Museum genomics reveals temporal genetic stasis and global genetic diversity in *Arabidopsis thaliana*

Running title: Museum genomics of Arabidopsis

Lua Lopez[1,2], Patricia L. M. Lang[3], Stephanie Marciniak[4], Logan Kistler[5], Sergio M. Latorre[6], Asnake Haile[1,7], Eleanna Vasquez Cerda[1], Diana Gamba[1], Yuxing Xu[1], Patrick Woods[8], Mistire Yifru[7], Jeffrey Kerby[9], John K. McKay[8], Christopher G. Oakley[10], Jon Ågren[11], Tigist Wondimu[7], Collins Bulafu[12], George H. Perry[1,4], Hernán A. Burbano[6], Jesse R. Lasky[1,13]*

[1]Department of Biology
Pennsylvania State University
University Park, PA
USA

[2]Department of Biology
California State University, San Bernardino
San Bernardino, CA
USA

[3]Department of Plant and Microbial Biology
University of California, Berkeley
Berkeley, CA
USA

[4]Department of Anthropology
Pennsylvania State University
University Park, PA
USA

[5]Smithsonian Institution
Washington, DC
USA

[6]Centre for Life's Origins and Evolution
Department of Genetics, Evolution and Environment
University College London
London, England
UK

[7]Department of Plant Biology and Biodiversity Management
Addis Ababa University
Addis Ababa
Ethiopia

[8]Department of Soil and Crop Sciences
Colorado State University
Ft. Collins, CO
USA

[9]Aarhus Institute of Advanced Studies
Aarhus
Denmark

[10] Department of Botany and Plant Pathology,
and The Center for Plant Biology
Purdue University
West Lafayette, IN
USA

[11]Department of Ecology and Genetics
Uppsala University
Uppsala
Sweden

[12]Department of Plant Sciences, Microbiology, and Biotechnology
Makarere University
Kampala
Uganda

[13]PAC Herbarium
Pennsylvania State University
University Park, PA
USA

*Correspondence: lasky@psu.edu

Abstract

Global patterns of population genetic variation through time offer a window into evolutionary processes that maintain diversity. Over time, lineages may expand or contract their distribution, causing turnover in population genetic composition. At individual loci, migration, drift, and selection (among other processes) may affect allele frequencies. Museum specimens of widely distributed species offer a unique window into the genetics of understudied populations and changes over time. Here, we sequenced genomes of 130 herbarium specimens and 91 new field collections of *Arabidopsis thaliana* and combined these with published genomes. We sought a broader view of genomic diversity across the species, and to test if population genomic composition is changing through time. We documented extensive and previously uncharacterized diversity in a range of populations in Africa, populations that are under threat from anthropogenic climate change. Through time, we did not find dramatic changes in genomic composition of populations. Instead, we found a pattern of genetic change every 100 years of the same magnitude seen when comparing Eurasian populations that are 185 km apart, potentially due to a combination of drift and changing selection. We found only mixed signals of polygenic adaptation at phenology and physiology QTL. We did find that genes conserved across eudicots show altered levels of directional allele frequency change, potentially due to variable purifying and background selection. Our study highlights how museum specimens can reveal new dimensions of population diversity and show how wild populations are evolving in recent history.

Keywords: time series genomics, genomic diversity, museum specimens, allele frequency change, population structure

Introduction

The genomic composition and diversity of species across their geographic range can reveal the relative importance of processes such as demography, drift, and spatially varying selection. The evolutionary history of a species is reflected in its present-day patterns of relatedness across the genome, and these patterns will change in the future as demography and environments change (Fulgione & Hancock, 2018; C.-R. Lee et al., 2017; Olalde et al., 2018; Rhoné et al., 2020). For example, regions that differ in their history of colonization may show distinct patterns of diversity and genetic turnover through space (Lasky et al., 2024; Petkova et al., 2016). Furthermore, knowledge of geographic patterns in diversity and population structure can be used to optimize sampling designs for studies of genetic variation in phenotype (Fulgione & Hancock, 2018).

While biologists have often studied geographic patterns in genotype and phenotype, their study of the temporal genomic dynamics has developed only more recently, largely due to a previous lack of data. As global environments are changing, species range-wide study of temporal dynamics in population genomics could reveal corresponding evolutionary changes. However, the breadth of sampling required through space and time is a logistical hurdle. Recent studies have used ancient human sequences to demonstrate dramatic turnovers in population ancestry in some regions, as new lineages replaced older ones (Allentoft et al., 2024; Kennett et al., 2022; LaPolice et al., 2024; Olalde et al., 2018), or consistent geographic population structure through time (Antonio et al., 2024). For non-human species, long-term field studies have offered powerful windows into temporal dynamics (Bergland et al., 2014; Lynch et al., 2024; Troth et al., 2018). Global natural history collections made over the last couple of centuries are an underutilized but incredibly valuable resource for biology, holding a wealth of information on genotype and phenotype (Burbano & Gutaker, 2023; Lopez et al., 2020). For example, researchers have used museum specimens to show demographic turnover in a plant species where there was dramatic expansion of tetraploids across parts of Europe formerly inhabited only by diploids (Rosche et al., 2024). Recent population studies have also used sequences of museum specimens to identify changes in genetic diversity over time (Bi et al., 2019; Gauthier et al., 2020) and the history of spread of plant pathogens (Yoshida et al., 2013). Furthermore, museum collections often hold samples from a wide geographic area, allowing greater spatial coverage in evolutionary studies.

Arabidopsis thaliana, the model plant (hereafter referred to as "Arabidopsis"), is a useful system for studying drivers of geographic and temporal turnover in genotype. Arabidopsis is an annual, largely self-fertilizing plant species with a broad native distribution in Eurasia and Africa (Fulgione & Hancock, 2018; Yim et al., 2023). Populations of Arabidopsis occur across a wide range of climatic regions and different environments (e.g., from the Mediterranean coast to alpine systems) (Ågren & Schemske, 2012; Elfarargi et al., 2023; Gamba et al., 2024; Montesinos et al., 2009; Yim et al., 2023). Global studies on *A. thaliana* have revealed high levels of genetic and phenotypic diversity among populations (Koornneef et al., 2004; Mitchell-Olds, 2001) and this species has become a model for eco-evolutionary studies (Takou et al., 2019; Turner et al., 2020; Vasseur et al., 2018; Wuest & Niklaus, 2018).

Along with life history variation across contrasting environments, *A. thaliana* displays substantial large-scale population structure (Beck et al., 2008; Lasky et al., 2012; Schmid et al., 2006; Sharbel et al., 2000). The largest genetic divisions within this species can be found between most Eurasian genotypes and the so-called relict lineages associated with distinct mountain ranges in Africa, Atlantic islands, and some areas of Mediterranean Europe (Alonso-Blanco et al., 2016; Durvasula et al., 2017; Fulgione & Hancock, 2018; Gamba et al., 2024; C.-R. Lee et al., 2017). However, many such regions, especially in Africa (Yim et al., 2023), have been little-sampled and it is likely there are additional uncharacterized unique lineages. Isolation by distance has also been detected at regional and local scales (Hesen et al., 2024; Lasky et al., 2024; Picó et al., 2008; Tyagi et al., 2016) suggesting that Arabidopsis gene flow is rather limited in space.

Plant populations are often adapted to their local environment (Ågren & Schemske, 2012; Fournier-Level et al., 2011; Hereford, 2009; Lasky et al., 2018; Leimu & Fischer, 2008). Individual genes underlying environmental adaptation in plants have been characterized for a number of these traits. For example, local adaptation has been mapped to natural variation in flowering time and phenology genes (Fulgione et al., 2022; Martínez-Berdeja et al., 2020), genes affecting freezing tolerance (G. Lee et al., 2024; Monroe et al., 2016), trichomes (Arteaga et al., 2022), floral pigmentation (Todesco et al., 2022), and resistance to parasites and pathogens (Bellis et al., 2020; Karasov et al., 2014). However, these loci have been primarily identified and studied in the context of existing contemporary diversity, while loci underlying adaptation to environmental change through time are less well known (but see (Exposito-Alonso, Becker, et al., 2018; Franks et al., 2016; Lang et al., 2024; Troth et al., 2018). Furthermore, across the vast Arabidopsis range major elevational clines in some traits and loci reverse in direction between different mountains, indicating alternative local adaptation strategies (Gamba et al., 2024). This strongly suggests consideration of regional patterns is important to understand spatiotemporal patterns of genomic variation in Arabidopsis.

Environments are always changing, but current anthropogenic changes are especially rapid, and wild organisms are responding in some dramatic ways. Resurrection and long-term experimental studies have shown evolutionary changes in quantitative traits and allele frequencies (Anstett et al., 2024; Hamann et al., 2018; Sekor & Franks, 2018) and studies of museum specimens have shown changes in traits in response to environmental change (DeLeo et al., 2020; MacLean et al., 2018; Ng et al., 2023). In Arabidopsis we documented temporal phenotypic changes in herbarium specimens from the last two centuries, namely in collection date, leaf C:N, and leaf $d^{15}N$ (DeLeo et al., 2020). The degree to which these changes are genetic versus plastic, and potential genetic changes are oligogenic versus polygenic, remains unknown. Recently, we used herbarium sequences to show that alleles associated with decreased stomatal density were rising in frequency at multiple loci (Lang et al., 2024; Latorre et al., 2022; Lopez et al., 2022). However, the degree of ancestry turnover in Arabidopsis populations and the potential change in quantitative trait loci (QTL) for ecologically important traits remains unexplored.

Multiple types of selection can influence temporal allele frequency dynamics. Abiotic and biotic environments can undergo a sustained directional shift, such as greenhouse gas-

184 induced warming, causing selection for new traits (Lynch & Lande, 1993). As a result, the
185 underlying allele frequencies may show directional shifts, including sweeps (Hayward & Sella,
186 2022; Höllinger et al., 2019). Next, new beneficial mutations may rise in frequency, with some
187 sweeping to fixation (Patwa & Wahl, 2008). Other types of selection (e.g. frequency
188 dependence) might cause more complicated temporal dynamics, which we set aside for our
189 current study (Fijarczyk & Babik, 2015; Siewert & Voight, 2017). Here, we focus on directional
190 allele frequency shifts. By identifying genes showing dramatic allele frequency shifts,
191 researchers may gain insight into mechanisms of adaptation to directional shifts in environment.
192 Of course, drift will also cause allele frequency shifts over time, thus it may be necessary to
193 identify loci with strong shifts relative to the genomic background (Akbari et al., 2024).
194 Despite changing environments, it is likely that many traits and loci do not experience
195 changes in selection and instead are subject to consistent stabilizing or purifying selection. For
196 example, many genes are highly conserved in structure over speciation events and many millions
197 of years (Kachroo et al., 2015; Margoliash, 1963). However, it is important to recognize that
198 purifying selection will still lead to non-random change in frequencies of segregating alleles over
199 time, i.e. decreases in deleterious allele frequency. In simulations, (Buffalo & Coop, 2020)
200 demonstrated how background selection caused genome-wide positive temporal autocorrelation
201 in allele frequencies (i.e. directional allele frequency change) as deleterious variants are selected
202 against. Thus, conserved genes might show some evidence of directional allele frequency
203 change. However, it is unknown how purifying and background selection impact locus-specific
204 temporal change.
205 Here, we sequenced 130 herbarium specimens of *Arabidopsis thaliana* and combined
206 these with new sequencing of recent field collected accessions and seedbank accessions, and
207 published sequences of natural inbred lines and herbarium specimens (Alonso-Blanco et al.,
208 2016; Durvasula et al., 2017). These genotypes greatly expand the geographic range of
209 sequenced Arabidopsis genomes into regions of Africa and Asia, and were collected from nature
210 across a period of nearly 200 years. These data were used for several goals. First, we sought to
211 generate a more complete view of Arabidopsis genetic diversity and population structure.
212 Second, we sought to determine how population genetic composition has changed over time
213 across the species range. Toward these goals we focused on the following questions:
214 1. What is the geography of population structure and relatedness across Africa and Eurasia?
215 Are African populations diverse and genetically distinct from each other and from those
216 in Eurasia (i.e. "relicts")?
217 2. Has geographic population structure remained static? Or have new lineages spread across
218 the Arabidopsis range over the last 150 years?
219 3. Is there genetic turnover through time, indicating isolation-by-time?
220 4. Are QTL for ecologically important traits enriched for temporal turnover in allele
221 frequency, suggesting changing selection?

5. Do conserved genes show lower temporal turnover than non-conserved loci, suggesting continued conservation? Or do conserved genes show higher turnover due to purifying and background selection?

Methods

Sampling of herbarium specimens, wild individuals, and published data

We obtained genomes of 527 *Arabidopsis thaliana* samples for this study (Table S1). This includes data newly generated (n=225) from multiple sources: herbarium specimens with ancient DNA (aDNA) protocols in a clean lab (n=130), new collections from extant wild populations (n=91), and stock center accessions of (n=4) naturally inbred lines. We also used 302 published genomes from stock center lines (n=199), herbarium (n=38) and natural populations (n=65) (Alonso-Blanco et al., 2016; Durvasula et al., 2017; Lang et al., 2024). The published genomes from stock center lines were mostly collected in the last two decades (Figure 1). However, a few dozen were originally collected in western Europe in earlier decades. For temporal analyses, these accessions were included based on the year they were collected in the wild, following (Wilczek et al., 2014).

Herbarium sampling was greatest for Germany, Spain, and Norway (n>30 accessions each), with additional sampling from Russia, Ethiopia, and Britain (n>10 accessions each). To uncover population genomic patterns in unstudied regions, we included samples from regions previously unrepresented in the literature: 5 countries (Kenya, Rwanda, Eritrea, the Democratic Republic of Congo, Nepal) and a large Mediterranean island (Sardinia). We also sequenced new populations from north of the Arctic circle, up to >68º latitude in Norway. The herbarium samples ranged in year of collection from 1817 (Germany), 1838 (Ethiopia), 1860 (Russia) to as recent as 2011 (Tanzania) and 2010 (Spain). To compare with herbarium specimens from the same region, we made new field collections of 91 individuals, 21 from Ethiopia (in 2017/2018), 9 from Uganda (in 2018), and 61 from Norway (in 2009/2010).

Publicly available data were obtained from 199 accessions of the 1001 Genomes project, selected based on geographic proximity to herbarium specimens (Alonso-Blanco et al., 2016) and from 70 African samples (65 individuals from wild populations and 5 herbarium specimens) of (Durvasula et al., 2017). In total, our samples encompass 32 countries in Eurasia and Africa, representing a large part of the native range of Arabidopsis (Figure 1) (Yim et al., 2023). We then down-sampled the modern collections from Morocco and Norway to balance with the sampling from herbarium specimens.

DNA extraction, library preparation and sequencing: Newly generated data

130 newly obtained samples from herbaria were transported into a dedicated ancient DNA clean laboratory at Pennsylvania State University, USA. DNA was extracted and using specialized protocols detailed in Supplemental Material. After DNA extraction, we investigated the DNA fragmentation pattern for each sample and proceeded to perform library preparation based on the

262 particular pattern of each sample. Using the following criterion, samples were categorized as high
263 or low DNA fragmentation. Lowly fragmented samples showed DNA above the 500 bp ladder and
264 were subjected to a short shearing step (30 seconds) using a M220 Focused Ultrasonicator
265 (Covaris) prior to library preparation (detailed library preparation protocol in Supplemental
266 Material), while highly fragmented samples were unsheared.
267 A total of 95 fresh samples were included in this study (61 field samples from Norway, 30
268 from Africa and 4 from the INRA stock center). Thirteen fresh leaf samples were collected from
269 the field in East Africa while the remaining 17 were collected as seeds. Details for DNA extraction
270 and library preparation are found in the Supplemental Material. Samples were sequenced on
271 Illumina platforms to obtain PE 150bp reads.
272
273 *Data preprocessing and* de novo *SNP calling*
274 Full details on filtering, mapping, and SNP calling are found in the Supplemental Materials.
275 Herbarium samples were mostly newly generated data (130). To these we added five samples from
276 the African genomes (Durvasula et al., 2017) and 33 German genomes (Lang et al., 2024) also
277 collected from herbarium vouchers. Raw data for the African genomes was obtained from the
278 European Nucleotide archive (ENA) of the European Molecular Biology Lab - European
279 Bioinformatics Institute (EMBL - EBI) and German genomes from (Lang et al., 2024). Raw read
280 sequence data were assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects) to
281 confirm that they met our quality standards. Genome sequencing coverage ranged from 81.41% to
282 98.91%. Mean depth of the covered portion of the genome showed a wide range, from 1 to 42.6
283 (median=8.4, Table S2). For all herbarium samples cytosine deamination profiles characteristic of
284 ancient DNA were verified using mapDamage 2.0 (Jónsson et al., 2013) (Figure S1). SNP
285 discovery was done using a pseudohaplotype approach (see Supplemental Materials for more
286 detail) (Kistler et al., 2018).
287
288 *Statistical analysis*
289 *Population structure* - We first analyzed global population structure by generating a set of high
290 coverage and unlinked SNPs, i.e. those that remained after excluding those with >5% missing
291 calls and filtering for linkage disequilibrium (PLINK *indep-pairwise* filtering SNPs with R^2>0.1
292 in 50 SNP windows with 10 SNP steps between windows) (Chang et al., 2015). We calculated
293 pairwise genetic distances between all samples (identity-by-state/Hamming distance, PLINK, flat
294 missingness correction) to generate a neighbor joining tree and multidimensional scaling k=2
295 plot. We also calculated genetic distances and MDS in two dimensions separately after repeating
296 the missingness and linkage filtering for two regional sets: 1) east and south African genotypes
297 and 2) Eurasian genotypes.
298 To further characterize population structure, we implemented ADMIXTURE genetic
299 clustering, across a range of cluster numbers (1 to 15) (Alexander et al., 2009). We applied
300 ADMIXTURE to all global samples combined as well as a subset of south and east African
301 samples, the region sampled well by our new data but previously the least studied.

Isolation by distance - We tested for geographic population structure and isolation by distance in spatially explicit analyses. To account for non-independence among pairwise distance measures involving the same individual, we fit mixed-effects models that include a random effect for each individual. We then compared them with null models based on AIC (ResistanceGA, (Peterman, 2018), these results were consistent when we excluded samples with >25% missing SNP calls).

To better understand the spatial scales of isolation by distance and its heterogeneity among the previously well studied Eurasian populations compared to the newly sampled regions of Africa, we calculated wavelet genetic dissimilarity (Lasky et al., 2024) which quantifies scale-specific genetic turnover among populations. That is, wavelet dissimilarity is a measure of genetic distance at a specific spatial scale (determined by the dilation of the wavelet function). Wavelet genetic dissimilarity increases at greater scales under patterns of isolation by distance, patterns that may differ between regions based on their level of gene flow and population age (Lasky et al., 2024). Within Eurasia and two African regions, we calculated the average wavelet dissimilarity among sampled locations at a range of scales from 1-500 km.

Isolation by time - We studied whether geographic population structure was stable over the period of the study (1817-2018). To test for major turnover in genetic composition, we focused on the well-sampled regions in Europe and asked whether different genetic clusters emerged within regions over time. We implemented ADMIXTURE on Eurasian genotypes and chose K=6 because it resolved many regional differences in genotype before a sharp increase in cross-validation error at K=7 (Figure S3). We then divided the well-sampled areas into five discrete regions: Iberia, Britain, Norway, Northern Germany, and Southern Germany, each dominated by a different genetic cluster. We then used non-parametric Spearman's rank correlations to test whether the proportional assignment of genotypes for the most dominant ADMIXTURE cluster within each region changed over time.

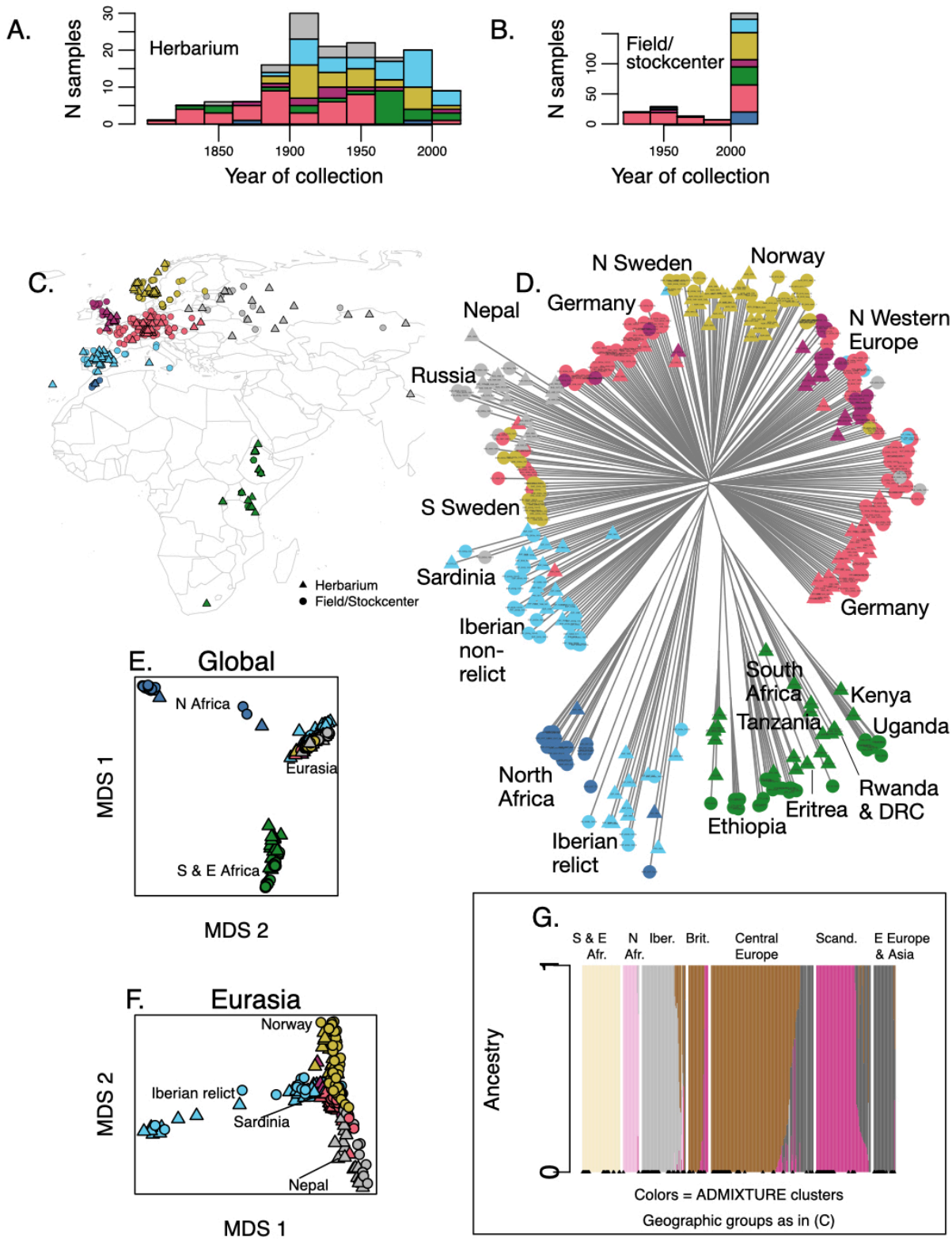
To test for more continuous temporal genomic change, we fit mixed effects models as with geographic distance, but added a covariate for separation in time (year of collection) to explain pairwise genetic distances. As with the pure geographic models, these models include individual-specific random effects to account for individuals that may have particularly low or high genetic distance compared to all others. For three regions that were the best sampled (Germany, Norway, and Iberia), we also tested for turnover through time within the region while accounting for geographic distance.

Change in allele frequency at specific loci – We tested for evidence of directional allele frequency shifts in QTL underlying traits that are potentially important in environmental adaptation in Arabidopsis and show evidence of temporal phenotypic change. We previously documented change in Arabidopsis herbarium phenotypes, including many of the sequenced specimens here (DeLeo et al. 2020). In particular, we found an increase in the date of collection and photothermal units accumulated by date of collection, signifying phenological shifts where

342 plants were collected later in growing seasons across eastern Europe and central Asia (DeLeo et
343 al. 2020). We also found an increase in leaf C:N ratio over time in a region from the
344 Mediterranean to central Europe (DeLeo et al. 2020). While DeLeo et al. (2020) found no change
345 in ^{13}C discrimination ($\Delta^{13}C$) over time, Lang et al. (2024) recently showed changes in the
346 frequency of alleles associated with stomatal traits over time, thus we decided to include putative
347 QTL for $\Delta^{13}C$. We used mixed-model GWAS results for flowering time from (Alonso-Blanco et
348 al., 2016), for germination/dormancy from (Martínez-Berdeja et al., 2020), and for leaf C:N and
349 $\delta^{13}C$ from (Gamba et al., 2024). For flowering time (the trait measured on the most genotypes),
350 we further implemented regional GWAS separately for Iberia, Germany, and Fennoscandia
351 (Norway, Sweden, Finland) (Alonso-Blanco et al., 2016). We used univariate linear mixed
352 models in GEMMA v.0.98.3 (Zhou & Stephens, 2012) excluding loci with MAF < 0.05,
353 calculating Wald test P-values.
354 To test for changes in selection over time at putative QTL for these traits, we first
355 characterized temporal directional allele frequency trends across the genome. We used several
356 analyses with different strengths. We first used non-parametric Wilcoxon tests determining if
357 alternate alleles had different median years of sample collection (function 'wilcox.test' in base
358 R) (R Core Team, 2023). Additionally, we used multiple logistic regression to test for allele
359 frequency change over years while controlling for linear effects of latitude and longitude
360 (function 'glm' in base R). Finally, we also included a test that included kinship random effects,
361 to identify loci with patterns deviating from the genomic background change through time
362 (Akbari et al., 2024), while also accounting for spatial trends associated with latitude and
363 longitude (using "mmer" function in the "sommer" package) (Covarrubias-Pazaran, 2016).
364 We tested for these allele frequency temporal dynamics across the native range (Eurasia
365 and Africa), and also within well-sampled regions (Iberia, Germany, Norway), because different
366 regions may undergo different changes in selection or phenotypic change (DeLeo et al. 2020).
367 We tested whether top GWAS loci were enriched for temporal allele frequency shifts. We
368 determined the strength of temporal change in SNPs (defined by the bottom 0.05, 0.1, 0.25 *p*-
369 value quantiles for the temporal allele frequency change tests) among the GWAS top SNPs (25,
370 50, or 100 SNPs, after removing nearby SNPs within 25 kb). We compared this quantile to those
371 generated from a null distribution generated from 1,000 circular permutations (around the
372 genome) of SNPs' trait temporal change p-values.
373
374 *Purifying and background selection* - We tested whether there was evidence of distinct
375 evolutionary processes acting on genic DNA versus non-genic DNA, and we also tested
376 conserved versus non-conserved genes. First, we compared the 0.01 p-value quantile of temporal
377 allele frequency change for genic versus non-genic markers. To generate a null distribution for
378 this comparison, we conducted 1,000 permutations of annotation of genic versus non-genic
379 (circularly around the genome).
380 Next, we used a list of Arabidopsis genes from orthogroups conserved across eudicots
381 identified by (Sun et al., 2018): 16,799 conserved versus 10,856 non-conserved genes.

382 Conserved genes were defined based on orthogrouping of all genes from Arabidopsis and six
383 other eudicot species (divergence >120 mya, *Ipomoea nil, Solanum tuberosum, Solanum*
384 *lycopersicum, Capsicum annuum, Coffea canephora, Mimulus guttatus*) (Sun et al., 2018).
385 Orthogroups that contained genes from at least six of these seven species were defined as
386 conserved orthogroups. Arabidopsis genes within these conserved orthogroups were classified as
387 conserved, while those not in conserved orthogroups were classified as non-conserved.
388 We first compared the 0.01 quantile of temporal allele frequency changes between all
389 genic SNPs in conserved versus non-conserved genes, using permutations of SNP annotation.
390 The site frequency spectrum likely differs between genic and intergenic SNPs, and between
391 SNPs in conserved versus non-conserved genes, potentially influencing power to detect temporal
392 change. Thus, we also tested these genomic contexts for enrichment of temporal allele frequency
393 changes for "rare" ($0.05 \geq MAF < 0.1$) and "common" ($MAF > 0.4$) SNPs in separate analyses.
394 We also conducted gene-level tests. Because longer genes tend to show greater
395 conservation (potentially stronger purifying selection) than shorter genes (Lipman et al., 2002),
396 we divided genes into bins based on amino acid lengths. The effect of length on conservation
397 was evident in the data from (Sun et al., 2018): non-conserved genes' median length was 226
398 amino acids while the conserved median was 416 (Wilcoxon test, $p<10^{-16}$). To obtain a test
399 statistic for temporal change of each gene, we calculated the lower 0.1 tail (or 0.25, results were
400 qualitatively the same) p-values for temporal change for SNPs in the coding region. We then
401 permuted gene conservation status circularly around the genome 1000 times and developed a
402 null distribution of median difference in gene p-values between conserved vs non-conserved
403 genes, matching for gene length (in five bins: <200, 201-400, 401-600, 601-800, and 801-1000
404 amino acids long). We then compared the null permutations to the observed in a two-tailed test
405 for each gene length bin. Because little is known about how background selection might
406 influence variation in temporal allele frequency change across the genome, we conducted
407 forward genetic simulations of purifying and background selection (see Supplemental Material
408 for more detail).
409
410 *Genomic prediction of trait change* - As an alternate approach to testing for directional evolution
411 of ecologically important traits, we also used whole-genome prediction. Whole genome
412 prediction models can work well for polygenic traits and assume an infinitesimal model of trait
413 variation (Hill & Kirkpatrick, 2010) and have previously been successfully employed to predict
414 local adaptation in Arabidopsis (Gienapp et al., 2017). To create a genomic prediction model, we
415 used a training dataset of 421 sequenced *A. thaliana* genotypes (Alonso-Blanco et al., 2016).
416 Each genotype had flowering time data at constant temperatures of 10ºC (FT10) and 16ºC
417 (FT16) (Alonso-Blanco et al., 2016). The kinship matrix was generated with PLINK v1.90b6.26
418 (Chang et al., 2015) using a minor allele frequency filter of 0.05. The 'kin.blup' function from
419 the rrBLUP package (Endelman, 2011) was used to conduct genomic predictions for flowering
420 time. To assess the accuracy, or extent of overfitting in our genomic prediction model, we cross-
421 validated our data for FT10 and FT16 by conducting tenfold genomic predictions. We then used

422 multiple linear regression models to test for changes in predicted flowering times across years
423 within each of three regions (Iberia, Germany, and Norway), while accounting for elevation,
424 latitude, and longitude.
425
426 **Results**
427
428 *An expanded view of Arabidopsis population structure in its native range*
429 We first analyzed global population structure using a set of high coverage, unlinked 201,299
430 SNPs. An MDS plot in two dimensions separates genotypes from North Africa, East and South
431 Africa, and Eurasia (Figure 1C). ADMIXTURE cross-validation error was minimized at K=6.
432 At K=6 the samples split into North African, South and East African, Iberian, British/Central
433 European, Scandinavian, and Eastern European/Asian groups (Figure 1E, see other K values in
434 Figure S2).
435
436


A.
N samples
Herbarium
Year of collection
B.
N samples
Field/
stockcenter
Year of collection
C.
Herbarium
Field/Stockcenter
D.
N Sweden
Norway
Germany
Nepal
N Western
Europe
Russia
S Sweden
Sardinia
Germany
Iberian
non-
relict
South
Africa
Kenya
Tanzania
Uganda
North
Africa
Rwanda
& DRC
Eritrea
Ethiopia
Iberian
relict
E. Global
N Africa
Eurasia
S & E Africa
MDS 1
MDS 2
F. Eurasia
Norway
Iberian relict
Sardinia
Nepal
MDS 2
MDS 1
G.
S & E
Afr.
N
Afr.
Iber.
Brit.
Central
Europe
Scand.
E Europe
& Asia
Ancestry
Colors = ADMIXTURE clusters
Geographic groups as in (C)

437
438

Figure 1. Sequenced accessions showing their distribution over time (A-B) and colored according to geography (C) or ADMIXTURE cluster (G) and with shapes indicating whether they were herbarium specimens (triangles) or sequenced from stock center or newly collected accessions (circles). Neighbor joining tree (D) includes sample codes with country abbreviation and year collected from the wild (Table S1). Two-dimensional MDS for the entire dataset (E) highlights the divergence between Africa and Eurasia and (F) for Eurasia highlights the distinct nature of the Iberian relicts. Specimens from countries of note are labeled on the tree. Cross validation error lowest for ADMIXTURE K=6 so these results are shown (G), with triangles at bottom indicating herbarium specimens. Shown are only accessions included in population genetic analyses (after filtering for balance between modern collections and herbarium).

The Sardinian sample was from low elevation near the coast and was assigned to a mix of Eurasian non-relict and North African/Iberian relict ancestries. The Nepalese sample was estimated to have primarily Eurasian non-relict ancestry, but with small portions of ancestry from both African groups, suggesting potential ancestry of a relict group. The MDS in two dimensions of Eurasian populations showed the Nepalese sample is toward one extreme alongside Russian, Kazakh, and Ukrainian populations (Figure 1D).

African Arabidopsis populations are understudied but clearly genetically distinct (Figure 1B, 1C) (Durvasula et al., 2017). Because these African populations are under threat from future climate change (Yim et al., 2023), and because the bulk of previously unstudied countries contributed by our herbarium samples were from East Africa, we conducted a focused analysis on African populations. A neighbor joining tree showed major differentiation among mountains (Figure 2A). This differentiation is consistent with the overall rarity of Arabidopsis and low suitability in low elevations in East Africa (Yim et al., 2023), suggesting very little gene flow between mountains over long periods. However, the Tanzanian (mostly lower elevation mountains), Mt Suswa (Kenya), and South African populations were more similar to each other than were genotypes across other individual mountains, suggesting greater or more recent gene flow in these more southerly regions (Figure 2). There were signals of geographic barriers, with Ethiopian populations on either side of the Rift Valley being most different from each other (Figure 2).

An MDS plot of East/South African samples (n=52) divided Horn of Africa populations (Ethiopia & Eritrea) from those further south along the first dimension, while the second dimension separated populations on either side of the Rift Valley in Ethiopia from each other as well as those from further south (Figure 2C). ADMIXTURE K = 4 had the lowest cross validation error, splitting high elevation Ethiopian populations in Bale and Choke into two groups on either side of the rift valley, Mt Elgon & Mt Kenya shield volcanoes in Uganda & Kenya as another group. The Simien mountains, lower elevation samples in the Horn along with Mt Suswa (Kenya), all Tanzanian populations, Mt Karismbi (Democratic Republic of Congo & Rwanda) and South Africa were all in another group. The Rwandan & DRC samples are both

from Mt Karisimbi, which straddles these nations' borders and these two genotypes were highly similar.

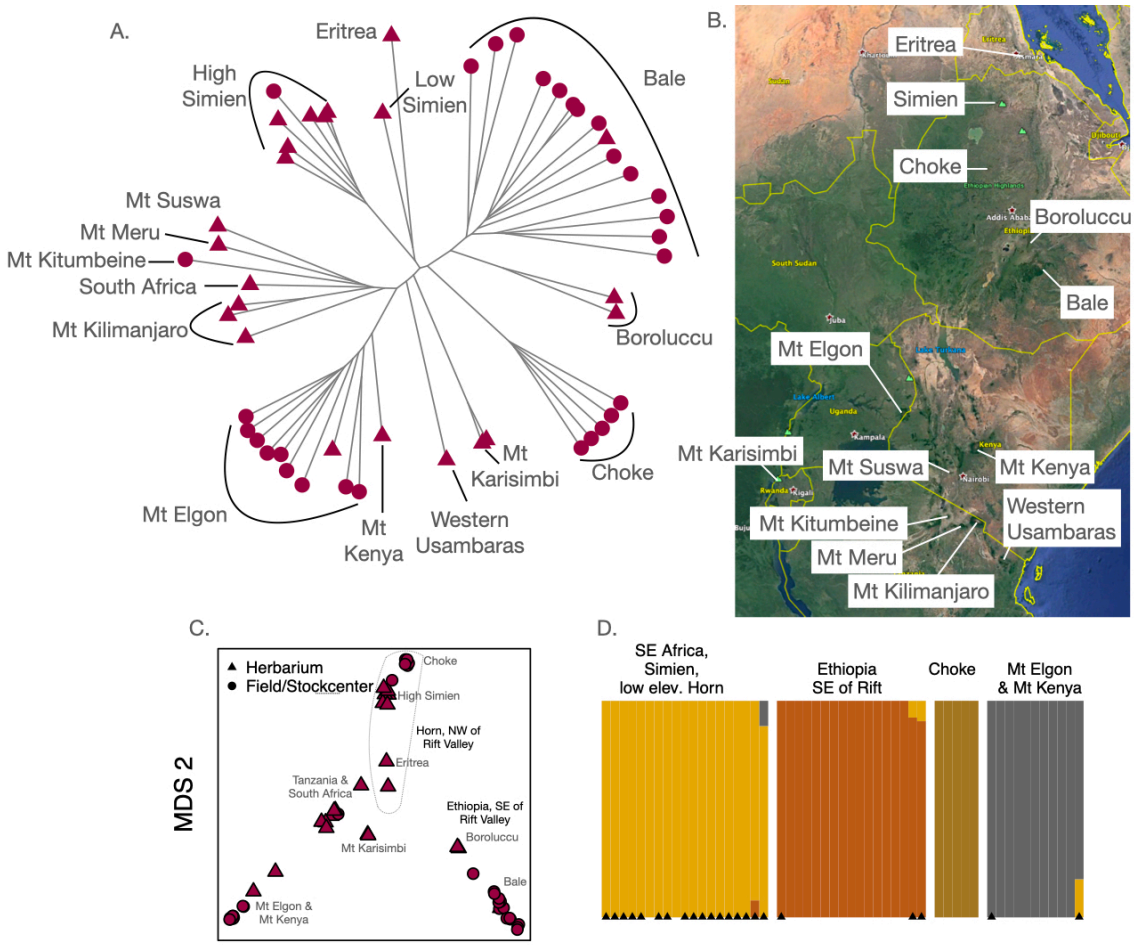


Figure 2. Accessions from East and South Africa, with shapes indicating whether they were herbarium specimens (triangles) or sequenced from fresh tissue from stock center or newly collected accessions (circles). Mountains of origin are labeled on the neighbor-joining tree (A) and map (B). (C) MDS in two dimensions with mountains labeled. (D) Cross validation error lowest for ADMIXTURE K=4 so these results are shown, with triangles at bottom indicating herbarium specimens.

Patterns of population structure to assess the provenance of a historically important specimen

Genotypes from different elevations of the same mountain were typically more closely related than they were to genotypes from similar elevations on different mountains (Figure 2), as is true for most mountain ranges in Eurasia and Africa (Gamba et al., 2024). However, an exception was our oldest specimen in the region, from 1838, collected by Wilhelm G. Schimper at a location in the Simien Mountains. The location, according to Schimper, was "Demerki," (sometimes "Demerkit") a name not apparently in current use, or a corruption of another name. The "Demerki" sample falls in the neighbor-joining tree alongside a 1969 Eritrean sample from 1800 m asl (see "low elevation Simien" in Figure 2). ADMIXTURE on African accessions at

K=5 (only slightly lower cross validation error than the K=4 shown in Figure 2) also assigns the 'Demerki' and Eritrean genotypes 30-40% ancestry from the group of genotypes from Tanzania, many at lower elevations (unlike the higher elevation Simien genotypes which show no such ancestry). Because of this similarity with lower elevation genotypes, we infer that "Demerki" was likely a site in Simien at lower elevation than 3500 m.

Based on the low genetic similarity, the demographic history of these Horn of Africa populations <2500 m asl (including populations known from herbarium specimens from Djibouti & Somalia) (Yim et al., 2023) appears distinct from the mostly higher elevation populations elsewhere in East Africa. Similarly, the 1953 sample from 1750 m asl in the Western Usambaras mountains in Tanzania was highly genetically differentiated from the other Tanzanian populations at higher elevations (Figure 2). Thus, in East African low mountain ranges near the Indian Ocean, Arabidopsis populations seem distinct from populations nearby at much higher elevations, possibly due to a distinct history, such as greater or more recent gene flow among populations, and adaptations to low elevations.

Isolation by distance

We found significant isolation by distance globally including all accessions with coordinates (Figure 3, n = 404, linear mixed models with individual random effects, null AIC= -799279, distance AIC= -853089; general linear model of distance R^2=0.41). For non-relict accessions in Eurasia, we also found significant isolation by distance (n=303, null AIC = -509158, distance AIC= -526778; general linear model of distance R^2=0.36). Isolation by distance was also significant across East Africa (Figure 3, n=50, null AIC= -8998, distance AIC= -9765). Within the two most sampled East African mountains we also found significant isolation by distance: Mt Elgon, Uganda (max. distance = 18 km, n=10, null AIC= -432, distance AIC= -442) and the Bale Mts., Ethiopia (max. Distance = 43 km, n=14, null AIC= -821, distance AIC = -873).

Alonso-Blanco et al. (2016) found relict accessions showed a steeper increase in genetic distance over geographic distance, compared to non-relict accessions, suggesting longer histories of isolation for relicts. We compared East African samples to non-relict Eurasian samples and found a similar pattern: at distances > 250 km, East African samples were more genetically distinct than were non-relicts from Eurasia at the same distance (Figure 3). At a distance of ~700 km we found the greatest genetic distance in East Africa; separating populations from the Great Lakes region from those farther north, in the Horn of Africa. These were more genetically distinct than were accessions from Nepal versus Portugal, >7800 km apart (the most geographically distant non-relict Eurasian samples). These results were also supported by wavelet dissimilarity analysis, which quantified scale-specific genetic turnover among populations. Horn populations showed the greatest turnover at ~50-500 km scales, with Great Lakes populations just slightly lower (Figure 3). By contrast Eurasian non-relicts showed much lower genetic turnover up to ~500 km scales.

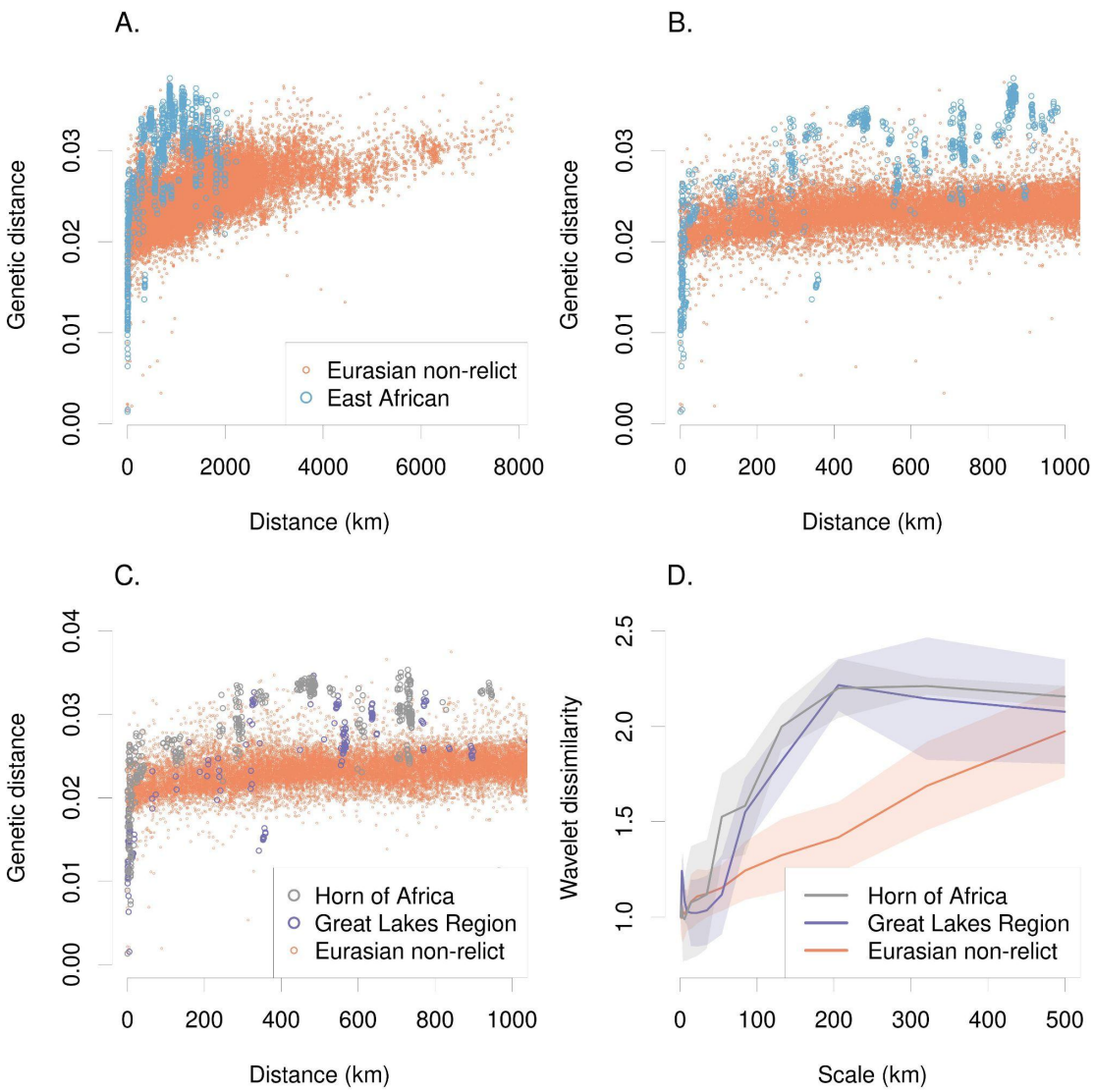


Figure 3. Isolation by distance across the range of Arabidopsis. (A) Pairwise genetic distances within regions, comparing Eurasian non-relicts to East African populations and (B) a truncated version focused on distances <1000 km. (C) Comparing pairwise genetic distance within two of the best sampled regions in East Africa with Eurasian non-relicts. (D) Scale-specific wavelet genetic dissimilarity (line = mean among sampled locations, ribbon = +-1 SD) showed the much greater genetic turnover at scales ~100-500 km for both East African regions compared to Eurasian non-relicts.

Genetic change through time and space

We saw no dramatic changes in ancestry over time. This can be observed in the neighbor-joining tree and ordinations (Figure 1) where herbarium specimens from a given region are most closely related to modern collections from the same region. Additionally, genetic clustering of Eurasian accessions by ADMIXTURE with K=6 showed little change in estimated ancestry assignments over the study period (Figure 4). None of the regionally dominant genetic clusters changed in the

proportion of sample composition over time (Spearman's rank correlation, year versus proportion of most regionally dominant cluster, all p>0.08; also true when using K=3, the K with lowest cross validation error in Eurasia). This indicates there were no major turnover events in population composition over the period of our study.

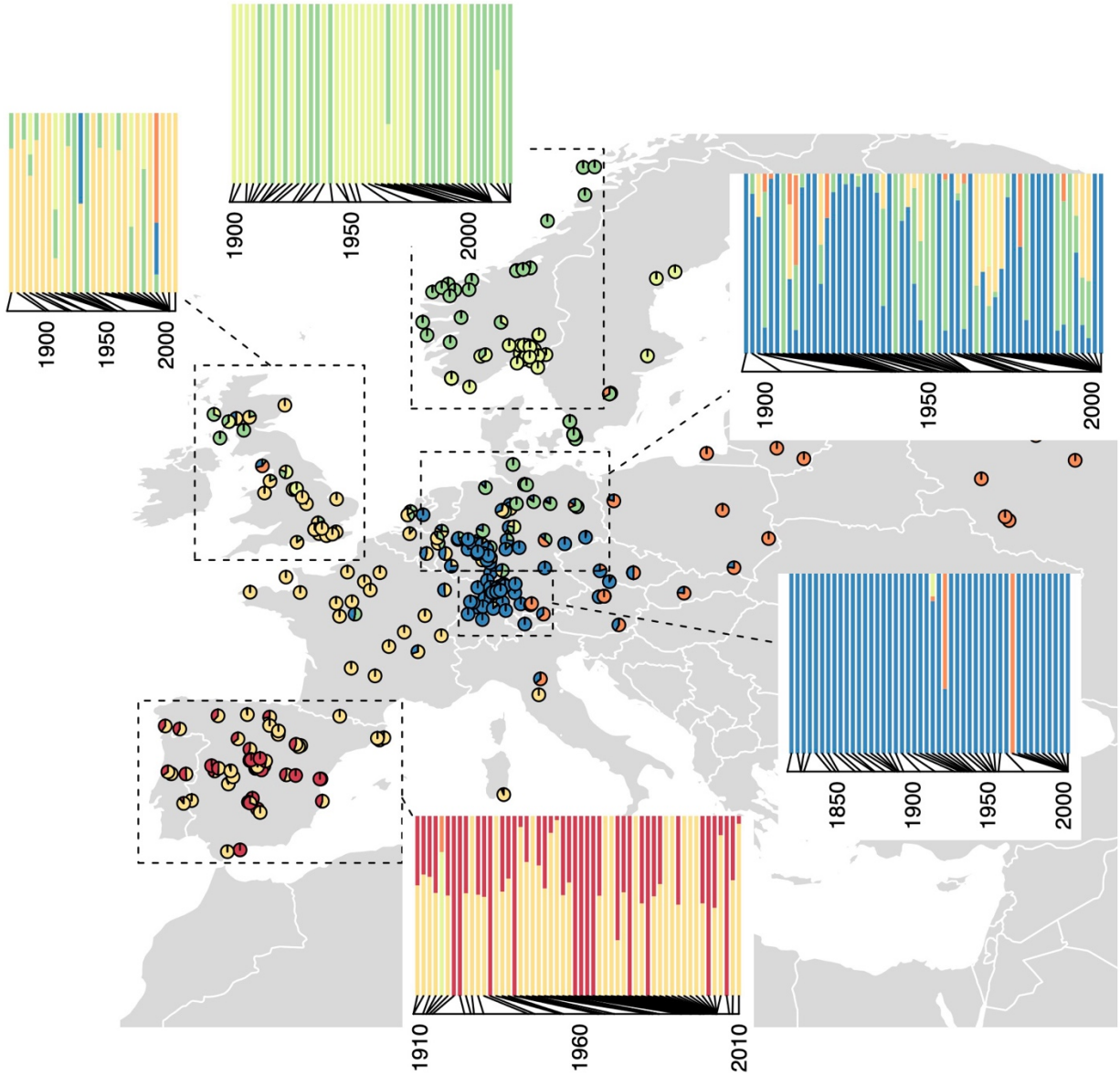


Figure 4. Genetic cluster membership over time (inset barplots) for regional subsets of genotypes (dashed boxes) with individual genotypes shown as pie charts indicated cluster assignment by ADMIXTURE with K=6. Each genetic cluster that was most common in a region showed no significant change in the proportion of assigned ancestry for local genotypes over time (Spearman's rank correlation test, all p>0.08). Herbarium specimens and stock center genotypes are included here.

To examine changes in the distributions of deeper levels of genomic differentiation, we tested for changes in the distribution of relicts in Iberia, where they were at modest frequency (13%, Alonso-Blanco et al. 2016). Alonso-Blanco et al. (2016) noted that Iberian relict accessions in the 1001 Genomes panel tended to be found in landscapes with low anthropogenic disturbance and dry summers. We found in our Iberian herbarium sequences that this pattern was consistent across the 20th century (Figure 5B). Dividing Iberian samples into those from 1908-1968 and those from 1971-2010 (leaving equal sample size groups), and adding the 1001 Genomes genotypes (Alonso-Blanco et al., 2016), we found that relicts were in significantly drier summer climates but with no significant change in this pattern over time (precipitation in warmest quarter two-way ANOVA, relict $p = 0.0358$, time group $p = 0.5111$, relict X time group interaction $p = 0.9458$).

Despite the stability in regionally dominant genetic clusters, we found evidence for modest continuous genetic turnover through time (Figure 5A). Among all regions, AIC favored a model of isolation by distance and isolation by time (n= 404, linear mixed model with distance only AIC = -853089, distance and time AIC = -853491; general linear model with distance only R^2=0.41, distance + time R^2 = 0.42). However, changes in genetic distance with time separation could partly result from changes in our relative sampling of different regions (e.g. 11% of herbarium samples before 1933, the median year, were relict lineages, vs. 32% 1933 and onward) or geographic heterogeneity in temporal change. To avoid this problem, we focused the analysis on some genetic clusters and geographic regions.

Among Eurasian non-relicts, a model with both time and geographic distance was again favored over a geography model only (n=319, linear mixed model with geographic distance only AIC = -577470, geographic and time distance AIC = -577688; general linear model with distance only R^2=0.36, distance + time R^2 = 0.38). To put the genetic turnover through time into a comparable context, we calculated the increase in genetic distance with 100 years separating samples in terms of the geographic distance expected to give the same genetic distance. While at first glance the slope of temporal genetic turnover in Figure 5 may seem slight, the estimated turnover across 100 years in Eurasian non-relicts was equivalent to that occurring over 185 km of geographic distance (for all accessions the figure was 299 km).

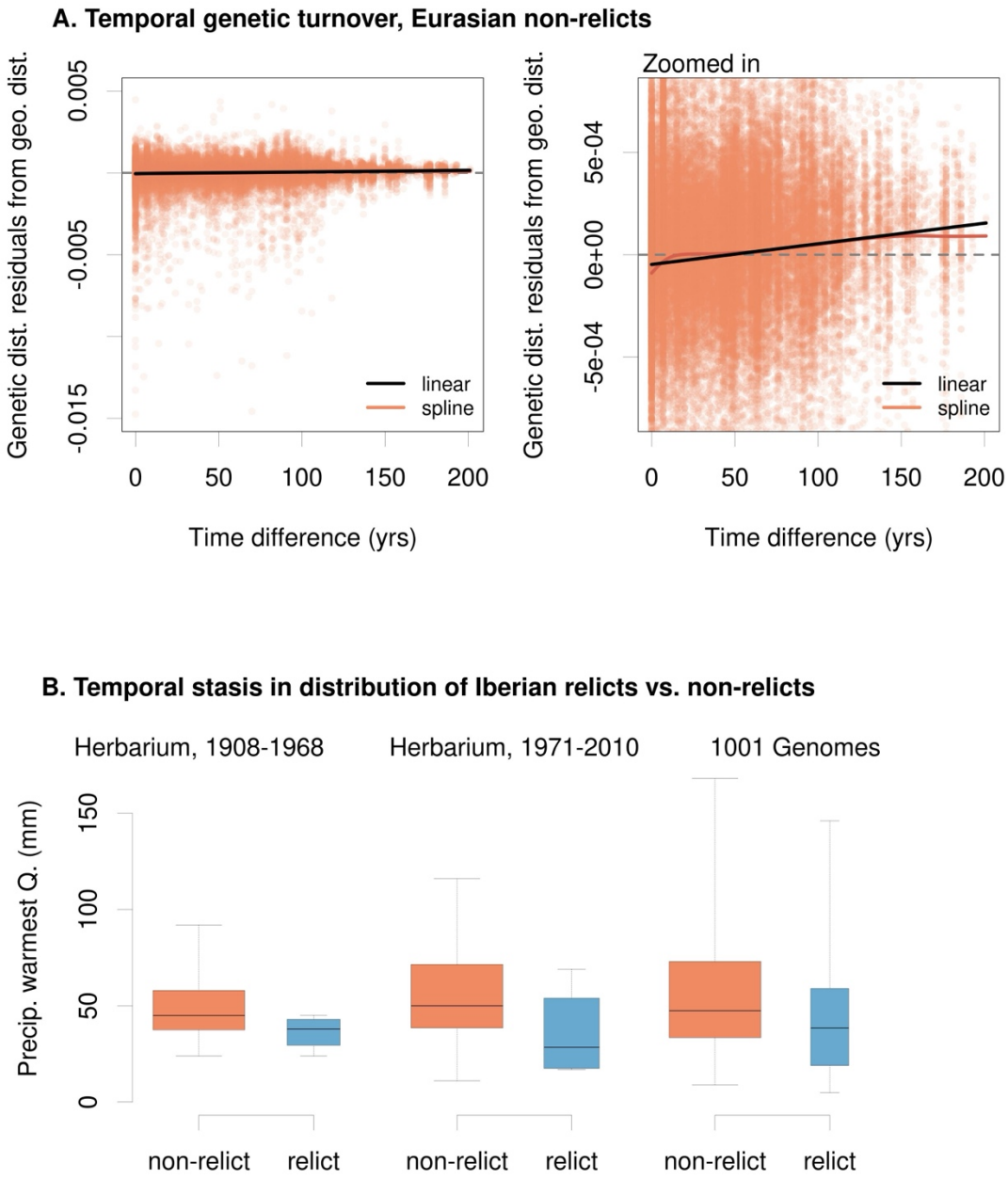


Figure 5. There is modest temporal genetic turnover (A) shown through residuals from a geographic mixed model vs. time difference, with a linear model (black) and spline (red) fitted, while (B) at coarser levels of relatedness the distribution of relicts in drier parts of Iberia remains consistent through time. Note the changing y-axis scales in (A) to show the statistically significant, albeit noisy, temporal turnover. (B) excludes Pyrenees because we did not have these populations sampled from herbaria.

In Germany (n = 94), the model with isolation by distance and time was favored, such that with 100 years separating samples the genetic distance was expected to be of the same magnitude as the genetic distance between samples 70 km apart in geographic distance (mixed model

geography AIC=-50286, geography + time AIC=-50314; general linear model with distance only R^2=0.12, distance + time R^2 = 0.13). For Norwegian (n=44) and non-relict Iberian samples (n=44) a model of isolation-by-distance (but not time) was favored by AIC (Norway: geography AIC=-9496, geography + time AIC=-9496; non-relict Iberia geography AIC=-9835, geography + time AIC=-9833).

Temporal change in allele frequency of putative quantitative trait loci

In general, we found little consistent evidence that putative QTL for several ecologically important traits (GWAS SNPs) were enriched for directional allele frequency change. A list of top SNPs showing directional allele frequency change and nearby genes is included (Table S3 For Eurasian genotypes (n=339; 1,209,375 SNPs) several traits showed nominally significant (α=0.05) enrichments for some combinations of top GWAS SNPs and allele frequency change statistical model, but none of these were significant after controlling for FDR=0.05 (Table S4). The same was true in regional analyses using species-wide GWAS (none significant at FDR=0.05, Tables S5-7). However, because the genetic basis of variation may change among regions, we also tested SNPs from regional GWAS. We found Iberian flowering time (10ºC) SNPs were strongly enriched in temporal allele frequency turnover (n=58 genotypes, top 25 SNPs, FDR=0.05, Table S8, Figure 6). We did not find any enrichment in Fennoscandia flowering time SNPs' turnover in Norway (n=45 genotypes, Table S9). By contrast, German flowering time GWAS SNPs showed significantly reduced change over time, for both 10ºC and 16ºC (n=96 genotypes, Table S10, Figure 6).

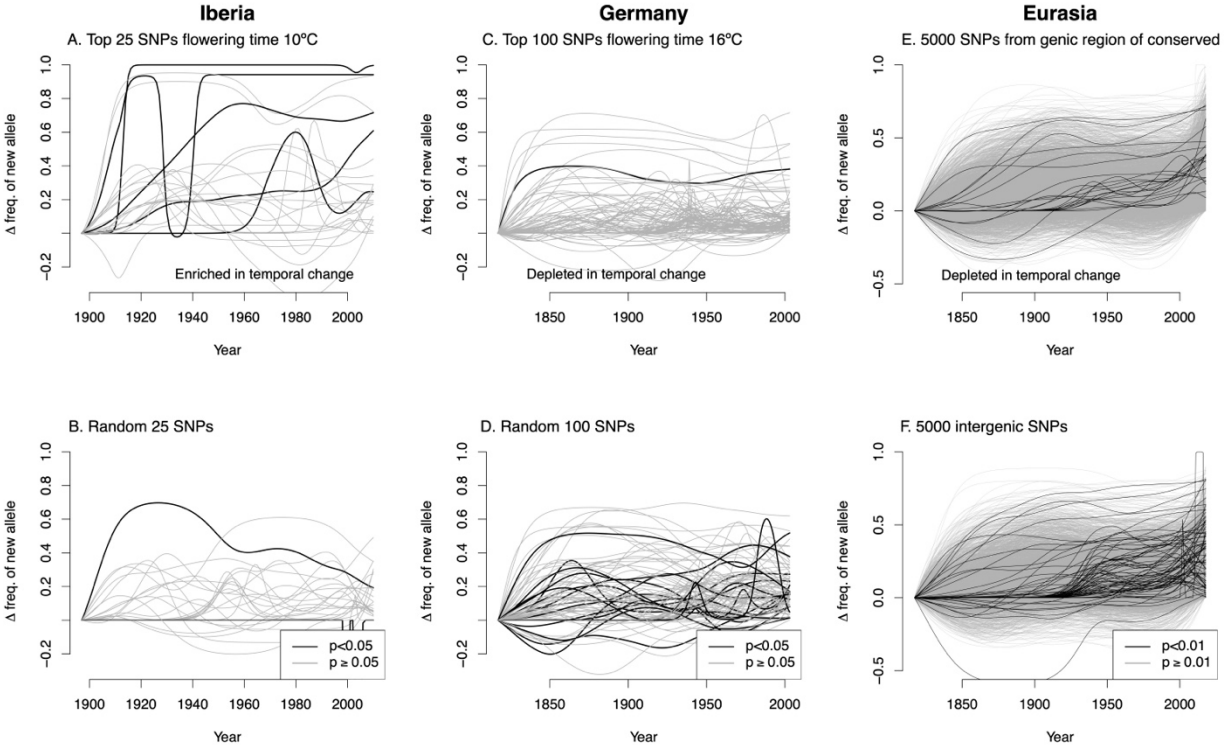


Figure 6. Allele frequency trajectories for different subsets of SNPs (lines) and regions, modeled using generalized additive models (GAMs) for visualization. Allele frequency trajectories are signed so that the y-axis indicates the increase in frequency of the allele more common in recent years (the 'new allele') compared to the beginning of the time series. SNPs with significant temporal allele frequency change in linear mixed models controlling for kinship and geography ($p<0.05$ for A-D, $p<0.01$ for E-F) are shown in black. The random SNPs and the genic/intergenic SNPs shown are single draws from the indicated categories.

Lower directional allele frequency change in genes, especially conserved genes

Next, we tested enrichment of SNPs in genic regions for directional allele frequency change. We found across Eurasia, genic SNPs showed significantly less directional allele frequency change than intergenic SNPs genes (permutation test for 0.01 quantile of logistic and mixed model year effects, $p=0.006$ and $p<0.002$, respectively, example in Figure 6E-F). Additionally, genic SNPs of conserved genes showed significantly less directional allele frequency change than genic SNPs of non-conserved genes (permutation test for 0.01 quantile of logistic and mixed model year effects, $p<0.002$ for both). These patterns were generally consistent within regions (Table S11). These results were consistent when we restricted the analysis to only SNPs with MAF 5-10% and those with MAF > 40% (Table S12).

At the gene level, we also found that the conserved genes showed evidence for different types of selection depending on gene length (Figure S4). Eudicot conserved genes less than 200 AA sometimes had higher directional allele frequency change ($p<0.05$ for some models and regions, Table S13). By contrast, eudicot conserved genes from 200-400 AA were not significantly different from non-conserved genes. Furthermore, genes in amino acid length bins from 400-1,000 showed significantly less directional allele frequency change than null expectations ($p<0.05$ many models and regions, Table S13). We found these patterns were consistent within Germany and Iberia, but not statistically significant in Norway (Table S13).

The forward genetic simulations we conducted were variable in whether they showed genic regions under background selection having greater or less temporal allele frequency changes estimated with our approach (Figure S5). This variability is in contrast to the more consistent effect of background selection of increasing genome-wide temporal autocorrelation in allele frequency (Buffalo & Coop, 2020). In our simulations, genic SNPs (where deleterious mutations could occur; intergenic SNPs were all neutral) showed a significant depletion in temporal allele frequency changes in some simulations, while in other simulations they showed a significant enrichment in temporal allele frequency changes over intergenic SNPs.

Temporal change in allele frequency: genomic prediction

Out-of-sample genomic predictions of flowering time were accurate with 10-fold cross validation (FT10 $r=0.70$, FT16 $r=0.68$). As a positive control, in genome-predicted flowering times we recovered a known cline: Iberia showed strong elevational clines (multiple linear regression; FT10 $t=5.3$, $p<10^{-6}$; FT16 $t=4.8$, $p<10^{-5}$) as observed from common garden

experiments (Gamba et al., 2024; Montesinos-Navarro et al., 2011). When adding a year term to regressions, however, we only found a slightly significant advancement in predicted flowering time at 10ºC over time in Germany (t=-2.5, p=0.0135) but otherwise we did not find any significant change in predicted phenology over time (Table S14). This general lack of clear evolution of flowering time is consistent with the collection date of herbarium specimens, which have not changed in western Europe, in contrast with eastern Europe to central Asia where dates were later in later years, regions not well sampled here (DeLeo et al., 2020).

Discussion

To understand the importance of genetic variation in adaptation and phenotypic variation, it is important to characterize broader patterns of relatedness and genome-wide variation across a species. However, generating comprehensive population genomic surveys of widespread species is challenging due to the logistics of studies at such scales. For example, the Arabidopsis 1001 Genomes project made great progress by resequencing genotypes from across most of the European range (Alonso-Blanco et al., 2016), but greatly under-sampled regions of Asia and Africa (Durvasula et al., 2017; Hsu et al., 2019; Roy, 2018; Zou et al., 2017). Historically, research on Arabidopsis natural genetic variation has had a strong European bias (Hoffmann, 2002; Koornneef et al., 2004). However, African populations in particular show high genetic divergence from each other and from Eurasian populations (Durvasula et al., 2017; Gamba et al., 2024) and here we add substantially to the known diversity and its distribution in Africa. Additionally, high elevation African populations (samples in this study mostly from >4000 m asl) occupy outlier environments (Gamba et al., 2024), suggesting a great resource of undiscovered locally adapted genes and traits. Our use of museum sequencing allowed us to rapidly fill in important gaps in unstudied populations of Arabidopsis.

Temporal change in population size and allele frequency has been often inferred indirectly using current snapshots of population genetic composition (Nielsen et al., 2005; Schiffels & Durbin, 2014; Slatkin & Hudson, 1991). An exception is the rapidly growing field of ancient human genetics, which is starting to reveal the sometimes dramatic changes in population structure and ancestry over the last several thousand years (Allentoft et al., 2024; LaPolice et al., 2024; Olalde et al., 2018). However, there are few similar studies in other species outside of agricultural species (Kreiner et al., 2022; Smith et al., 2019; Swarts et al., 2017; Verdugo et al., 2019). By contrast, museums house millions of wild specimens covering centuries of sampling, providing a major resource to learn about evolution in response to contemporary environmental changes (Burbano & Gutaker, 2023; Lopez et al., 2020). Here, we used Arabidopsis specimen sequences to demonstrate temporal population genomic turnover, although we see little evidence of dramatic turnover in ancestry of the type seen in humans. Nevertheless, this turnover is heterogeneous across the genome, identifiably so, despite the limited outcrossing in this species.

706 We showed how many genes conserved over many millions of years show distinct patterns of
707 turnover during the study period (1817-2018).
708
709 *Global population structure*
710 By sequencing museum specimens, we were able to fill in major gaps in global population
711 structure especially in Africa, and important additions in Norway, Sardinia, and Nepal. We found
712 that east and southern African Arabidopsis represent a group of populations highly distinct from
713 those in North Africa and Eurasia, with major divergence between populations from different
714 mountains across east Africa (Gamba et al., 2024). We showed that the level of genetic
715 divergence between African populations increases with spatial distance at a rate much greater
716 than that seen in Eurasia. This is consistent with the hypothesis that these populations have been
717 spread across East Africa for a long period of time and represent a potential origin of the
718 *Arabidopsis thaliana* species (Durvasula et al., 2017; Fulgione & Hancock, 2018).
719 The high divergence between mountains was especially true for high elevation
720 populations in the Horn of Africa where the East African Rift Valley separates the two groups of
721 greatest divergence, consistent with findings in *Trifolium* (Wondimu et al., 2014). Furthermore,
722 the apparently high diversity of Arabidopsis from Bale (Figure 2) is consistent with these
723 mountains also being a center of diversity for *Trifolium* and *Carduus* (Wondimu et al., 2014)
724 perhaps because of the topography of Bale: a high elevation plateau with the widest range of
725 Afroalpine habitat in the region. Within some of the relatively lower elevation populations in
726 southern Kenya and northern Tanzania, we found less divergence suggesting more recent gene
727 flow between populations. The separation of these genotypes from those nearby on Mt Kenya &
728 Mt Elgon is consistent with findings in *Dendrosenecio* in this region (Tusiime et al., 2020). Our
729 study helps set the stage for future investigation of adaptive and phenotypic variation in the
730 diverse populations of East Africa.
731 The sample collected by W.G. Schimper from "Demerki" Ethiopia likely has a low
732 elevation (<3000 m) provenance. Sources have disputed the location of the historically important
733 "Demerki" specimens, as being Däräsge Maryam at 2990 m (Dorothea McEwan pers. comm.) or
734 alternatively at 3500 m (Gillett, 1972), and many have georeferenced it at the peak of the massif
735 (4400 m asl), though the specimen itself does not indicate elevation. The location is of interest
736 given the extensive collections (>600 Schimper specimens labeled "Demerki" in GBIF),
737 including many types, made by Schimper in a time period from which few other specimens
738 survive (Gillett, 1972).
739 We also added a genome from a large Mediterranean island previously unstudied:
740 Sardinia, which was a mix of Eurasian non-relict and Iberian/Moroccan relict-like ancestries.
741 This relict ancestry is consistent with Sardinia being a refugium during glacial periods as
742 recently indicated in our paleoclimatic projection of Arabidopsis distribution (Yim et al., 2023)
743 and with previous findings that relict ancestry is greater in low elevations in the Mediterranean
744 (Alonso-Blanco et al., 2016). Thus, it is likely that with greater sampling additional diversity of
745 relict ancestry will be found across the Mediterranean. Previous studies had included very few

sequences of Norwegian genotypes. Here we show that they are a group of relatively similar genotypes with similar to others from Northern Europe including Sweden, Britain, and Germany

Temporal population turnover

One of the most powerful benefits of sequencing museum specimens is to be able to obtain longitudinal samples of populations through time at multiple locations and populations. Some recent examples of major turnover in ancestry include ancient humans in Europe (Allentoft et al., 2024; LaPolice et al., 2024; Olalde et al., 2018), *Centaurea* plants under expansion of a polyploid lineage (Rosche et al., 2024), and domesticated pigs following their introduction into Europe (Frantz et al., 2019). Notably, these examples either can travel large distances under locomotion (mammals) or wind and animal attachment (*Centaurea*). Indeed, using modeling based on temporal autocorrelation in allele frequency change (Simon & Coop, 2024) estimated that the great majority of human genetic turnover in European populations was attributable to gene flow, with some contribution from drift. In non-relict Eurasian Arabidopsis, we found more modest but detectable turnover through a period of 100 years equivalent to that estimated to occur between populations 185 km apart at the same point in time. This magnitude may seem high given Arabidopsis is primarily self-fertilizing and with little mechanism of seed dispersal. However, there is water dispersal by some genotypes using mucilage (Saez-Aguayo et al., 2014), and hitchhiking on humans, livestock, large mammals, and birds could play a role. Furthermore, individual Arabidopsis populations might show strong drift in landscapes where they colonize recently disturbed patches and potentially go extinct as succession proceeds (Baron et al., 2015; Lorts & Lasky, 2020). Further work is required to understand processes contributing to population genomic turnover. Additionally, a weakness of museum specimens is that they were not specifically sampled for this type of study, and so the heterogeneity in sampling through time may limit our ability to detect population genetic changes. Each region showed distinct spatiotemporal gaps in sampling (Figure 4), potentially limiting our ability to accurately characterize allele frequency and population structure dynamics. Furthermore, the need to remove only small parts of tissue from specimens and consequent relatively intensive lab work limits the ability to scale up sample sizes.

Locus-specific turnover in allele frequency

Outside of experimental evolution with microbes, there have been relatively few previous time-series studies dissecting patterns of directional allele frequency change over many generations at specific loci across the genome (Akbari et al., 2024; Exposito-Alonso, Vasseur, et al., 2018; Lang et al., 2024; Le et al., 2022; Mathieson et al., 2015). Our sequence data allowed us to test for evidence of non-random patterns in allele frequency change. In general, we did not see much evidence for changes in the frequency of GWAS SNPs for traits, such as flowering time, d13C, and seed dormancy, known to be important to local environmental adaptation in Arabidopsis (and presumably under shifting selection with environmental change) (Dittberner et al., 2018; Gamba et al., 2024; Martínez-Berdeja et al., 2020; Stinchcombe et al., 2004).

There may be multiple reasons we did not see more such evidence of adaptive temporal allele frequency shifts. GWAS appear to work well in many Arabidopsis traits (Alonso-Blanco et al., 2016; Atwell et al., 2010; Baxter et al., 2010) so it seems likely the GWAS results used contained true QTL, although false positives and negatives may remain. Limited recombination in Arabidopsis may limit resolution to see temporal changes across the genome, but between many individual Arabidopsis populations there are many past recombination events that limit LD and provide resolution for GWAS. Furthermore, whole genome predictions on flowering time at two temperatures did not show genome-wide shifts in these traits. One plausible explanation may be that while traits of plants in nature have changed over time (DeLeo et al., 2020), the shifts in selection and genetic architecture of traits may be sufficiently heterogeneous among populations to limit our ability to see evidence of adaptation across regions and continents (Gamba et al., 2024; Lopez-Arboleda et al., 2021). Finally, the method of looking at enrichment in GWAS SNPs for temporal allele frequency shifts may be of limited power. Our strategy for testing for adaptation at these loci is similar to that of (Akbari et al., 2024), although we do not rely on determination of a significance threshold for calling individual loci as under selection. In Arabidopsis, (Lang et al., 2024) developed a polygenic score for individual stomatal genes to show shifts in allele frequency suggesting shifts to decreased stomatal density, a potential strategy to be implemented in the future for more traits.

We did identify the top flowering time GWAS SNPs in Iberia as enriched for shifts in allele frequency over time, potentially signifying changes in selection. Flowering time is a trait often under shifting selection among environments in many species (Ågren et al., 2017; Munguía-Rosas et al., 2011) and has been identified as a trait showing temporal change following environmental events (Franks et al., 2007). However, it is hard to identify the specific change in selection pressures potentially acting in this case. While climate has changed in Iberia over this period (Corte-Real et al., 1998; Esteban-Parra et al., 1998), land use may also influence selection on flowering time in Arabidopsis. And there has been dramatic land use change, such as the dramatic abandonment of farmland in the mountains of Spain over the 20th century, >90% in some regions (Lasanta et al., 2017).

The loci showing the strongest allele frequency turnover identified several genes with known roles in environmental response (Table S3). For example, the second SNP with the most significant change in Eurasia was closest to plant U-box E3 ligase 12 (PUB12) which plays a role in immunity and abscisic acid signaling (Kong et al., 2015; Yamaguchi et al., 2017). The top SNP in Germany was *EXTRA LARGE G PROTEIN 3* (*XLG3*) which is involved in pathogen associated molecular pattern signaling (Y. Wang et al., 2023). The number 5 SNP in Norway is in a cluster of *SMALL AUXIN UP RNA* (*SAUR26/27/28*) genes that regulate temperature responsive growth affected by known *cis*-regulatory variants and with allele frequency clines along temperature gradients (Z. Wang et al., 2019). These loci merit further investigation to determine the effects of this allelic variation.

Purifying and background selection and temporal shifts in allele frequency

We found that genic regions and conserved genes with longer amino acid sequences showed low temporal allele frequency change over time, potentially due to purifying selection and background selection on linked sites. The former pattern stands in contrast to the theoretical genome-wide results of (Buffalo & Coop, 2020) who found that background selection causes temporal autocorrelation in allele frequency change across the genome. Regardless of the genome-wide pattern, in Arabidopsis there are clearly differences among genomic contexts in their temporal dynamics. It is notable that genic SNPs are enriched in signals of local adaptation in Arabidopsis (Hancock et al., 2011; Lasky et al., 2012, 2018), opposite to what we found for temporal dynamics, perhaps because temporal allele frequency changes across the genome are dominated by background selection and not adaptation to changing environments.

One hypothesis to explain our finding is the following. Alleles at appreciable frequency for our analysis (i.e. MAF>5%) in conserved genes are neutral, background selection limits their diversity within populations, but many of these neutral (e.g. synonymous) variants are fixed between populations. Over time, background selection limits change in the locally fixed alleles at these loci but the allele frequency in our analysis (due to variation between populations) remains the same. Thus low temporal allele frequency turnover in conserved genes. In less conserved (shorter) genes, variants of appreciable frequency are neutral or slightly deleterious (leading to consistent decrease in frequency as predicted by Buffalo & Coop 2020) or under positive selection potentially due to environmental change (leading to consistent increase in frequency). Relatedly, (Cvijović et al., 2018) showed in an asexual model how background selection leads to distinct trajectories of neutral allele frequency depending on initial frequency, and so differences among loci in the site frequency spectrum can lead to distinct temporal trajectories. However, more theory is required to develop predictions for differences in allele frequency trajectories across genomes subject to spatially variable purifying, background, and positive selection. The fact that conserved Arabidopsis genes were depleted in temporal allele frequency changes suggests Arabidopsis demographic and genome parameters may be in a space where background selection reduces temporal allele frequency change. Other species with different demography and genomes may show distinct patterns.

Conclusions

Natural history collections contain a vast wealth of samples that can generate insights into how organisms and populations are changing over time. However, in recent years their survival has been threatened as institutions cut support. Our results highlight the continued vitality of these collections and potential benefits to directly revealing evolutionary change through time. Arabidopsis is changing as likely all species, and future work will dissect ecological and evolutionary mechanisms driving this change.

Acknowledgements

This research was made possible by assistance from several curators and museums, specifically Real Jardín Botánico, Kew Gardens, Komarov Botanical Garden, Oslo Museum of Natural

History, Stuttgart State Museum of Natural History, and Herbarium Tubingense. Plant material was exported from Uganda with permission of The Ministry of Agriculture, Animal Industry and Fisheries' Plant Quarantine and Inspection Services, permit UQIS 4414/93/PC (E). Material was exported from Ethiopia with permission of the Ethiopian Biodiversity Institute, Ref. no. EBE71/7065/2018. Material was imported to the USA under USDA APHIS permits P37-17-01651 and P37-18-00230. Sequencing of Norwegian accessions was supported by NSF DEB-1743273 to CGO and JKM. We thank D. Weigel and the Max Planck Society for funding the sequencing of German herbarium specimens. LL is partially supported by California State University, San Bernardino, and by National Science Foundation award BIO-BRC 2217793. Funding was provided by National Institutes of Health grant R35GM138300 to JRL.

Author contributions

The study was conceived and designed by LL and JRL. LL, PL, AH, MY, PW, JK, CB, JRL collected data. LL, PL, SL, YX, EC, DG, and JRL analyzed data. LL and JRL led writing. All authors contributed to writing.

References

Ågren, J., Oakley, C. G., Lundemo, S., & Schemske, D. W. (2017). Adaptive divergence in flowering time among natural populations of Arabidopsis thaliana: Estimates of selection and QTL mapping. *Evolution*, *71*(3), 550–564. https://doi.org/10.1111/evo.13126

Ågren, J., & Schemske, D. W. (2012). Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytologist*, *194*(4), 1112–1122. https://doi.org/10.1111/j.1469-8137.2012.04112.x

Akbari, A., Barton, A. R., Gazal, S., Li, Z., Kariminejad, M., Perry, A., Zeng, Y., Mittnik, A., Patterson, N., Mah, M., Zhou, X., Price, A. L., Lander, E. S., Pinhasi, R., Rohland, N., Mallick, S., & Reich, D. (2024). *Pervasive findings of directional selection realize the promise of ancient DNA to elucidate human adaptation* (p. 2024.09.14.613021). bioRxiv. https://doi.org/10.1101/2024.09.14.613021

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655. https://doi.org/10.1101/gr.094052.109

Allentoft, M. E., Sikora, M., Fischer, A., Sjögren, K.-G., Ingason, A., Macleod, R., Rosengren, A., Schulz Paulsson, B., Jørkov, M. L. S., Novosolov, M., Stenderup, J., Price, T. D., Fischer Mortensen, M., Nielsen, A. B., Ulfeldt Hede, M., Sørensen, L., Nielsen, P. O., Rasmussen, P., Jensen, T. Z. T., … Willerslev, E. (2024). 100 ancient genomes show repeated population turnovers in Neolithic Denmark. *Nature*, *625*(7994), 329–337. https://doi.org/10.1038/s41586-023-06862-3

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., Cao, J., Chae, E., Dezwaan, T. M., Ding, W., Ecker, J. R., Exposito-Alonso, M., Farlow, A., Fitz, J., Gan, X., Grimm, D. G., Hancock, A. M., Henz, S. R., Holm, S., … Zhou, X. (2016).

1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, *166*, 481–491. https://doi.org/10.1016/j.cell.2016.05.063

Anstett, D. N., Anstett, J., Sheth, S. N., Moxley, D., Jahani, M., Huang, K., Todesco, M., Jordan, R., Lazaro-Guevara, J. M., Rieseberg, L. H., & Angert, A. L. (2024). *Evolutionary rescue during extreme drought* (p. 2024.10.24.619808). bioRxiv. https://doi.org/10.1101/2024.10.24.619808

Antonio, M. L., Weiß, C. L., Gao, Z., Sawyer, S., Oberreiter, V., Moots, H. M., Spence, J. P., Cheronet, O., Zagorc, B., Praxmarer, E., Özdoğan, K. T., Demetz, L., Gelabert, P., Fernandes, D., Lucci, M., Alihodžić, T., Amrani, S., Avetisyan, P., Baillif-Ducros, C., … Pritchard, J. K. (2024). Stable population structure in Europe since the Iron Age, despite high mobility. *eLife*, *13*, e79714. https://doi.org/10.7554/eLife.79714

Arteaga, N., Méndez-Vigo, B., Fuster-Pons, A., Savic, M., Murillo-Sánchez, A., Picó, F. X., & Alonso-Blanco, C. (2022). Differential environmental and genomic architectures shape the natural diversity for trichome patterning and morphology in different Arabidopsis organs. *Plant, Cell & Environment*, *45*(10), 3018–3035. https://doi.org/10.1111/pce.14308

Atwell, S., Huang, Y. S., Vilhjalmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Muliyati, N. W., Zhang, X., Amer, M. A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, *465*(7298), 627–631. https://doi.org/10.1038/nature08800

Baron, E., Richirt, J., Villoutreix, R., Amsellem, L., & Roux, F. (2015). The genetics of intra- and interspecific competitive response and effect in a local population of an annual plant species. *Functional Ecology*, *29*(10), 1361–1370. https://doi.org/10.1111/1365-2435.12436

Baxter, I., Brazelton, J. N., Yu, D., Huang, Y. S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J. O., Nordborg, M., Vitek, O., & Salt, D. E. (2010). A Coastal Cline in Sodium Accumulation in Arabidopsis thaliana Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. *PLOS Genetics*, *6*(11), e1001193. https://doi.org/10.1371/journal.pgen.1001193

Beck, J. B., Schmuths, H., & Schaal, B. A. (2008). Native range genetic variation in Arabidopsis thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. *Molecular Ecology*, *17*(3), 902–915. https://doi.org/10.1111/j.1365-294X.2007.03615.x

Bellis, E. S., Kelly, E. A., Lorts, C. M., Gao, H., DeLeo, V. L., Rouhan, G., Budden, A., Bhaskara, G. B., Hu, Z., Muscarella, R., Timko, M. P., Nebie, B., Runo, S. M., Chilcoat, N. D., Juenger, T. E., Morris, G. P., dePamphilis, C. W., & Lasky, J. R. (2020). Genomics of sorghum local adaptation to a parasitic plant. *Proceedings of the National Academy of Sciences*, *117*(8), 4243–4251. https://doi.org/10.1073/pnas.1908707117

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genet*, *10*(11), e1004775. https://doi.org/10.1371/journal.pgen.1004775

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., Moritz, C., & Good, J. M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLOS Genetics*, *15*(5), e1008119. https://doi.org/10.1371/journal.pgen.1008119

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Buffalo, V., & Coop, G. (2020). Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proceedings of the National Academy of Sciences*, *117*(34), 20672–20680. https://doi.org/10.1073/pnas.1919039117

Burbano, H. A., & Gutaker, R. M. (2023). Ancient DNA genomics and the renaissance of herbaria. *Science*, *382*(6666), 59–63. https://doi.org/10.1126/science.adi1180

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Corte-Real, J., Qian, B., & Xu, H. (1998). Regional climate change in Portugal: Precipitation variability associated with large-scale atmospheric circulation. *International Journal of Climatology*, *18*(6), 619–635. https://doi.org/10.1002/(SICI)1097-0088(199805)18:6<619::AID-JOC271>3.0.CO;2-T

Covarrubias-Pazaran, G. (2016). Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *PLOS ONE*, *11*(6), e0156744. https://doi.org/10.1371/journal.pone.0156744

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics*, *209*(4), 1235–1278. https://doi.org/10.1534/genetics.118.301058

DeLeo, V. L., Menge, D. N. L., Hanks, E. M., Juenger, T. E., & Lasky, J. R. (2020). Effects of two centuries of global environmental variation on phenology and physiology of Arabidopsis thaliana. *Global Change Biology*, *26*(2), 523–538. https://doi.org/10.1111/gcb.14880

Dittberner, H., Korte, A., Mettler-Altmann, T., Weber, A. P. M., Monroe, G., & Meaux, J. de. (2018). Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana. *Molecular Ecology*, *27*(20), 4052–4065. https://doi.org/10.1111/mec.14838

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., & Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 201616736. https://doi.org/10.1073/pnas.1616736114

Elfarargi, A. F., Gilbault, E., Döring, N., Neto, C., Fulgione, A., Weber, A. P. M., Loudet, O., & Hancock, A. M. (2023). Genomic Basis of Adaptation to a Novel Precipitation Regime. *Molecular Biology and Evolution*, *40*(3), msad031. https://doi.org/10.1093/molbev/msad031

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, *4*, 250–255.

Esteban-Parra, M. J., Rodrigo, F. S., & Castro-Diez, Y. (1998). Spatial and temporal patterns of precipitation in Spain for the period 1880–1992. *International Journal of Climatology*, *18*(14), 1557–1574. https://doi.org/10.1002/(SICI)1097-0088(19981130)18:14<1557::AID-JOC328>3.0.CO;2-J

Exposito-Alonso, M., Becker, C., Schuenemann, V. J., Reiter, E., Setzer, C., Slovak, R., Brachi, B., Hagmann, J., Grimm, D. G., Chen, J., Busch, W., Bergelson, J., Ness, R. W., Krause, J., Burbano, H. A., & Weigel, D. (2018). The rate and potential relevance of new mutations in a colonizing plant lineage. *PLOS Genetics*, *14*(2), e1007155. https://doi.org/10.1371/journal.pgen.1007155

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018). Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nature Ecology & Evolution*, *2*(2), 352–358. https://doi.org/10.1038/s41559-017-0423-0

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, *24*(14), 3529–3545. https://doi.org/10.1111/mec.13226

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A Map of Local Adaptation in Arabidopsis thaliana. *Science*, *334*(6052), 86–89.

Franks, S. J., Kane, N. C., O'Hara, N. B., Tittes, S., & Rest, J. S. (2016). Rapid genome-wide evolution in Brassica rapa populations following drought revealed by sequencing of ancestral and descendant gene pools. *Molecular Ecology*, n/a-n/a. https://doi.org/10.1111/mec.13615

Franks, S. J., Sim, S., & Weis, A. E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences*, *104*(4), 1278–1282. https://doi.org/10.1073/pnas.0608379104

Frantz, L. A. F., Haile, J., Lin, A. T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A., Mullin, V. E., Daly, K. G., Battista, V. M., Price, M., Gron, K. J., Alexandri, P., Arbogast, R.-M., Arbuckle, B., Bălăşescu, A., Barnett, R., Bartosiewicz, L., … Larson, G. (2019). Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. *Proceedings of the National Academy of Sciences*, *116*(35), 17231–17238. https://doi.org/10.1073/pnas.1901169116

Fulgione, A., & Hancock, A. M. (2018). Archaic lineages broaden our view on the history of Arabidopsis thaliana. *New Phytologist*, *219*(4), 1194–1198. https://doi.org/10.1111/nph.15244

Fulgione, A., Neto, C., Elfarargi, A. F., Tergemina, E., Ansari, S., Göktay, M., Dinis, H., Döring, N., Flood, P. J., Rodriguez-Pacheco, S., Walden, N., Koch, M. A., Roux, F., Hermisson, J., & Hancock, A. M. (2022). Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-28800-z

Gamba, D., Lorts, C. M., Haile, A., Sahay, S., Lopez, L., Xia, T., Takou, M., Kulesza, E., Elango, D., Kerby, J., Yifru, M., Bulafu, C. E., Wondimu, T., Glowacka, K., & Lasky, J. R. (2024). The genomics and physiology of abiotic stressors associated with global elevational

gradients in Arabidopsis thaliana. *New Phytologist*, *244*, 2062–2077. https://doi.org/10.1111/nph.20138

Gauthier, J., Pajkovic, M., Neuenschwander, S., Kaila, L., Schmid, S., Orlando, L., & Alvarez, N. (2020). Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Molecular Ecology Resources*, *20*(5), 1191–1205. https://doi.org/10.1111/1755-0998.13167

Gienapp, P., Fior, S., Guillaume, F., Lasky, J. R., Sork, V. L., & Csilléry, K. (2017). Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends in Ecology & Evolution*, *32*(12), 897–908. https://doi.org/10.1016/j.tree.2017.09.004

Gillett, J. B. (1972). W. G. Schimper's Botanical Collecting Localities in Ethiopia. *Kew Bulletin*, *27*(1), 115–128. https://doi.org/10.2307/4117875

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Hamann, E., Weis, A. E., & Franks, S. J. (2018). Two decades of evolutionary changes in Brassica rapa in response to fluctuations in precipitation and severe drought. *Evolution; International Journal of Organic Evolution*, *72*(12), 2682–2696. https://doi.org/10.1111/evo.13631

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F., & Bergelson, J. (2011). Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science*, *334*(6052), 83–86.

Hayward, L. K., & Sella, G. (2022). Polygenic adaptation after a sudden change in environment. *eLife*, *11*, e66697. https://doi.org/10.7554/eLife.66697

Hereford, J. (2009). A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *The American Naturalist*, *173*(5), 579–588.

Hesen, V., Boele, Y., Holmer, R., Boesten, R., Wijfjes, R., Aarts, M. G. M., Putten, W. H. van der, Scheres, B., & Willemsen, V. (2024). *Steep logarithmic increase of genetic variation in natural Arabidopsis thaliana accessions across geographic scales* (p. 2024.04.26.591275). bioRxiv. https://doi.org/10.1101/2024.04.26.591275

Hill, W. G., & Kirkpatrick, M. (2010). What Animal Breeding Has Taught Us about Evolution. *Annual Review of Ecology, Evolution, and Systematics*, *41*(Volume 41, 2010), 1–19. https://doi.org/10.1146/annurev-ecolsys-102209-144728

Hoffmann, M. H. (2002). Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). *Journal of Biogeography*, *29*(1), 125–134.

Höllinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLOS Genetics*, *15*(3), e1008035. https://doi.org/10.1371/journal.pgen.1008035

Hsu, C.-W., Lo, C.-Y., & Lee, C.-R. (2019). On the postglacial spread of human commensal Arabidopsis thaliana: Journey to the East. *New Phytologist*, *222*(3), 1447–1457. https://doi.org/10.1111/nph.15682

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kachroo, A. H., Laurent, J. M., Yellman, C. M., Meyer, A. G., Wilke, C. O., & Marcotte, E. M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science (New York, N.Y.)*, *348*(6237), 921. https://doi.org/10.1126/science.aaa0769

Karasov, T. L., Kniskern, J. M., Gao, L., DeYoung, B. J., Ding, J., Dubiella, U., Lastra, R. O., Nallu, S., Roux, F., Innes, R. W., Barrett, L. G., Hudson, R. R., & Bergelson, J. (2014).

The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature*, *512*(7515), 436–440. https://doi.org/10.1038/nature13439

Kennett, D. J., Lipson, M., Prufer, K. M., Mora-Marín, D., George, R. J., Rohland, N., Robinson, M., Trask, W. R., Edgar, H. H. J., Hill, E. C., Ray, E. E., Lynch, P., Moes, E., O'Donnell, L., Harper, T. K., Kate, E. J., Ramos, J., Morris, J., Gutierrez, S. M., … Reich, D. (2022). South-to-north migration preceded the advent of intensive farming in the Maya region. *Nature Communications*, *13*(1), 1530. https://doi.org/10.1038/s41467-022-29158-y

Kistler, L., Maezumi, S. Y., Gregorio de Souza, J., Przelomska, N. A. S., Malaquias Costa, F., Smith, O., Loiselle, H., Ramos-Madrigal, J., Wales, N., Ribeiro, E. R., Morrison, R. R., Grimaldo, C., Prous, A. P., Arriaza, B., Gilbert, M. T. P., de Oliveira Freitas, F., & Allaby, R. G. (2018). Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science (New York, N.Y.)*, *362*(6420), 1309–1313. https://doi.org/10.1126/science.aav0207

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., & Ding, L. (2009). VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, *25*(17), 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

Kong, L., Cheng, J., Zhu, Y., Ding, Y., Meng, J., Chen, Z., Xie, Q., Guo, Y., Li, J., Yang, S., & Gong, Z. (2015). Degradation of the ABA co-receptor ABI1 by PUB12/13 U-box E3 ligases. *Nature Communications*, *6*(1), 8630. https://doi.org/10.1038/ncomms9630

Koornneef, M., Alonso-Blanco, C., & Vreugdenhil, D. (2004). Naturally occurring genetic variation in Arabidopsis thaliana. *Annual Review of Plant Biology*, *55*, 141–172. https://doi.org/10.1146/annurev.arplant.55.031903.141605

Kreiner, J. M., Latorre, S. M., Burbano, H. A., Stinchcombe, J. R., Otto, S. P., Weigel, D., & Wright, S. I. (2022). Rapid weed adaptation and range expansion in response to

agriculture over the past two centuries. *Science (New York, N.Y.)*, *378*(6624), 1079–1085. https://doi.org/10.1126/science.abo7293

Lang, P. L. M., Erberich, J. M., Lopez, L., Weiß, C. L., Fung, H. F., Latorre, S. M., Lasky, J. R., Burbano, H. A., Expósito-Alonso, M., & Bergmann, D. C. (2024). Century-long timelines of herbarium genomes predict plant stomatal response to climate change. *Nature Ecology & Evolution*, 1–13. https://doi.org/10.1038/s41559-024-02481-x

LaPolice, T. M., Williams, M. P., & Huber, C. D. (2024). *The European Neolithic Expansion: A Model Revealing Intense Assortative Mating and Restricted Cultural Transmission* (p. 2024.04.29.591653). bioRxiv. https://doi.org/10.1101/2024.04.29.591653

Lasanta, T., Arnáez, J., Pascual, N., Ruiz-Flaño, P., Errea, M. P., & Lana-Renault, N. (2017). Space–time process and drivers of land abandonment in Europe. *CATENA*, *149*, 810–823. https://doi.org/10.1016/j.catena.2016.02.024

Lasky, J. R., Des Marais, D. L., McKay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation of Arabidopsis thaliana: The roles of geography and climate. *Molecular Ecology*, *21*(22), 5512–5529. https://doi.org/10.1111/j.1365-294X.2012.05709.x

Lasky, J. R., Forester, B. R., & Reimherr, M. (2018). Coherent synthesis of genomic associations with phenotypes and home environments. *Molecular Ecology Resources*, *18*(1), 91–106. https://doi.org/10.1101/051110

Lasky, J. R., Takou, M., Gamba, D., & Keitt, T. H. (2024). Estimating scale-specific and localized spatial patterns in allele frequency. *Genetics*, *227*(3), iyae082. https://doi.org/10.1093/genetics/iyae082

Latorre, S. M., Lang, P. L. M., & Burbano, H. A. (2022). *Historical Arabidopsis thaliana genomes from Germany* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7156189

Le, M. K., Smith, O. S., Akbari, A., Harpak, A., Reich, D., & Narasimhan, V. M. (2022). *1,000 ancient genomes uncover 10,000 years of natural selection in Europe* (p. 2022.08.24.505188). bioRxiv. https://doi.org/10.1101/2022.08.24.505188

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. *Nature Communications*, *8*(1), 14458. https://doi.org/10.1038/ncomms14458

Lee, G., Sanderson, B. J., Ellis, T. J., Dilkes, B. P., McKay, J. K., Ågren, J., & Oakley, C. G. (2024). A large-effect fitness trade-off across environments is explained by a single mutation affecting cold acclimation. *Proceedings of the National Academy of Sciences*, *121*(6), e2317461121. https://doi.org/10.1073/pnas.2317461121

Leimu, R., & Fischer, M. (2008). A Meta-Analysis of Local Adaptation in Plants. *PLoS ONE*, *3*(12), e4010. https://doi.org/10.1371/journal.pone.0004010

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* (arXiv:1303.3997). arXiv. https://doi.org/10.48550/arXiv.1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.*, *25*. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. PMC. https://doi.org/10.1093/bioinformatics/btp352

Lipman, D. J., Souvorov, A., Koonin, E. V., Panchenko, A. R., & Tatusova, T. A. (2002). The relationship of protein conservation and sequence length. *BMC Evolutionary Biology*, *2*(1), 20. https://doi.org/10.1186/1471-2148-2-20

Lopez, L., Marciniak, S., Perry, G. H., & Lasky, J. R. (2022). *Historical Arabidopsis thaliana genomes from across its native range.* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7187528

Lopez, L., Turner, K. g., Bellis, E. s., & Lasky, J. r. (2020). Genomics of Natural History Collections for Understanding Evolution in the Wild. *Molecular Ecology Resources*, *n/a*(n/a). https://doi.org/10.1111/1755-0998.13245

Lopez-Arboleda, W. A., Reinert, S., Nordborg, M., & Korte, A. (2021). Global Genetic Heterogeneity in Adaptive Traits. *Molecular Biology and Evolution*, *38*(11), 4822–4831. https://doi.org/10.1093/molbev/msab208

Lorts, C. M., & Lasky, J. R. (2020). Competition × drought interactions change phenotypic plasticity and the direction of selection on Arabidopsis traits. *New Phytologist*, *n/a*(n/a). https://doi.org/10.1111/nph.16593

Lynch, M., & Lande, R. (1993). Evolution and extinction in response to environmental change. In J. Kingsolver & P. Kareiva (Eds.), *Biotic Interactions and Global Change* (pp. 234–250). Sinauer Associates, Inc. http://www.indiana.edu/~lynchlab/PDF/Lynch57.pdf

Lynch, M., Wei, W., Ye, Z., & Pfrender, M. (2024). The genome-wide signature of short-term temporal selection. *Proceedings of the National Academy of Sciences*, *121*(28), e2307107121. https://doi.org/10.1073/pnas.2307107121

MacLean, H. J., Nielsen, M. E., Kingsolver, J. G., & Buckley, L. B. (2018). Using museum specimens to track morphological shifts through climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1763), 20170404. https://doi.org/10.1098/rstb.2017.0404

Margoliash, E. (1963). Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences*, *50*(4), 672–679. https://doi.org/10.1073/pnas.50.4.672

Martínez-Berdeja, A., Stitzer, M. C., Taylor, M. A., Okada, M., Ezcurra, E., Runcie, D. E., & Schmitt, J. (2020). Functional variants of DOG1 control seed chilling responses and

variation in seasonal life-history strategies in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *117*(5), 2526–2534. https://doi.org/10.1073/pnas.1912451117

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*, 499–503.

McKay, J., Woods, P., Agren, J., Lundemo, S., & Oakley, C. G. (2025). *Collection and Sequencing of Norwegian populations of Arabidopsis thaliana* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.15593806

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Mitchell-Olds, T. (2001). Arabidopsis thaliana and its wild relatives: A model system for ecology and evolution. *Trends in Ecology & Evolution*, *16*(12), 693–700. https://doi.org/doi: DOI: 10.1016/S0169-5347(01)02291-1

Monroe, J. G., McGovern, C., Lasky, J. R., Grogan, K., Beck, J., & McKay, J. K. (2016). Adaptation to warmer climates by parallel functional evolution of CBF genes in Arabidopsis thaliana. *Molecular Ecology*, *25*(15), 3632–3644. https://doi.org/10.1111/mec.13711

Montesinos, A., Tonsor, S. J., Alonso-Blanco, C., & Picó, F. X. (2009). Demographic and Genetic Patterns of Variation among Populations of Arabidopsis thaliana from Contrasting Native Environments. *PLoS ONE*, *4*(9), e7213. https://doi.org/10.1371/journal.pone.0007213

Montesinos-Navarro, A., Wig, J., Xavier Pico, F., & Tonsor, S. J. (2011). Arabidopsis thaliana populations show clinal variation in a climatic gradient associated with altitude. *New Phytologist*, *189*(1), 282–294.

Munguía-Rosas, M. A., Ollerton, J., Parra-Tabla, V., & De-Nova, J. A. (2011). Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecology Letters*, *14*(5), 511–521. https://doi.org/10.1111/j.1461-0248.2011.01601.x

Ng, M., McCormick, A., Utz, R. M., & Heberling, J. M. (2023). Herbarium specimens reveal century-long trait shifts in poison ivy due to anthropogenic CO2 emissions. *American Journal of Botany*, *110*(9), e16225. https://doi.org/10.1002/ajb2.16225

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*(11), 1566–1575. https://doi.org/10.1101/gr.4252305

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T. K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., … Reich, D. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*, *555*(7695), 190–196. https://doi.org/10.1038/nature25738

Patwa, Z., & Wahl, L. M. (2008). The fixation probability of beneficial mutations. *Journal of the Royal Society Interface*, *5*(28), 1279–1289. https://doi.org/10.1098/rsif.2008.0248

Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*(6), 1638–1647. https://doi.org/10.1111/2041-210X.12984

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Picó, F. X., Méndez-Vigo, B., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2008). Natural genetic variation of Arabidopsis thaliana is geographically structured in the Iberian peninsula. *Genetics*, *180*(2), 1009–1021. https://doi.org/10.1534/genetics.108.089581

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Renaud, G., Stenzel, U., & Kelso, J. (2014). leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Research*, *42*(18), e141. https://doi.org/10.1093/nar/gku699

Rhoné, B., Defrance, D., Berthouly-Salazar, C., Mariac, C., Cubry, P., Couderc, M., Dequincey, A., Assoumanne, A., Kane, N. A., Sultan, B., Barnaud, A., & Vigouroux, Y. (2020). Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-19066-4

Rosche, C., Broennimann, O., Novikov, A., Mrázová, V., Boiko, G. V., Danihelka, J., Gastner, M. T., Guisan, A., Kožić, K., Lehnert, M., Müller-Schärer, H., Nagy, D. U., Remelgado, R., Ronikier, M., Selke, J. A., Shiyan, N. M., Suchan, T., Thoma, A. E., Zdvořák, P., & Mráz, P. (2024). Herbarium specimens reveal a cryptic invasion of polyploid Centaurea stoebe in Europe. *New Phytologist*, *n/a*(n/a). https://doi.org/10.1111/nph.20212

Roy, S. (2018). Arabidopsis natural variants and the Indian scenario. *Current Science*, *114*(2), 263–265.

Saez-Aguayo, S., Rondeau-Mouro, C., Macquet, A., Kronholm, I., Ralet, M.-C., Berger, A., Sallé, C., Poulain, D., Granier, F., Botran, L., Loudet, O., Meaux, J. de, Marion-Poll, A., & North, H. M. (2014). Local Evolution of Seed Flotation in Arabidopsis. *PLOS Genetics*, *10*(3), e1004221. https://doi.org/10.1371/journal.pgen.1004221

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Schmid, K., Törjék, O., Meyer, R., Schmuths, H., Hoffmann, M., & Altmann, T. (2006). Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *TAG Theoretical and Applied Genetics*, *112*(6), 1104–1114.

Seguin-Orlando, A., Schubert, M., Clary, J., Stagegaard, J., Alberdi, M. T., Prado, J. L., Prieto, A., Willerslev, E., & Orlando, L. (2013). Ligation Bias in Illumina Next-Generation DNA Libraries: Implications for Sequencing Ancient Genomes. *PLOS ONE*, *8*(10), e78575. https://doi.org/10.1371/journal.pone.0078575

Sekor, M., & Franks, S. (2018). An experimentally introduced population of Brassica rapa (Brassicaceae). 2. Rapid evolution of phenotypic traits. *Plant Ecology and Evolution*, *151*(3), Article 3. https://doi.org/10.5091/plecevo.2018.1401

Sharbel, T. F., Haubold, B., & Mitchell-Olds, T. (2000). Genetic isolation by distance in Arabidopsis thaliana: Biogeography and postglacial colonization of Europe. *Molecular Ecology*, *9*(12), 2109–2118.

Siewert, K. M., & Voight, B. F. (2017). Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Molecular Biology and Evolution*, *34*(11), 2996–3005. https://doi.org/10.1093/molbev/msx209

Simon, A., & Coop, G. (2024). The contribution of gene flow, selection, and genetic drift to five thousand years of human allele frequency change. *Proceedings of the National Academy of Sciences*, *121*(9), e2312377121. https://doi.org/10.1073/pnas.2312377121

Slatkin, M., & Hudson, R. R. (1991). Pairwise Comparisons of Mitochondrial DNA Sequences in Stable and Exponentially Growing Populations. *Genetics*, *129*(2), 555. https://doi.org/10.1093/genetics/129.2.555

Smith, O., Nicholson, W. V., Kistler, L., Mace, E., Clapham, A., Rose, P., Stevens, C., Ware, R., Samavedam, S., Barker, G., Jordan, D., Fuller, D. Q., & Allaby, R. G. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. *Nature Plants*, *5*(4), 369–379. https://doi.org/10.1038/s41477-019-0397-9

Staats, M., Cuenca, A., Richardson, J. E., Ginkel, R. V., Petersen, G., Seberg, O., & Bakker, F. T. (2011). DNA Damage in Plant Herbarium Tissue. *PLOS ONE*, *6*(12), e28448. https://doi.org/10.1371/journal.pone.0028448

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., Purugganan, M. D., & Schmitt, J. (2004). A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4712–4717. https://doi.org/10.1073/pnas.0306401101

Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., Shen, G., Qi, J., Qin, Y., Li, J., Wang, L., Chang, W., Guo, Z., Baldwin, I. T., & Wu, J. (2018). Large-scale gene losses underlie the genome evolution of parasitic plant Cuscuta australis. *Nature Communications*, *9*(1), 2683. https://doi.org/10.1038/s41467-018-04721-8

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., Kruse-Peeples, M., Lepak, N., Prim, L., Romay, M. C., Ross-Ibarra, J., Sanchez-Gonzalez, J. de J., Schmidt, C., Schuenemann, V. J., Krause, J., Matson, R. G., Weigel, D., Buckler, E. S., & Burbano, H. A. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science*, *357*(6350), 512–515. https://doi.org/10.1126/science.aam9425

Takou, M., Wieters, B., Kopriva, S., Coupland, G., Linstädter, A., & De Meaux, J. (2019). Linking genes with ecological strategies in Arabidopsis thaliana. *Journal of Experimental Botany*, *70*(4), 1141–1151. https://doi.org/10.1093/jxb/ery447

Todesco, M., Bercovich, N., Kim, A., Imerovski, I., Owens, G. L., Dorado Ruiz, Ó., Holalu, S. V., Madilao, L. L., Jahani, M., Légaré, J.-S., Blackman, B. K., & Rieseberg, L. H. (2022). Genetic basis and dual adaptive role of floral pigmentation in sunflowers. *eLife*, *11*, e72072. https://doi.org/10.7554/eLife.72072

Troth, A., Puzey, J. R., Kim, R. S., Willis, J. H., & Kelly, J. K. (2018). Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science*, *361*(6401), 475–478. https://doi.org/10.1126/science.aat5760

Turner, K. G., Lorts, C. M., Haile, A. T., & Lasky, J. R. (2020). Effects of genomic and functional diversity on stand-level productivity and performance of non-native Arabidopsis. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1937), 20202041. https://doi.org/10.1098/rspb.2020.2041

Tusiime, F. M., Gizaw, A., Gussarova, G., Nemomissa, S., Popp, M., Masao, C. A., Wondimu, T., Abdi, A. A., Mirré, V., Muwanika, V., Eilu, G., & Brochmann, C. (2020). Afro-alpine flagships revisited: Parallel adaptation, intermountain admixture and shallow genetic structuring in the giant senecios (Dendrosenecio). *PLoS ONE*, *15*(3), e0228979. https://doi.org/10.1371/journal.pone.0228979

Tyagi, A., Singh, S., Mishra, P., Singh, A., Tripathi, A. M., Jena, S. N., & Roy, S. (2016). Genetic diversity and population structure of Arabidopsis thaliana along an altitudinal gradient. *AoB PLANTS*, *8*, plv145. https://doi.org/10.1093/aobpla/plv145

Vasseur, F., Sartori, K., Baron, E., Fort, F., Kazakou, E., Segrestin, J., Garnier, E., Vile, D., & Violle, C. (2018). Climate as a driver of adaptive variations in ecological strategies in Arabidopsis thaliana. *Annals of Botany*, *122*(6), 935–945. https://doi.org/10.1093/aob/mcy165

Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Maisano Delser, P., Hare, A. J., Burger, J., Collins, M. J., Kehati, R., Hesse, P., Fulton, D., Sauer, E. W., Mohaseb, F. A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., … Bradley, D. G.

(2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science*, *365*(6449), 173–176. https://doi.org/10.1126/science.aav1002

Wales, N., & Kistler, L. (2019). Extraction of Ancient DNA from Plant Remains. *Methods in Molecular Biology (Clifton, N.J.)*, *1963*, 45–55. https://doi.org/10.1007/978-1-4939-9176-1_6

Wang, Y., Zhang, H., Wang, P., Zhong, H., Liu, W., Zhang, S., Xiong, L., Wu, Y., & Xia, Y. (2023). Arabidopsis EXTRA-LARGE G PROTEIN 1 (XLG1) functions together with XLG2 and XLG3 in PAMP-triggered MAPK activation and immunity. *Journal of Integrative Plant Biology*, *65*(3), 825–837. https://doi.org/10.1111/jipb.13391

Wang, Z., Yang, L., Liu, Z., Lu, M., Wang, M., Sun, Q., Lan, Y., Shi, T., Wu, D., & Hua, J. (2019). Natural variations of growth thermo-responsiveness determined by SAUR26/27/28 proteins in Arabidopsis thaliana. *New Phytologist*, *224*(1), 291–305. https://doi.org/10.1111/nph.15956

Wilczek, A. M., Cooper, M. D., Korves, T. M., & Schmitt, J. (2014). Lagging adaptation to warming climate in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *111*(22), 7906–7913. https://doi.org/10.1073/pnas.1406314111

Wondimu, T., Gizaw, A., Tusiime, F. M., Masao, C. A., Abdi, A. A., Gussarova, G., Popp, M., Nemomissa, S., & Brochmann, C. (2014). Crossing barriers in an extremely fragmented system: Two case studies in the afro-alpine sky island flora. *Plant Systematics and Evolution*, *300*(3), 415–430. https://doi.org/10.1007/s00606-013-0892-9

Wuest, S. E., & Niklaus, P. A. (2018). A plant biodiversity effect resolved to a single chromosomal region. *Nature Ecology & Evolution*, *2*(12), Article 12. https://doi.org/10.1038/s41559-018-0708-y

Yamaguchi, K., Mezaki, H., Fujiwara, M., Hara, Y., & Kawasaki, T. (2017). Arabidopsis ubiquitin ligase PUB12 interacts with and negatively regulates Chitin Elicitor Receptor Kinase 1 (CERK1). *PLOS ONE*, *12*(11), e0188886. https://doi.org/10.1371/journal.pone.0188886

Yim, C., Bellis, E. S., DeLeo, V. L., Gamba, D., Muscarella, R., & Lasky, J. R. (2023). Climate biogeography of Arabidopsis thaliana: Linking distribution models and individual variation. *Journal of Biogeography*, *51*(n/a), 560–574. https://doi.org/10.1111/jbi.14737

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *eLife*, *2*, e00731. https://doi.org/10.7554/eLife.00731

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

Zou, Y.-P., Hou, X.-H., Wu, Q., Chen, J.-F., Li, Z.-W., Han, T.-S., Niu, X.-M., Yang, L., Xu, Y.-C., Zhang, J., Zhang, F.-M., Tan, D., Tian, Z., Gu, H., & Guo, Y.-L. (2017). Adaptation of Arabidopsis thaliana to the Yangtze River basin. *Genome Biology*, *18*(1), 239. https://doi.org/10.1186/s13059-017-1378-9

Data accessibility

Genetic data:

Individual genotype data are available on DataDryad (DOI: 10.5061/dryad.31zcrjdz1 **).** Raw reads are available at NCBI SRA as project number PRJNA1269104.

Sample metadata:

Metadata can be found in supplementary tables of this manuscript.

Code:

Scripts are available on GitHub (https://github.com/jesserlasky/Arabidopsis_herbarium) and on DataDryad (DOI: 10.5061/dryad.31zcrjdz1 **)**.

1388 **Supplemental Material**
1389
1390 *DNA extraction, library preparation and sequencing: Newly generated data*
1391 *Herbarium samples* – One hundred and thirty newly obtained samples from herbaria were
1392 transported into a dedicated ancient DNA laboratory at Pennsylvania State University, USA. To
1393 avoid introducing contamination, the plastic bags containing the samples were cleaned with a 2%
1394 bleach solution. The surface of each sample was gently cleaned of potential soil and dust particles
1395 with a thin brush. The brush was cleaned in between samples by submerging the tip in a 2% bleach
1396 solution and then thoroughly rinsing it with milli-Q water (Millipore Corporation). DNA extraction
1397 was done using one medium-size leaf of *Arabidopsis thaliana*. Prior to extraction, tissue was
1398 homogenized using a TissueLyser (Qiagen). Each sample followed two cycles of homogenization,
1399 the first cycle was done for 30 seconds and the second was 15 seconds to ensure all tissue had been
1400 pulverized. Then, the DNA was extracted using the PTB protocol recommended for ancient
1401 samples with minor modifications (Wales & Kistler, 2019). Modifications were made to the length
1402 of the digestion (24 hours), to the final amount of Buffer AW1 (5x) and the amount of the elution
1403 buffer (25 µl). After extraction, DNA concentration was determined in a Qubit fluorometer using
1404 the Qubit dsDNA high-sensitivity (HS) Assay Kit (ThermoFisher Scientific). In addition, 5 µl of
1405 DNA from each sample were run in a 1% agarose gel to visually inspect the degree of DNA
1406 fragmentation.
1407 After DNA extraction, we proceeded to library preparation. DNA obtained from herbarium
1408 samples is expected to be highly fragmented. Thus, no shearing step is done before library
1409 preparation. However, since our study includes both samples from recent years (~ 10 years ago)
1410 and samples collected one or two centuries ago, we performed library preparation based on the
1411 particular fragment-size of each sample. Samples that showed DNA above the 500 bp ladder band
1412 were subjected to a short shearing step (30 seconds) using a M220 Focused Ultrasonicator
1413 (Covaris). Samples lacking DNA above the 500 bp ladder band were processed for library
1414 preparation as recommended for ancient samples (no shearing step). Herbarium-preserved samples
1415 are known to show fragmentation and damage rapidly after preservation (Staats et al., 2011). Thus,
1416 to capture short fragments and maximize complexity despite low input concentration we followed
1417 a library preparation method that relies on blunt-end ligation of custom adapters (Meyer & Kircher,
1418 2010). This method is known to be more efficient and less-biased for chemically damaged and
1419 fragmented DNA than the AT-overhang adapter ligation commonly used in commercially
1420 available library preparation kits (Seguin-Orlando et al., 2013). One modification was
1421 implemented for the non-sheared samples; cleaning steps during the library preparation process
1422 were done using first the QIAquickNucleotide Removal Kit (Qiagen) and then the MinElute PCR
1423 Purification Kit (Qiagen) instead of AMPure XP/SPRIselect magnetic beads (Beckman Coulter).
1424 Final DNA content in the libraries was quantified using a Qubit fluorometer with the Qubit dsDNA
1425 high-sensitivity (HS) Assay Kit (ThermoFisher Scientific). Since samples had individual
1426 identifying barcodes, a pooled equimolar aliquot was prepared for sequencing. Each sequencing
1427 run contained approximately 40 randomly selected samples, including extraction and library

preparation blanks to ensure that no contamination had happened during any step. The 42 samples used for the pilot study were sequenced PE at 100bp in a HiSeq 2500 platform. The remaining samples were sequenced at 75bp PE in a NextSeq 550 Illumina sequencer.

Recent field collections - A total of 95 fresh samples were included in this study (61 field samples from Norway (McKay et al., 2025), 30 from East Africa and 4 from the INRA stock center). Thirteen fresh leaf samples were collected from the field in East Africa while the remaining 17 were collected as seeds. Leaf material was immediately desiccated in silica gel within sealed plastic bags. The samples were transported to the laboratory and stored at 4ºC until DNA extraction. Seed samples from East Africa, together with seeds from the 4 INRA stock center lines, were cultivated in growth chambers, and leaf tissue was dried in silica gel. Prior to DNA extraction all samples were flash-frozen in liquid Nitrogen and then homogenized as done with the herbarium samples. Next, DNA was extracted using the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's recommendations. The extracted DNA concentrations were determined in a Qubit fluorometer using the Qubit dsDNA high-sensitivity (HS) Assay Kit (ThermoFisher Scientific). DNA from the field-collected leaf samples were sent to BGI (China) for library preparation and sequencing. These samples were sequenced with paired end (PE) 150bp reads on a HiSeq X Ten Illumina machine. The remaining samples were processed by Genomics and Bioinformatics Service at Texas A&M AgriLife, where library preparation was conducted and sequencing was done on a NovaSeq 6000 Illumina platform with PE 150bp reads. Finally, the Norwegian field samples were collected as seeds and cultivated in a Colorado State University greenhouse. A single plant was sampled from each family once plants reached the vegetative state, and leaf tissue was from that plant was used for subsequent DNA extraction using the Qiagen DNeasy Plant Mini Kit (Valencia CA, USA). Extracted DNA was then quantified using a Qubit Fluorometer (ThermoFisher Scientific). Whole genome sequencing (WGS) libraries of the extracted DNA were prepared at the University of Colorado Boulder sequencing core. These WGS libraries were then paired-end (2 x 150 bp) whole genome sequenced at the University of Colorado Anschutz Medical Campus using an Illumina HiSeq.

Data preprocessing and de novo *SNP calling*

Herbarium samples preprocessing - Herbarium samples were mostly newly generated data (130). To these we added five samples from the African genomes (Durvasula et al., 2017) and 33 German genomes (Lang et al., 2024) also collected from herbarium vouchers. Raw data for the African genomes was obtained from the European Nucleotide archive (ENA) of the European Molecular Biology Lab - European Bioinformatics Institute (EMBL - EBI) and German genomes from (Lang et al., 2024). Raw read sequence data were assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects) to confirm that they met our quality standards. For all herbarium samples cytosine deamination profiles characteristic of ancient DNA were verified using mapDamage 2.0 (Jónsson et al., 2013) (Figure S1). Samples processed using the traditional aDNA library preparation and samples sheared before library preparation were

1468 checked with MapDamage 2.0 separately to ensure that all meet the ancient DNA criteria. Reads
1469 were trimmed using Leehom with the -ancientdna flag (Renaud et al., 2014). Merged and
1470 unmerged reads were then concatenated to maximize genome coverage. Trimmed reads were
1471 mapped to the *Arabidopsis thaliana* nuclear genome (TAIR v. 10) with the BWA v.0.7.17 *-aln*
1472 function since this option has a higher mapping rate for short fragments compared with alternatives
1473 (Li, 2013; Li & Durbin, 2009). After the reads from each sample were mapped to the nuclear
1474 genome, SAM_{TOOLS} v.1.5 was used to transform the mapped SAM files to BAM, sort the BAM
1475 files, remove duplicates, and filter the reads (minimum mapping quality of 30 and 30 base pair
1476 minimum length) (Li et al., 2009). Finally, SAM_{TOOLS} *flagstat* was used to count the number of
1477 alignments for every sample mapped to the *A. thaliana* nuclear genome (Li et al., 2009).
1478
1479 *Fresh samples and publicly available data preprocessing* - Raw data from 199 samples (Alonso-
1480 Blanco et al., 2016) were downloaded from the Sequence Read Archive (SRA) of the National
1481 Center for Biotechnology Information (NCBI). Additionally, the 65 African samples (all
1482 Moroccan except one from Tanzania) were obtained from the European Nucleotide archive
1483 (ENA) of the European Molecular Biology Lab - European Bioinformatics Institute (EMBL -
1484 EBI). For all the publicly available samples and the ones generated from fresh tissue, raw read
1485 sequence data were assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects)
1486 to confirm that they met our quality standards. Raw paired-end reads were trimmed and filtered
1487 using Trimmomatic-0.38 to remove adapter sequences, short reads, and low-quality reads from
1488 raw sequence data (Bolger et al., 2014). To remove adapter sequences, we used seed matches
1489 with a maximum of three mismatches and an initial length of 10. These seeds were extended and
1490 clipped when paired-end reads had a score of 30 or lower. Additionally, leading and trailing
1491 bases of low quality (Phred quality score <20) or N were removed. Likewise, reads where the
1492 quality of a 4-base-wide sliding window dropped below 15 at the 3' end were removed. Finally,
1493 any remaining reads that were shorter than 75 bases long were eliminated. Trimmed reads were
1494 then mapped to the *Arabidopsis thaliana* nuclear genome (TAIR v. 10) with BWA v.0.7.17 *-mem*
1495 flag (Li, 2013; Li & Durbin, 2009). After the reads from each sample were mapped to the nuclear
1496 genome, SAM_{TOOLS} v.1.5 was used to convert the mapped SAM files to BAM, sort the BAM
1497 files, remove duplicates, and filter the reads (minimum mapping quality of 30 and 30 base pair
1498 minimum length) (Li et al., 2009). Finally, SAM_{TOOLS} *-flagstat* was used to count the number of
1499 alignments for every sample mapper to the *A. thaliana* nuclear genome (Li et al., 2009).
1500
1501 de novo *SNP calling* - SNP discovery was done using a pseudohaplotype approach (Kistler et al.,
1502 2018). A combination of SAM_{TOOLS} *mpileup* and VarScan v2.3.9 *mpileup2snp* was used to
1503 discover all candidate variant positions in all samples with minimum mapping quality of 20, a
1504 minimum read depth of 2 and at least 2 supporting reads to call a variant (Koboldt et al., 2009; Li
1505 et al., 2009). To exclude paralogs and repetitive regions, variable sites in the top 10% of coverage
1506 were filtered out and the remaining sites were summarized in a bed file. The bed file was used to
1507 create individual pseudohaplotype files for each sample of variable sites with a minimum coverage

of 2 and a maximum of 70 (these coverage thresholds were selected based on the specific coverage data from all samples included in this study). After filtering sites by coverage, individual files were combined into plink tped and tfam files and culled to biallelic sites. Finally, PLINK v1.90 was used create the final vcf file containing the selected variable sites (Purcell et al., 2007).

Simulating purifying and background selection

We used the SLiM simulation software to simulate a Wright-Fisher model with semi-realistic genome structure (genes with introns and exons, and intergenic space) and a combination of neutral and deleterious mutations, with deleterious mutations only occurring in genic sequences (Haller & Messer, 2019). We simulated a stepping-stone model of 10 populations with a migration rate of 0.005 and used a selfing rate of 95%. We ran models for 100k generations, and tested simulations with different combinations of genome size, deleterious mutation distribution of fitness effects, and recombination rates. We then tested each SNP of at least 5% MAF in a sample of 200 individuals across the last 100 generation in regression models with time (as in the Arabidopsis data). We then tested for enrichment of the temporal allele frequency change statistics for mutations in different genomic contexts (genic vs intergenic).

Figure S1: Historical sample authentication. Merged (on the left) and unmerged reads (on the right) were run separately. Black indicates samples that were not sheared (aDNA) and red indicates samples that were shared (sheared) before library preparation due to differences among samples in the existing degree of fragmentation. A) Fraction of C-to-T converted base-pairs along sequencing reads and the B) correlation of C-to-T conversion at the first base of a read respective to sample's age. These patterns show the characteristic damage done to DNA fragments accumulated over time.

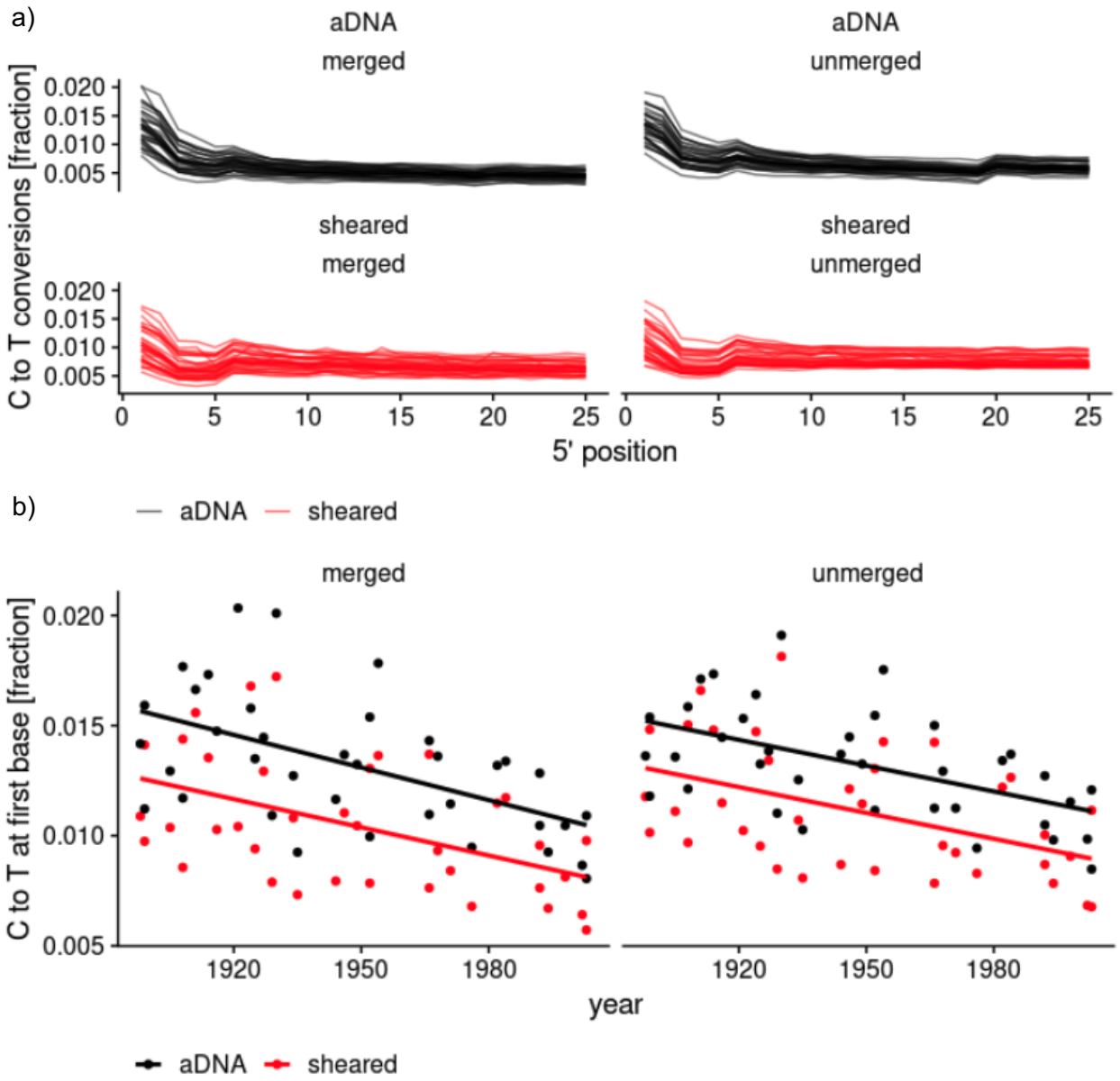
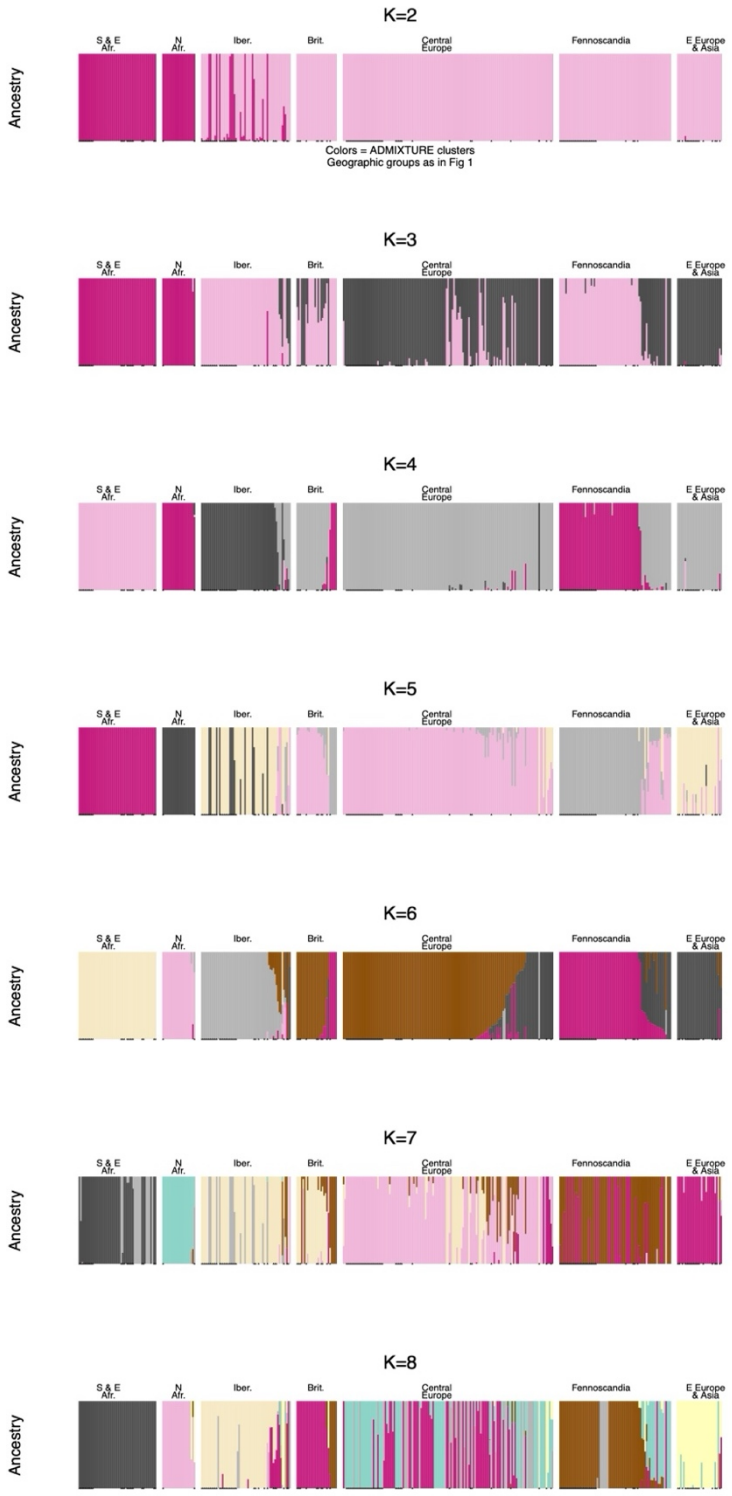
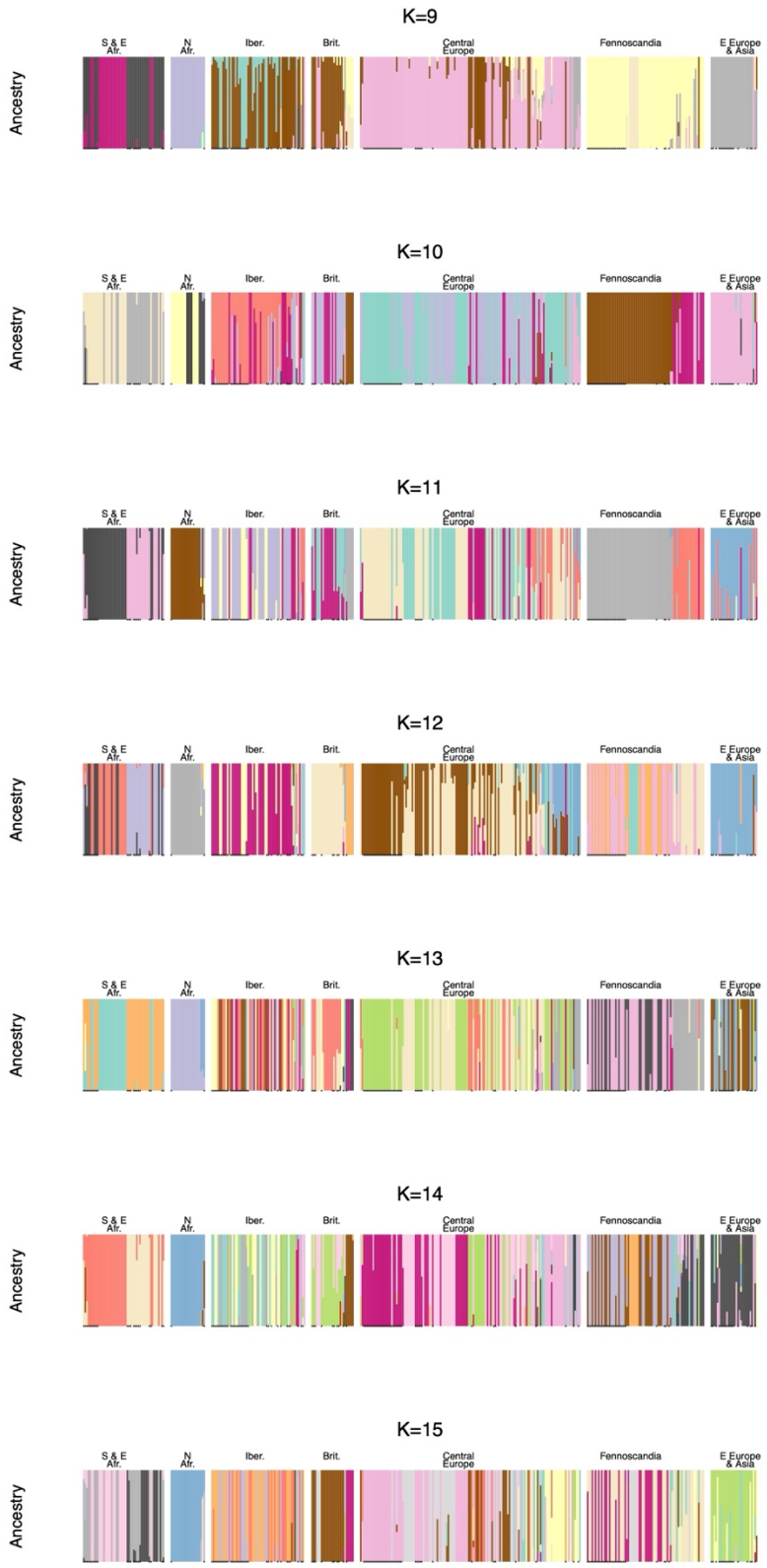


Figure S2. ADMIXTURE results for K=2-15, with samples grouped along the x-axis by geographic region as in Figure 1. Colors indicate assignment to ancestral clusters.




K=9
K=10
K=11
K=12
K=13
K=14
K=15
Ancestry
S & E Afr.
N Afr.
Iber.
Brit.
Central Europe
Fennoscandia
E Europe & Asia

1544

Figure S3. Results of ADMIXTURE clustering with different K values **across Eurasia**. This clustering was primarily for visualization purposes and determining turnover in genotypes within regions in Figure 4. For this reason, we used K=6 in analyses, though we also conducted a version of analysis with K=3. This K resolved different regional genotypic clusters but came before a sharp increase in cross validation error at K=7.

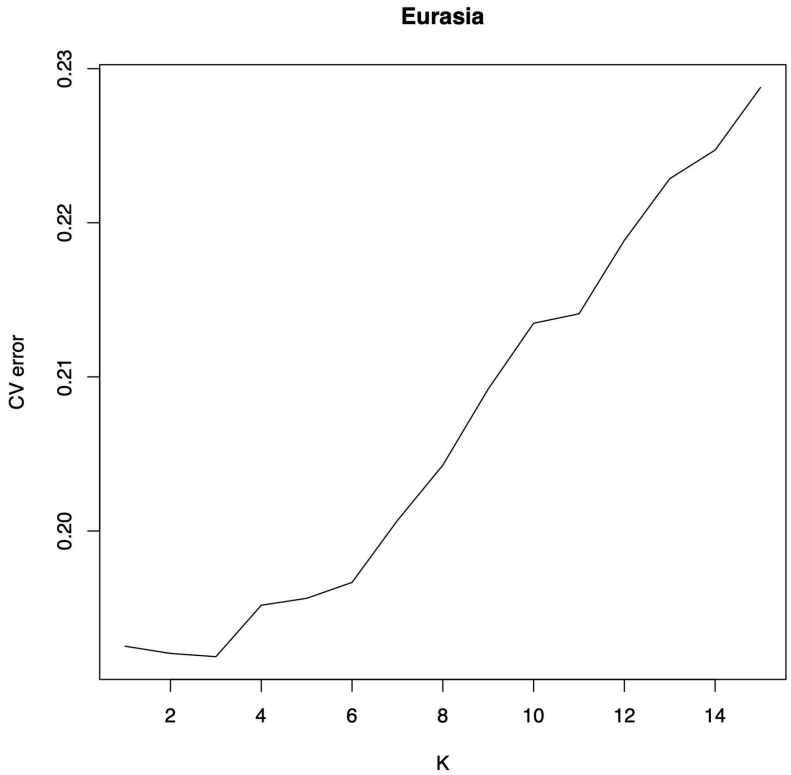


Figure S4. Comparing the observed difference in temporal allele shifts in conserved ('cons') versus non-conserved ('nc') genes of different lengths (AA = amino acid) for Eurasian samples. For each gene we calculated the 0.1 quantile p-value for SNPs in the gene, and then we compared the median of these values between conserved and non-conserved genes. When the observed is *less* than zero and the null, it indicates conserved genes show greater temporal allele frequency shifts (as for genes less than or equal to 200 AA in length using the Wilcoxon and logistic tests that do not correct for population structure). When the observed is *greater* than zero and the null, it indicates non-conserved genes show greater temporal allele frequency shifts (as for genes greater than 600 AA in length). The overall trends are similar between tests except for small genes, where the greater turnover in conserved small genes is only seen in tests that do not correct for population structure (Wilcoxon and logistic) while the mixed model turnover is not different for small, conserved genes compared to non-conserved. This discrepancy for small genes suggests that the high turnover for conserved small genes follows the genome-wide pattern for turnover, a turnover possibly due to drift being stronger due to their background selection. However, this hypothesis does not seem consistent with what is observed for longer conserved genes (less turnover than expected).

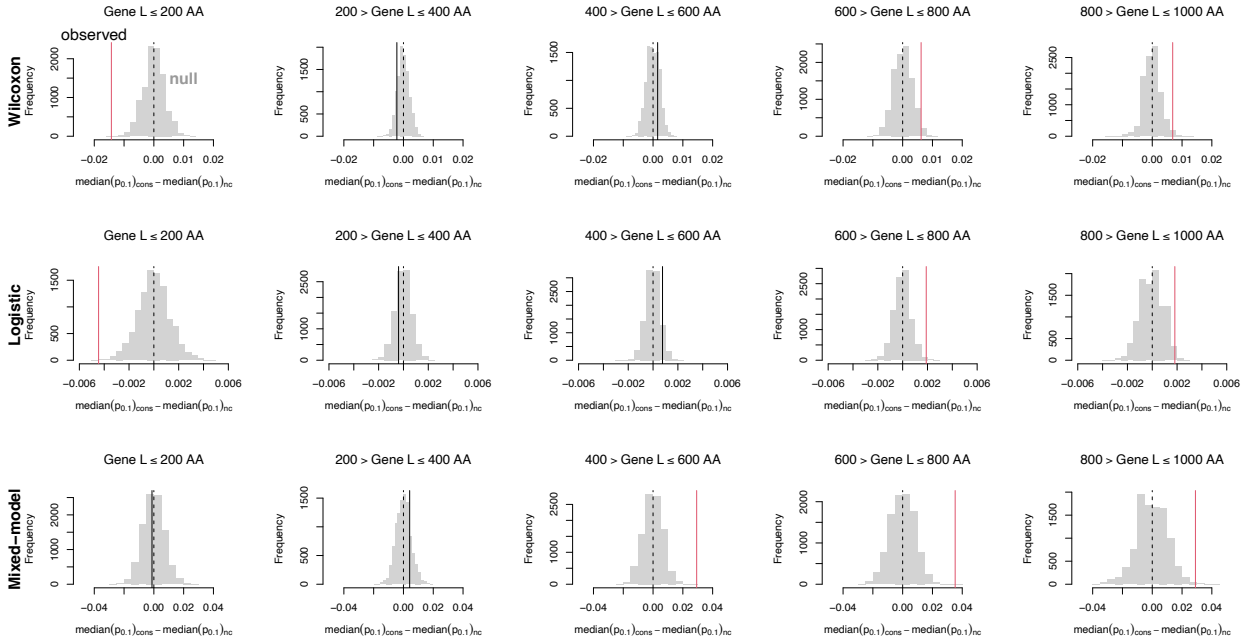
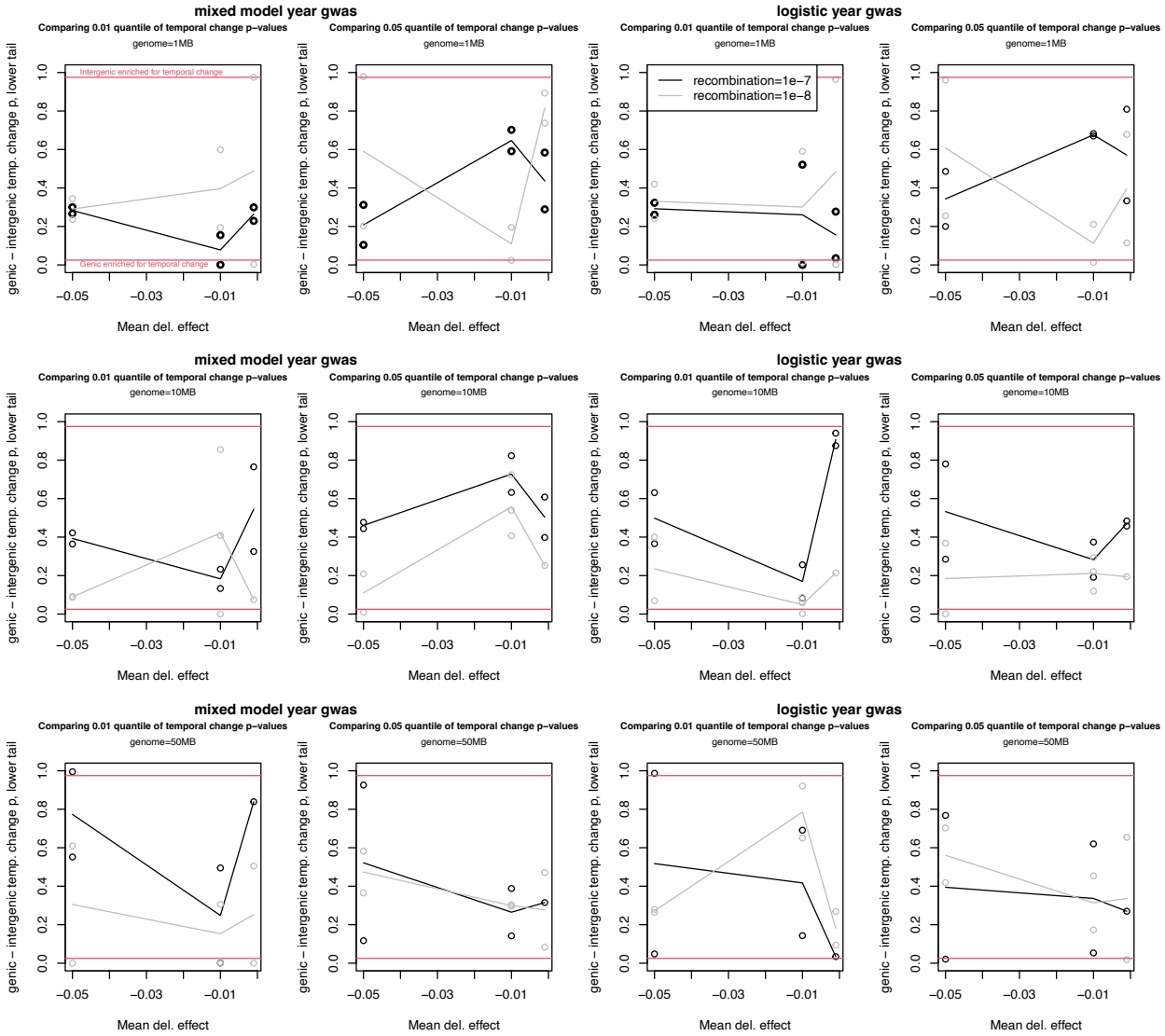



Figure S5. Results from simulations of purifying and background selection show no clear effect on the enrichment of genic SNPs for temporal allele frequency change. Purifying selection acts on a subset of coding SNPs and a subset of intronic SNPs, with neutrality for other SNPs. The y-axes show the lower tail of a permutation based null-model test, where the test statistic is the genic SNPs' 0.01 p-value quantile minus the intergenic SNPs' 0.01 quantile (1st and 3rd columns) or 0.05 quantile (2nd and 4th columns). Both linear mixed models testing year correlations with SNPs while accounting for kinship (1st and 2nd columns), or simple logistic year versus SNP allele (3rd and 4th columns) models are shown. Dots show individual simulation results, lines show means.