Ohio State University), Neville Millar (Kellogg Biological Station), Eric Knox (Indiana University), Timothy Morton and Joy Bergelson (University of Chicago), Alan Fryday, Bethany Huot and Sheng Yang He (Michigan State University), and Christine Niezgoda (Field Museum of Natural History) for help during sample collection; Suresh Balasubramanian (Monash University) for seeds of Irish accessions; Sang-Tae Kim, Julian Regalado, Dino Jolic, Danielle Seymour, Jörg Hagmann and Jorge Quintana for help with setting up the RAD-seq analysis pipeline and for help throughout the project; Christa Lanz, Julia Hildebrandt for help with sequencing; and Stephan Schiffels (MPI for Science of Human History), Talia Karasov and Richard Neher for discussions. We thank anonymous reviewers for their constructive comments that helped make substantial improvements to the manuscript. This work was funded by the DFG through ERA-CAPS Project 1001GenomesPlus and the Max Planck Society.

**References**

1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in
Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.

Colautti RI, Lau JA. 2015. Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Mol. Ecol.* 24:1999–2017.

Dai CL, Vazifeh MM, Yeang CH-H, Tachet R, Wells RS, Vilar MG, Daly MJ, Ratti C, Martin AR. 2020. Population Histories of the United States Revealed through Fine-Scale Migration and Haplotype Analysis. *Am. J. Hum. Genet.* 106:371–388.

Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J. 2013. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* 93:687–696.

Dias-Alves T, Mairal J, Blum MGB. 2018. Loter: A software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* [Internet]. Available from: http://dx.doi.org/10.1093/molbev/msy126

Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD. 2015. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Mol. Ecol.* 24:2095–2111.

Dlugosch KM, Parker IM. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* 17:431–449.

Durka W, Bossdorf O, Prati D, Auge H. 2005. Molecular evidence for multiple introductions of garlic mustard (Alliaria petiolata, Brassicaceae) to North America. *Mol. Ecol.* 14:1697–1706.

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Xavier Picó F, Alonso-Blanco C, et al. 2017. African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* 114:5213–5218.

Ebenesersdóttir SS, Sandoval-Velasco M, Gunnarsdóttir ED, Jagadeesan A, Guðmundsdóttir VB, Thordardóttir EL, Einarsdóttir MS, Moore KHS, Sigurðsson Á, Magnúsdóttir DN, et al. 2018. Ancient genomes from Iceland reveal the making of a human population. *Science* 360:1028–1032.

Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M, Facon B. 2016. Is There a Genetic Paradox of Biological Invasion? *Annu. Rev. Ecol. Evol. Syst.* 47:51–72.

Exposito-Alonso M, Becker C, Schuenemann VJ, Reiter E, Setzer C, Slovak R, Brachi B, Hagmann J, Grimm DG, Chen J, et al. 2018. The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS Genet.* 14:e1007155.

Exposito-Alonso M, Exposito-Alonso M, Gómez Rodríguez R, Barragán C, Capovilla G, Chae E, Devos J, Dogan ES, Friedemann C, Gross C, et al. 2019. Natural selection on the Arabidopsis thaliana genome in present and future climates. *Nature* [Internet]. Available from: https://doi.org/10.1038/s41586-019-1520-9

Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. 2018. Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nat Ecol Evol* 2:352–358.

Facon B, Jarne P, Pointier JP, David P. 2005. Hybridization and invasiveness in the freshwater snail Melanoides tuberculata: hybrid vigour is more important than increase in genetic variance. *J. Evol. Biol.* 18:524–535.

Facon B, Pointier J-P, Jarne P, Sarda V, David P. 2008. High Genetic Variance in Life-History Strategies within Invasive Populations by Way of Multiple Introductions. *Curr. Biol.* 18:363–367.
Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* 31:1275–1291.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37:4302–4315.

Fisher RA. 1930. The Genetical Theory of Natural Selection. OxfordClarendon Press

Flegontov P, Altmüller NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkhoshbacht N, Candilio F, Culleton BJ, et al. 2019. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570:236–240.

Fulgione A, Hancock AM. 2018. Archaic lineages broaden our view on the history of Arabidopsis thaliana. *New Phytol.* 219:1194–1198.

Flegontov P, Atkinson NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkhoshbacht N, Candilio F, Culleton BJ, et al. 2019. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570:236–240.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet.* 11:e1005004.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet.* 11:e1005004.

Gelman A. 2006. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics* 48:432–435.

Heller R, Maynard Smith J. 1978. Does Muller’s ratchet work with selfing? *Genet. Res.* 32:289–293.

Heng Li RD. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754.

Hoffmann MH. 2002. Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). *J. Biogeogr.* 29:125–134.

Holub EB. 2001. The arms race is ancient history in Arabidopsis, the wildflower. *Nat. Rev. Genet.* 2:516–527.

Holub EB, Beynon JL. 1997. Symbiology of Mouse-Ear Cress (Arabidopsis Thaliana) and Oomycetes. In: Andrews JH, Tommerup IC, Callow JA, editors. Advances in Botanical Research. Vol. 24. Academic Press. p. 227–273.

Hsu C-W, Lo C-Y, Lee C-R. 2019. On the postglacial spread of human commensal Arabidopsis thaliana: journey to the East. *New Phytol.* 222:1447–1457.

Hufbauer RA, Facon B, Ravnigé V, Turgeon J, Foucaud J, Lee CE, Rey O, Estoup A. 2012. Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evol. Appl.* 5:89–101.
Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182.

Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, Kersten S, Lundberg DS, Neumann M, Regalado J, et al. 2018. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host Microbe* 24:168–179.e4.

Karasov TL, Kniskern JM, Gao L, DeYoung BJ, Ding J, Dubiella U, Lastra RO, Nallu S, Roux F, Innes RW, et al. 2014. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* 512:436–440.

Kasulin L, Rowan B, León RJC, Schuenemann VJ, Weigel D, Botto JF. 2017. A single haplotype hyposensitive to light and requiring strong vernalization dominates Arabidopsis thaliana populations in Patagonia, Argentina. *Mol. Ecol.* [Internet]. Available from: http://dx.doi.org/10.1111/mec.14107

Keller SR, Taylor DR. 2010. Genomic admixture increases fitness during a biological invasion. *J. Evol. Biol.* 23:1720–1731.

Kirkpatrick M, Jarne P. 2000. The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load. *Am. Nat.* 155:154–167.

Koch E, Slusarenko A. 1990. Arabidopsis is susceptible to infection by a downy mildew fungus. *Plant Cell* 2:437–445.

Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Neuffer B, Wright SI, Weigel D. 2019. Long-term balancing selection drives evolution of immunity genes in Capsella. *Elife* [Internet] 8. Available from: http://dx.doi.org/10.7554/eLife.43606

Kofler R, Schlötterer C. 2012. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28:2084–2085.

Kolbe JJ, Glor RE, Rodríguez Schettino L, Lara AC, Larson A, Losos JB. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177–181.

Kono TJY, Lei L, Shih C-H, Hoffman PJ, Morrell PL, Fay JC. 2018. Comparative Genomics Approaches Accurately Predict Deleterious Variants in Plants. *G3* 8:3321–3329.

Koski MH, Layman NC, Prior CJ, Busch JW, Galloway LF. 2019. Selfing ability and drift load evolve with range expansion. *Evol Lett* 3:500–512.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4:1073–1081.

Laflamme B, Dillon MM, Martel A, Almeida RND, Desveaux D, Guttmann DS. 2020. The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. *Science* 367:763–768.

La Sorte FA, Mckinney ML, Pyšek P. 2007. Compositional similarity among urban floras within and across continents: biogeographical consequences of human-mediated biotic interchange: Intercontinental compositional similarity. *Glob. Chang. Biol.* 13:913–921.

Lavergne S, Molofsky J. 2007. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc. Natl. Acad. Sci. U. S. A.* 104:3883–3888.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype
data. *PLoS Genet.* 8:e1002453.

Lee C-R, Svardal H, Farlow A, Exposito-Alonso M, Ding W, Novikova P, Alonso-Blanco C, Weigel D, Nordborg M. 2017. On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat. Commun.* 8:14458.

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B, Wellcome Trust Case Control Consortium 2, et al. 2015. The fine-scale genetic structure of the British population. *Nature* 519:309–314.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* 17:142–152.

Lynch WO. 1915. The Flow of Colonists To and From Indiana Before the Civil War. *Indiana Mag. Hist.* 11:1–7.

Maclean CA, Chue Hong NP, Prendergast JGD. 2015. *hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets*. *Mol. Biol. Evol.* 32:3027–3029.

Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32:244–257.

McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* [Internet]. Available from: http://arxiv.org/abs/1802.03426

Miles A, Bot P io, Rodrigues MF, Ralph P, Harding N, Pisupati R, Rae S. 2020. cggh/scikit-allel: v1.3.1. Available from: https://zenodo.org/record/3935797

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17:240–248.

Moest M, Van Belleghem SM, James JE, Salazar C, Martin SH, Barker SL, Moreira GRP, Mérot C, Joron M, Nadeau NJ, et al. 2020. Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. *PLoS Biol.* 18:e3000597.

Montinaro F, Busby GBJ, Pascali VL, Myers S, Hellenthal G, Capelli C. 2015. Unravelling the hidden ancestry of American admixed populations. *Nat. Commun.* 6:6596.

Morris WF, Ehrlén J, Dahlgren JP, Loomis AK, Louthan AM. 2020. Biotic and anthropogenic forces rival climatic/abiotic factors in determining global plant population growth and fitness. *Proc. Natl. Acad. Sci. U. S. A.* 117:1107–1112.

Neuffer B, Hurka H. 1999. Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. *Mol. Ecol.* 8:1667–1681.

Noël E, Jarne P, Glémin S, MacKenzie A, Segard A, Sarda V, David P. 2017. Experimental Evidence for the Negative Effects of Self-Fertilization on the Adaptive Potential of Populations. *Curr. Biol.* 27:237–242.
Noël L, Moores TL, van Der Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JD. 1999. Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of Arabidopsis. Plant Cell 11:2099–2112.

Olson MV. 1999. When less is more: gene loss as an engine of evolutionary change. Am. J. Hum. Genet. 64:18–23.

O’Reilly PF, Birney E, Balding DJ. 2008. Confounding between recombination and selection, and the Ped/Pop method for detecting selection. Genome Res. 18:1304–1313.

Passel JS, Fix M. 1994. US immigration in a global context. Indiana Journal of Global Legal Studies 2:5–19.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. Genetics 192:1065–1093.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L. 2013. On the accumulation of deleterious mutations during range expansions. Mol. Ecol. 22:5972–5982.

Petr M, Vernot B, Kelso J. 2019. admixr-R package for reproducible analyses using ADMIXTOOLS. Bioinformatics 35:3194–3195.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossshord O, Byers D, Donohue K, et al. 2010. The scale of population structure in Arabidopsis thaliana. PLoS Genet. 6:e1000843.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505:87–91.

Rédei GP. 1992. A Heuristic Glance at the Past of Arabidopsis Genetics. In: Koncz, C and Chua*, N-H and Schell, J, editor. Methods in Arabidopsis Research. WORLD SCIENTIFIC. p. 1–15.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, et al. 2012. Reconstructing Native American population history. Nature 488:370.

Rieseberg LH. 1997. Hybrid Origins of Plant Species. Annu. Rev. Ecol. Syst. 28:359–389.

Rius M, Darling JA. 2014. How important is intraspecific genetic admixture to the success of colonising populations? Trends Ecol. Evol. 29:233–242.

Rose LE, Bittner-Eddy PD, Langley CH, Holub EB, Michelmore RW, Beynon JL. 2004. The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in Arabidopsis thaliana. Genetics 166:1517–1527.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918.

Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, Loe L, Clarke R, Lyons A, Mortimer R, Sayer D, et al. 2016. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nat.
Schmickl R, Jørgensen MH, Brysting AK, Koch MA. 2010. The evolutionary history of the Arabidopsis lyrata complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evol. Biol.* 10:98.

Schrieber K, Lachmuth S. 2017. The Genetic Paradox of Invasions revisited: the potential role of inbreeding × environment interactions in invasion success: The Genetic Paradox revisited. *Biol Rev* 92:939–952.

Slusarenko AJ, Schlaich NL. 2003. Downy mildew of Arabidopsis thaliana caused by *Hyaloperonospora parasitica* (formerly *Peronospora parasitica*). *Mol. Plant Pathol.* 4:159–170.

Smith AL, Hodkinson TR, Villellas J, Catford JA, Csergő AM, Blomberg SP, Crone EE, Ehrlén J, García MB, Laine A-L, et al. 2020. Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proc. Natl. Acad. Sci. U. S. A.* 117:4218–4227.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23:23–35.

Stebbins GL. 1959. The Role of Hybridisation in Evolution. *Proceedings of the American Philosophical Society* 103:231–251.

Stephan W. 2019. Selective Sweeps. *Genetics* 211:5–13.

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. *agriGO* v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45:W122–W129.

Van de Weyer A-L, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JDG, Dangl JL, Weigel D, Bemm F. 2019. A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. *Cell* 178:1260–1272.e14.

Vatsiou AI, Bazin E, Gaggiotti OE. 2016. Detection of selective sweeps in structured populations: a comparison of recent methods. *Mol. Ecol.* 25:89–103.

Verhoeven KJF, Macel M, Wolfe LM, Biere A. 2011. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc. Biol. Sci.* 278:2–8.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.

Walsh B, Lynch M. 2018. Evolution and Selection of Quantitative Traits. Oxford University Press

Willi Y. 2013. Mutational meltdown in selfing Arabidopsis lyrata. *Evolution* 67:806–815.

Winter M, Kühn I, La Sorte FA, Schweiger O, Nentwig W, Klotz S. 2010. The role of non-native plants and vertebrates in defining patterns of compositional dissimilarity within and across continents: Compositional dissimilarity at large scales. *Glob. Ecol. Biogeogr.* 19:332–342.