

1 Global patterns of natural selection inferred using ancient DNA

2 Laura L. Colbran^{1,*}, Jonathan Terhorst^{2,*}, and Iain Mathieson^{1,*}

3 ¹Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.

4 ²Department of Statistics, University of Michigan, Ann Arbor, MI, 48109, USA.

5 *Correspondence to LLC (laura.colbran@pennmedicine.upenn.edu), JT (jonth@umich.edu) or IM
6 (mathi@pennmedicine.upenn.edu)

7 Abstract

8 Ancient DNA has revolutionized our understanding of human history, and is now yielding important
9 insights into evolution and natural selection. However, studies of selection using ancient DNA have
10 largely been limited to Europe, excluding populations in other parts of the world. While many selec-
11 tive pressures were local to specific populations others, for example those related to the development
12 of agriculture, may have been universal. By studying a broader range of global populations, we can
13 identify examples of local adaption but also more general principles of adaptation to climatic, social
14 and technological changes.

15 We therefore leverage ancient DNA to test for selection in 7244 individuals from 13 ancient
16 and 19 present-day populations across five regions: Europe, East Asia, South Asia, Africa and the
17 Americas. In each region, we tested for selection using multiple approaches that account for complex
18 demographic histories. We identify 31 genome-wide significant signals of selection, including both
19 known and novel loci. We find a high degree of shared signal across regions, suggesting extensive
20 parallel or shared adaptation. Using a novel admixture-aware time series method, we find that the
21 strength of selection on many variants changed over time, for example decreasing selection at *LCT*
22 in Europe and increasing selection at *ADH1B* in East Asia over the past few thousand years.

23 Finally, we developed a test for polygenic selection on complex traits by modeling the frequencies
24 of trait-associated alleles identified in GWAS. We tested for selection jointly across regions, avoiding
25 the confounding effect of population stratification by excluding the European or East Asian GWAS
26 population from the selection test. We find evidence for directional selection on pigmentation and
27 immune traits, and that strong stabilizing selection on female waist-hip ratio was universal across
28 human populations suggesting a fundamental constraint on human morphology.

29 Introduction

30 As modern humans migrated out of Africa and throughout the world over the last 50,000 years, they
31 encountered dramatically different environmental pressures, providing the opportunity for genetic adap-
32 tation to shape human variation. In the past 10,000 years (the Holocene), populations in different parts
33 of the world independently developed agriculture, urbanism and new forms of social organization. This
34 raises the question of how humans adapted to these new environments, how much those adaptations were
35 shared across populations, and how they affect human diversity today. Understanding which parts of
36 the genome have been subject to selection can highlight functionally important regions, and contribute
37 to understanding how they influence phenotype. Further, strongly selected regions can play a role in
38 medically-significant phenotypes, regardless of their current fitness cost, making patterns of selection
39 informative for understanding variation in health outcomes among populations.

40 The emerging science of ancient DNA has brought a new perspective to studies of human evolution.
41 The ability to observe allele frequency changes directly combined with a better understanding of demo-
42 graphic history has enabled powerful inference of natural selection [1], precise estimates of timing [2, 3],
43 and the ability to recover signatures of selection that have been obscured by admixture or genetic drift
44 [4, 5]. However, due to sampling limitations, large-scale ancient DNA selection studies have been limited
45 to Europe [1, 6, 7, 3, 8], limiting the scope for comparative evolutionary studies.

46 We therefore took advantage of the recent influx of non-European ancient DNA to conduct a scan for
47 selection across five regions of the world—Africa, East Asia, Europe, South Asia, and the Americas—
48 each with hundreds of ancient samples and present-day population data. Each of these regions has a
49 complex history of Holocene admixture and we adapted an allele frequency-based test [1] to account for
50 more general admixture history. We further adapted this test to detect polygenic selection, by modeling
51 expected frequencies of alleles associated with a complex trait. One of the primary advantages of ancient
52 DNA is that ancient individuals can be directly dated so we also developed a time series approach to
53 model the evolution of allele frequencies in admixed populations. We find evidence of both private and
54 shared targets of selection, demonstrating the power of ancient DNA to highlight the interconnected
55 nature of evolution in the recent human past.

56 Results

57 An allele frequency-based test for selection

58 Identifying natural selection in genomic data requires distinguishing its effects from the effects of ad-
59 mixture and genetic drift. The Holocene is characterized by high levels of admixture between previously
60 diverged populations, which can lead to both false positives and false negatives in selection scans [4, 5, 8].
61 We therefore developed a test where we model expected allele frequency based on genome-wide admix-
62 ture proportions, extending a previous approach [1] by modeling the ancient populations themselves as
63 admixtures of unobserved source populations. This approach is similar to that of Cheng et al. [9], except
64 that we treat source populations as independent. We perform a likelihood ratio test and then apply a
65 correction for the inflation in test statistics due to genetic drift by calculating P-values from a gamma
66 distribution fitted to the distribution of test statistics. We applied this method to test for selection at
67 1,150,640 variants captured by the 1240k reagent [10, 11] for five regions with a genetic record which in-
68 cludes both ancient and present-day samples: Africa, East Asia, Europe, South Asia, and Central/South
69 America (Fig. 1A). In each region, we defined populations of ancient and present-day individuals based on
70 genome-wide ancestry and inferred ancestry proportions from the unobserved source populations using
71 ADMIXTURE [12] (Supplementary Figures 1-5).

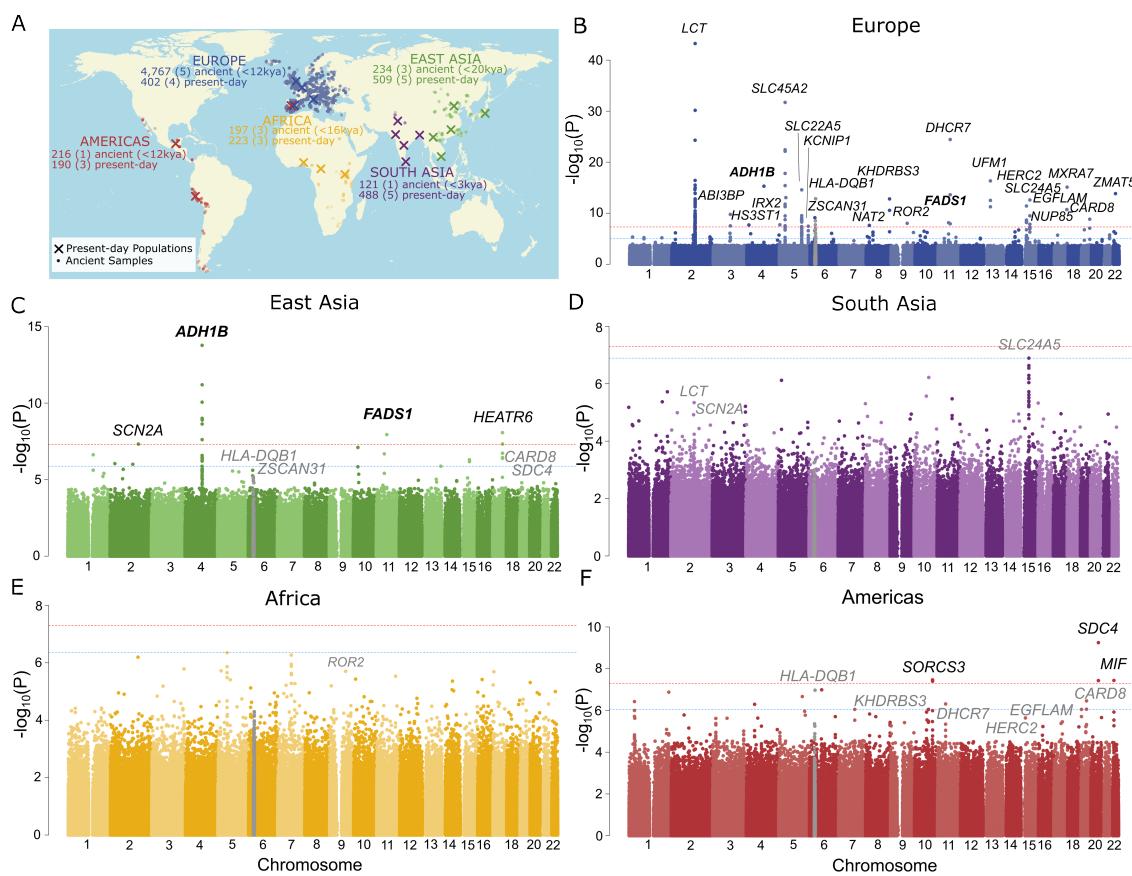


Figure 1: Scan for selection on 1240k variants based on admixture proportions for 5 regions highlighted in A). Manhattan plots of the results for B) Europe, C) East Asia, D) South Asia, E) Africa, and F) the Americas. Significant SNPs are filtered to only include those with at least one more nearby (at least 2 more for Europe), in high LD, and also passing FDR<0.1 (Methods). The HLA is indicated in grey, as it was treated independently of the rest of the genome. Peaks are labeled with the causal gene if known, or the nearest to the lead SNP. Bold text indicates loci passing genome-wide multiple testing in multiple regions, while grey text labels loci that are genome-wide significant in another region, and nominally significant in the indicated region after correcting for the total number of genome-wide significant peaks.

72 31 signals of selection with extensive sharing across continents

73 After filtering out SNPs with potential genotype errors, we identified 24, 4, and 3 genome-wide significant
 74 ($p < 5 \times 10^{-8}$) peaks in Europe, East Asia, and the Americas respectively, with none identified in South
 75 Asia or Africa (Fig. 1B-F, Supplementary Table 1, Supp. Fig. S1). The differing numbers of peaks
 76 reflect the relative power we have in each region; simulations show that we have the most power to
 77 detect selection in Europe—though still only around 50% for a selection coefficient of 0.02—and the least
 78 in Africa (Supplementary Note A). Comparison with other selection scans conducted using European
 79 data suggests that our ordering of loci is similar, and the main factor affecting the number of loci reported
 80 is the difference in the genome-wide significance threshold (Supplementary Note B).

81 The loci we detect include well-known signals of selection associated with diet, pigmentation and
 82 the immune system, as well as novel loci of unknown function (Supplementary Table 1, Supplementary
 83 Note C). Two loci were genome-wide significant in more than one region (*ADH1B* and *FADS1* in both
 84 Europe and East Asia), and twelve were genome-wide significant in one region and nominally significant
 85 (at $P < 0.05/31$) in at least one other. Variants passing FDR< 0.05 ($N = 215$) in Europe, where we had
 86 the highest power, are significantly enriched in the lower tail of the p-value distribution for all other
 87 non-African regions ($P = 9.0 \times 10^{-8}$, 1.5×10^{-5} , 9.2×10^{-11} and 0.45 for East Asia, the Americas,

88 South Asia, and Africa, respectively; Kolmogorov-Smirnov test). Overall, this suggests that a relatively
 89 high proportion of selected variants have shared signals across regions, reflecting either parallel selection
 90 on standing variation, selection on recurrent mutations at the same loci, or selection prior to the most
 91 recent shared ancestry.

92 Admixture-aware modeling of time series data

93 We extended our previous time-series modeling approach [bmws; 2] to take into account varying levels
 94 of admixture among individuals. We model allele frequencies evolving separately in each admixture
 95 component, allowing selection coefficients to vary over time and across admixture components in a
 96 regularized manner. We conducted simulations to validate the approach (Supplementary Text) then
 97 applied it to all significant loci. This allows us to identify changes in the strength of selection over time.
 98 For example, at *ADH1B*—an alcohol dehydrogenase gene where a variant associated with poor alcohol
 99 metabolism is known to have been under strong selection in East Asia [13–16]—we find a distinct increase
 100 in the selection coefficient between 100 and 150 generations ago (Fig. 2A-C). The *ADH1B* locus also
 101 shows evidence of selection in Europe (Supp. Fig. S2), also previously identified [8, 17]. However, the
 102 putatively causal East Asian variant (rs1229984) is rare in Europe and it remains unclear whether the
 103 phenotype under selection is shared.

104 At *LCT*, the strongly selected variant associated with adult lactase persistence in Europe [18], we
 105 infer strong selection up to about 100 generations ago, consistent with previous estimates of timing
 106 [2, 3] (Fig. 2D-F). Selection subsequently drops off into the present-day, consistent with several lines of
 107 evidence showing little or no recent selection [19] ,or phenotypic effects today [20, 21].

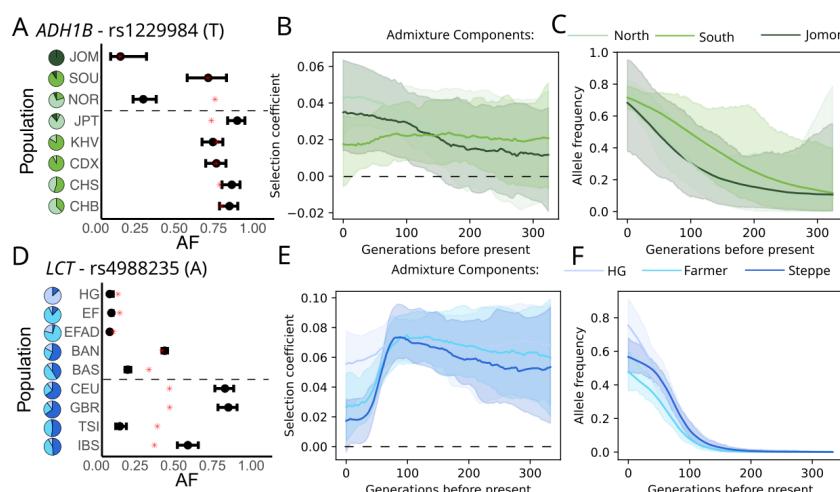


Figure 2: Ancestry-specific time series of the most strongly selected loci in East Asia and Europe. A) Observed allele frequencies; B) inferred selection coefficients; C) Allele frequency trajectories in admixture components at *ADH1B* in East Asia and D-F) *LCT* in Europe. We plotted the information for the derived allele of the lead SNP in each peak. In the frequency plots (first column), the bars represent 95% confidence intervals, while the red asterisk indicates the fitted frequency from the admixture model and the dashed line separates ancient (top) from present-day (bottom) groups. The pie charts indicate the average proportion of each admixture component for each population. See Methods for population abbreviations. In the plots of inferred selection coefficient and allele frequency (second and third columns), different colors distinguish the k source populations, solid lines represent fitted values, and shaded areas represent the 5-95% bootstrap confidence interval.

108 Parallel adaptation of fatty acid metabolism in Europe and East Asia

109 Selection for the same *FADS1* haplotype has been previously identified in both Europe [1] and East
110 Asia [22, 23]. The selected haplotype increases the rate of synthesis of plasma unsaturated fatty acids,
111 is believed to be an adaptation to a relatively plant-heavy agricultural diet [24], and appears to still be
112 under selection today in Europe [21]. We estimate that selection in Europe remained relatively constant
113 through time with a selection coefficient around 1% (Fig. 3A-C). In contrast, in East Asia, we find that
114 selection has intensified over time, with the allele frequency increasing substantially in the past 100
generations (Fig. 3D-F).

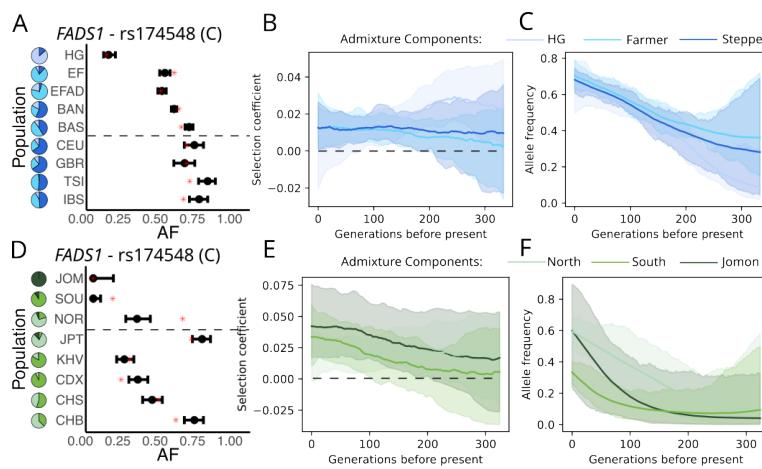


Figure 3: Details of the lead SNP at the *FADS1* locus in Europe and East Asia. A) Plot of observed Allele Frequencies, B) Inferred Selection coefficients, and C) AF trajectories in admixture components at *FADS1* in Europe, and in D-F) East Asia. We plotted the derived allele of rs174548, which is the lead SNP in Europe. Details of the plots are the same as in Fig. 2.

115

116 Selection on pigmentation in Europe

117 The SNP rs12913832 in *HERC2* is the major determinant of blue (versus brown) eye color through its
118 regulatory effects on *OCA2*, and is a known target of selection in Europe [25, 1]. We detect relatively
119 consistent selection and a persistent difference in frequency across ancestry groups (Fig. 4A-C), likely
120 reflecting the North-South frequency gradient that has persisted from the Mesolithic to the present-day
121 [26]. In contrast, at skin pigmentation loci *SLC45A2* and *SLC24A5*, we find consistent selection driving
122 allele frequencies almost to fixation. Selection on skin pigmentation is thought to reflect pressure for
123 increased vitamin D synthesis [27]. We find that selective pressure on *DHCR7* [28], which encodes a
124 key part of the vitamin D synthesis pathway intensified in the past 100 generations, consistent with our
125 previous findings from Britain [2] that selective pressure on vitamin D levels persisted or even intensified
126 into the past few thousand years.

127 Shared signals of selection related to immunity at the HLA and other loci

128 The human leukocyte antigen (HLA) region on chromosome 6 is highlighted in almost every selection
129 scan [29], due in part to the fact that it is one of the most gene-dense regions of the genome, containing
130 hundreds of phenotypic associations including many related to immune function [30]. However, it is
131 difficult to calibrate selection tests in the HLA region for a variety of reasons, including high levels of
132 linkage disequilibrium; a different population history than other portions of the genome (stemming from,
133 e.g. the effects of balancing selection), and the fact that high SNP density can lead to spurious visual

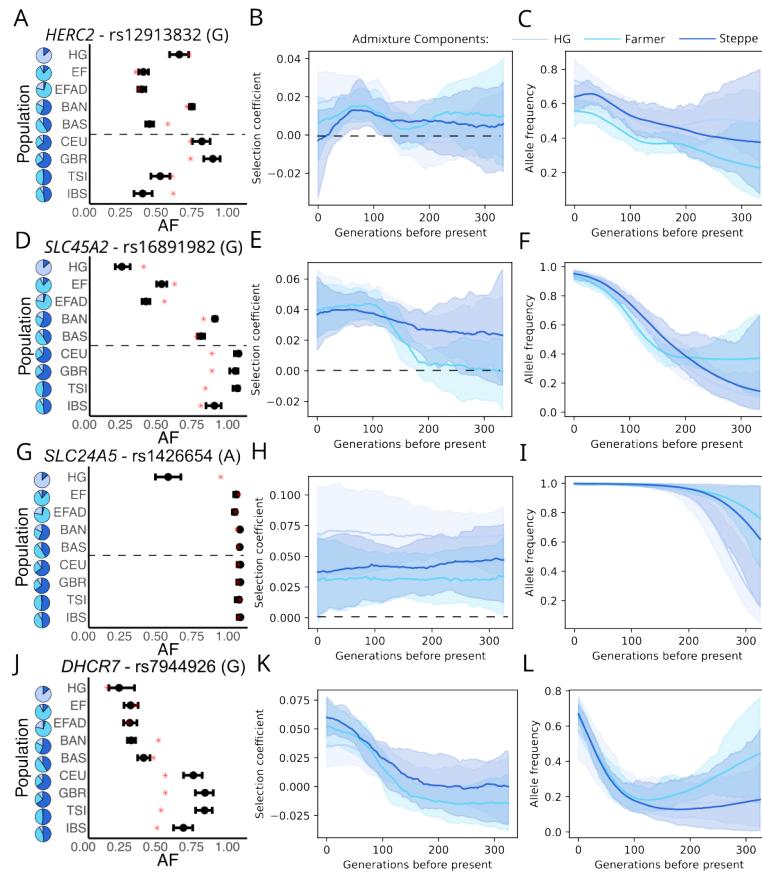


Figure 4: Details of the *HERC2*, *SLC45A2* and *SLC24A5* loci in Europe. A) Plot of observed Allele Frequencies, B) Inferred Selection coefficients, and C) AF trajectories in admixture components at *Herc2*, followed by D-F) *SLC45A2*, G-I) *SLC24A5*, and J-L) *DHCR7*. We plotted the information for the derived allele of the lead SNP in each peak. Details of the plots are the same as in Fig. 2.

134 peaks in genome-wide scans. Therefore, when testing for selection in the HLA, we utilized a separately
135 calibrated empirical null distribution, and also performed a 50kb window-based Bonferroni correction to
136 control for SNP density (Figure 5).

137 This approach yielded four HLA-wide significant loci: at *HLA-DQB1* in both Europe and the Americas,
138 replicating in East Asia; *HLA-B* in Europe and the Americas; *NOTCH4* in Europe; and *HLA-F*
139 in the Americas, replicating in East Asia. This suggests that these loci have been repeatedly targeted
140 by selection in different populations, either because of shared selective pressures related to agriculture
141 or other environmental transitions, or that they are common targets of selective sweeps responding to
142 novel pathogen exposures. For example, *HLA-DQB1* has previously been identified as a target of selec-
143 tion in Europe [8]. Our analysis suggests that the same locus has been repeatedly targeted in different
144 populations (Supp. Fig. S3), although we lack the resolution to tell whether it is the same haplotype.
145 Outside the HLA, several of our other signals are also likely related to immune function including *CARD8*
146 (Europe, replicated in East Asia and America), *MXRA7* (America) and *MIF* (America) (Supp. Fig. S1,
147 S2; Fig. 2G-I).

148 Universal stabilizing selection on female waist-hip ratio

149 Most human complex traits are highly polygenic. To study these traits, we extended our allele frequency-
150 based test to detect polygenic selection. We first identify trait-associated alleles based on summary

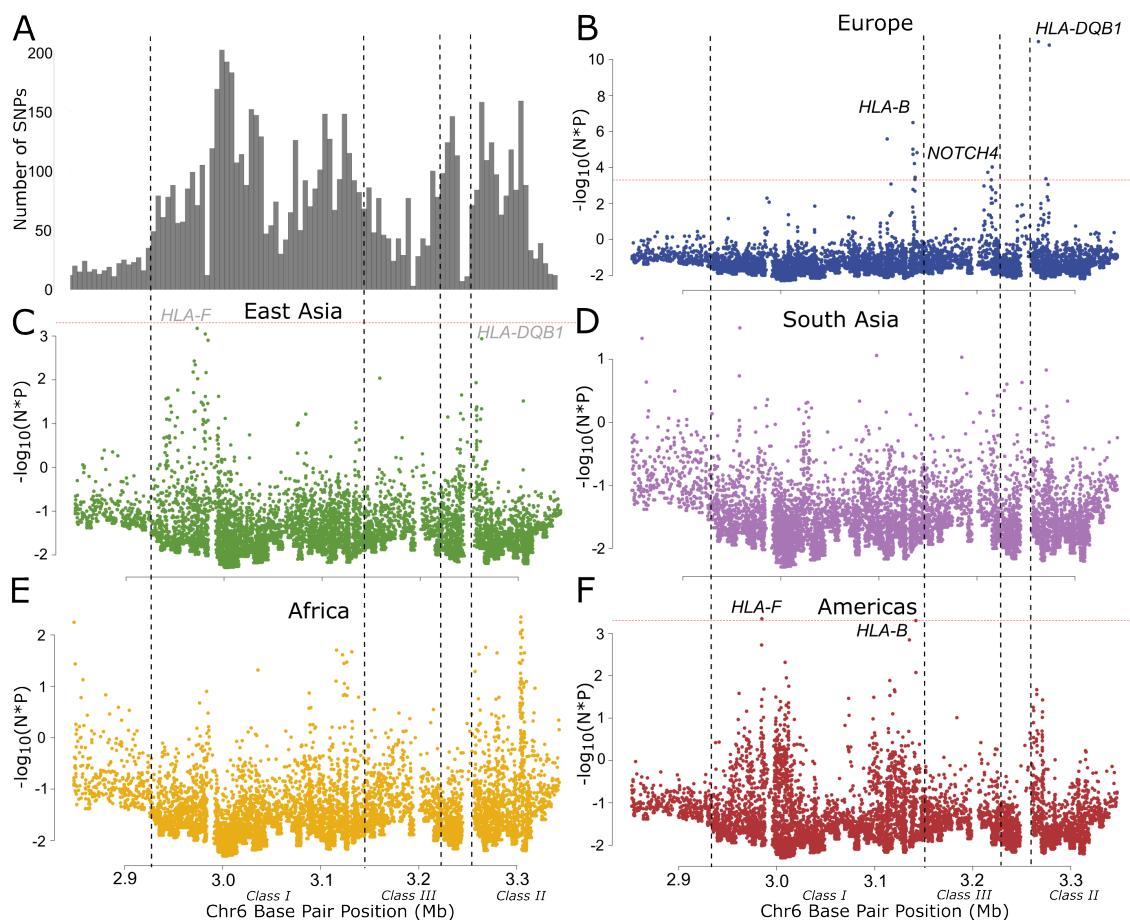


Figure 5: Manhattan plots corrected for SNP density focused on the MHC. We corrected for A) the distribution of SNPs across the MHC region in B) Europe, C) East Asia, D) South Asia, E) Africa, and F) the Americas. Each P-value in the manhattan plots is corrected by multiplying P by the number of SNPs in a tiled 50kb window, then the multiple testing threshold (red dashed line) corrects for the number of bins across the HLA. Dashed black boxes indicate the approximate locations of Class I, II and III HLA genes. Significant peaks are labelled in black with their nearest gene; grey labels indicate nominally-replicating sites.

151 statistics from genome-wide association studies (GWAS). For each ancient and modern population, we
 152 then count the proportion of trait increasing alleles, across all variants and individuals in each population.
 153 We treat this count as though it were an allele frequency, and perform the same likelihood ratio test,
 154 conditional on admixture proportions, that we used for the per-SNP analysis.

155 There are several advantages to this approach. First, since it only counts observed variants, it is
 156 insensitive to genotypes missing at random. Second, we only use effect directions, not effect sizes,
 157 which can be poorly estimated in GWAS (following [7]). Finally, we can compute a well-calibrated
 158 P-value by permuting the effect directions of variants. Small P-values indicate that the distribution of
 159 trait-associated variants is very different than expected based on genome-wide admixture proportions
 160 indicating directional selection in one or more populations. Conversely, large P-values indicate that
 161 the frequencies of trait-associated variants are “too close” to the genome-wide admixture proportions,
 162 indicating that the trait is under stabilizing selection.

163 A persistent problem with tests of polygenic selection is that they rely on variants identified in
 164 GWAS which can be confounded by population stratification, leading to false-positive signals of selection
 165 [31, 32]. One solution to this problem is to only perform selection tests on summary statistics obtained
 166 in a different population from the one on which the selection test is performed [33]. We therefore used

167 summary statistics obtained in European populations to test for selection in all populations except
 168 Europe (Fig. 6).

169 Testing 33 complex traits, we find a highly significant signal of stabilizing selection on waist-hip
 170 ratio (WHR; $P < 10^{-8}$), driven entirely by female waist-hip ratio (male WHR appears neutral; Fig. 6).
 171 We also find the same pattern when we test in Europe (Supp. Fig. S4; $P < 10^{-8}$). WHR has been
 172 observed to be under stabilizing selection in present-day populations in the UK Biobank [34]. These
 173 consistent results suggest that strong stabilizing selection on female WHR is a fundamental aspect of
 174 human biology, presumably due to an evolutionary conflict related to reproductive biology. We also find
 175 that Type 2 Diabetes is under significant stabilizing selection, which replicates when we use East Asian
 176 T2D summary statistics in non-East Asian populations (Supp. Fig. S5, S6, Supp. Table 2). Again,
 177 the basis for this is unclear, although the genetic correlation between T2D and birth weight—a classic
 178 example of a trait under stabilizing selection due to maternal-fetal conflict—is high, which might explain
 179 the signal.

180 We find significant directional selection on rheumatoid arthritis ($P = 1.5 \times 10^{-5}$) and a large but not
 181 significant signal of selection on tanning response ($P = 0.002$), both driven by East Asia (Supp. Table 3).
 182 If we test European summary statistics in Europe (Supplementary Fig. S4) we find directional selection
 183 on skin pigmentation, bone mineral density (BMD), body mass index (BMI) and mean corpuscular
 184 hemoglobin (MCH). Skin pigmentation does seem to reflect the effect of directional selection, as shown
 185 by other lines of evidence. The fact that we do not find it when testing non-European populations
 186 likely reflects the population-specificity of the loci driving adaptation in Europe. While BMD, BMI and
 187 MCH are all plausible targets of directional selection, we cannot rule out that these signals result from
 188 population stratification.

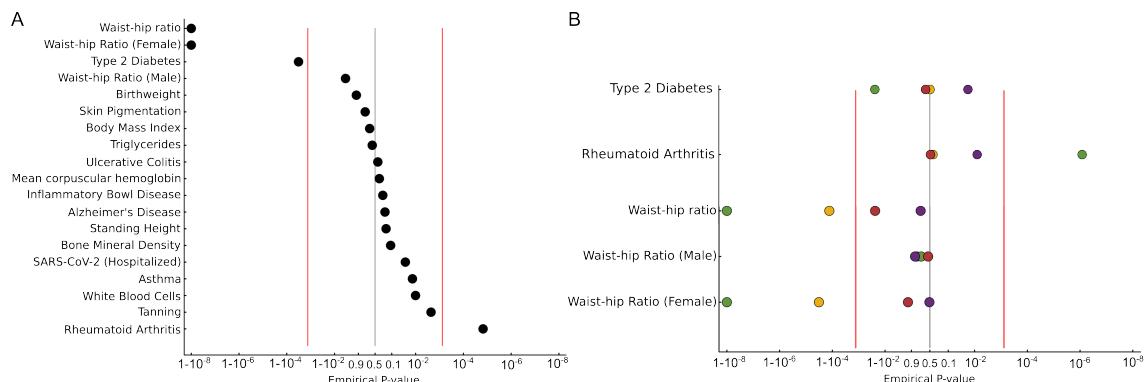


Figure 6: Several traits show evidence for polygenic selection. A) We tested for polygenic selection across SNPs from European-ancestry GWAS for 33 traits in jointly across 4 regions, excluding Europe to avoid the possibility of population stratification inducing false positives. B) Region-specific results for those traits with significant evidence for selection: Type 2 Diabetes, Rheumatoid Arthritis, and Waist-hip Ratio. The X-axis is plotted on a log scale, as $-\log_{10}(p) + \log_{10}(0.5)$ where $p < 0.5$ and $\log_{10}(1 - p) - \log_{10}(0.5)$ for $p > 0.5$. Red lines indicate a 2-sided Bonferroni correction threshold. 19 non-redundant traits are plotted here, see Supp. Fig. S4 for the complete results.

189 Discussion

190 Our results highlight several key aspects of recent natural selection in humans. First, we confirm that
 191 classic (i.e. strong, complete) selective sweeps are rare in recent human evolution [35]. Second, for many
 192 of the partial selective sweeps that do exist, the strength of selection varies with time, space and ancestry.
 193 Third, many signals of selection are shared between populations. In some cases (e.g. *FADS1*) the same

194 haplotype was selected in different populations, while in others (e.g. *ADH1B*, perhaps *HLA-DQB1*),
195 different haplotype at the same locus were selected. Similarly, the strongest signals of polygenic selection
196 we detected—of stabilizing selection on WHR and T2D—are largely shared across populations.

197 These shared signals of selection may reflect adaptation to shared selective pressures, for example
198 the adoption and development of agriculture independently in many parts of the world. For example,
199 the strongest signals of selection at individual loci (*LCT*, *ADH1B*, *FADS1*) are all plausibly related
200 to the consumption of agricultural products. Denser time series will allow greater temporal resolution
201 on the timing of selection and allow us to address the question of how tightly selection at different
202 loci was related to agriculture and other environmental changes. Shared signals may also reflect the
203 fact that some loci might be repeatedly targeted by sweeps. For example, genes involved in immunity
204 might experience repeated sweeps, or parallel sweeps in different populations, even if different pathogens
205 are involved. Finally, some shared signals might reflect imperfect modeling. For example, while the
206 European pigmentation locus *SLC24A5* nominally replicates in South Asia, that is likely due to recent
207 shared ancestry between Europe and South Asia which our approach does not model correctly due to
208 the limited South Asian ancient DNA sample size, rather than a real signal of independent selection in
209 South Asia. Indeed, time series analysis shows the selection coefficient s is not significantly different
210 from zero at any point in time after controlling for ancestry (Supp. Fig. S2).

211 Skin pigmentation in Europe is well-known to have been under selection, probably in response to
212 differing UV exposure at higher latitudes and the downstream effects on Vitamin D biosynthesis [27, 36].
213 We confirm this result, but we do not see such a clear signal in other populations. For example, in East
214 Asia, none of the genome-wide significant loci are clearly related to pigmentation and we see no polygenic
215 signal of selection on skin pigmentation using either European or East Asian summary statistics. We do
216 see a suggestive signal of polygenic selection on tanning response, raising the hypothesis that adaptation
217 to latitude in East Asia focused on plastic response rather than constitutive skin pigmentation as in
218 Europe.

219 The immune system is another key target of selection. While some HLA loci seem to be adaptive
220 in multiple regions, others appear to be population-specific, as are several of the non-HLA immune loci
221 we found. This is also reflected in the significant directional selection on loci influencing Rheumatoid
222 Arthritis in East Asia. RA is an autoimmune disease with a large role played by inflammatory pathway
223 including HLA genes [37]. While RA itself is likely not the source of selective pressure, it is an example
224 of a clinical trait being influenced by evolution; pro-inflammatory signals may be beneficial in the case
225 of an infection, but cause problems in other contexts.

226 We find that female waist-hip ratio (WHR) is consistently under stabilizing selection. However,
227 empirically many populations have higher female WHR than might be expected given the fact that lower
228 female WHR is generally associated with increased fertility. The advantage of a higher WHR is unclear
229 although it may lead to better health in some contexts (e.g. in times of resource scarcity) [38]. The
230 sex-specific selective pressure probably explains why the genetic correlation between sexes is lower for
231 WHR than for almost any other complex trait [39, 40].

232 The main limitation of our study is that we are still relatively underpowered, particularly outside
233 Europe. Even in Europe, power drops off rapidly for selection coefficients between 1-2%. While the
234 number of loci we detect is broadly consistent with other approaches using comparable significance
235 thresholds [1, 3, 5–7], it seems likely that there were many more variants under selection, with selection
236 coefficients less than 1%. Indeed, some other scans (e.g. [8]) find exactly that (Supplementary Note
237 C). However, the questions of how accurately those sweeps can be identified, how much they are shared
238 across populations, and how much they target *de novo* or standing variation remain unresolved.

239 Materials and Methods

240 A variant-level test for selection

241 To test for selection while taking into account admixture history, we explicitly model relationships
 242 between ancient and modern populations (Supp. Fig. S7). Specifically, we use admixture proportions to
 243 calculate the expected allele frequencies for a variant, and test whether observed frequencies match that
 244 expectation [1]. Unlike the previous iteration of this method, we infer admixture proportions in both
 245 ancient and modern populations relaxing the assumption that we have samples from unadmixed source
 246 populations and more accurately reflecting the fact that ancient populations were themselves products
 247 of admixture of previous peoples.

248 We quantify these relationships in a likelihood framework. The log-likelihood for a given variant
 249 under the alternative hypothesis that allele frequencies across populations are unconstrained is

$$\ell_1 = \sum_{i=1}^{A+B} x_i \log(f_i) + (N_i - x_i) \log(1 - f_i)$$

250 where A and B are the number of ancient and present-day populations respectively, x_i , N_i , and $f_i = x_i/N_i$
 251 are the observed alternate allele count, total allele counts and allele frequency for the variant in population
 252 i . To obtain the likelihood under the null hypothesis that allele frequencies are constrained, ℓ_0 , we
 253 calculate the expected frequencies in our set of populations as:

$$(f_1, \dots, f_{A+B}) = M \cdot Q$$

where

$$M = \begin{bmatrix} P_1^1 & \dots & P_1^k \\ \vdots & & \vdots \\ P_A^1 & \dots & P_A^k \\ P_{A+1}^1 & \dots & P_{A+1}^k \\ \vdots & & \vdots \\ P_{A+B}^1 & \dots & P_{A+B}^k \end{bmatrix}$$

and

$$Q = (q_1, \dots, q_k),$$

254 where M is a $(A+B) \times k$ matrix containing the admixture proportions for the $A+B$ ancient and present-
 255 day populations for each of the k source populations, and Q is a vector of length k containing the allele
 256 frequencies of the variant in each of the source populations. We estimate M using ADMIXTURE v1.3.0
 257 [12], after conducting LD pruning (200kb windows, 25-variant step, $r^2 > 0.4$). For the non-European
 258 samples, we ran ADMIXTURE in unconstrained mode, across a range of source population numbers
 259 from $K = 2$ to $K = 6$ and used the K with the minimum cross-validation error. (Within Europe we
 260 fixed $K = 3$ because of well-established evidence for three source populations.) Conditional on M , we
 261 fit Q to maximize

$$\ell_0 = \sum_{i=1}^{A+B} x_i \log(f_i) + (N_i - x_i) \log(1 - f_i),$$

262 thereby obtaining both the null likelihood and the expected frequencies in our source populations. Q is
 263 equivalent to the Q -matrix generated by ADMIXTURE but we fit it numerically for consistency with

264 the calculation of the alternative likelihood.

$$\ell_0 = \sum_{i=1}^{A+B} x_i \log(\hat{f}_i) + (N_i - x_i) \log(1 - \hat{f}_i).$$

265 The value of $2|\ell_1 - \ell_0|$ follows a χ^2 distribution with $A + B - k$ degrees of freedom. We additionally
266 controlled for inflation due to genetic drift or unmodeled ancestry variation using genomic control [41],
267 dividing the χ^2 statistics by a constant so that the median statistic matches the theoretical value, and
268 calculate a P-value based on the χ^2_{A+B-k} distribution.

269 We ran this scan for 1,150,640 variants captured by the 1240k reagent commonly used for ancient
270 DNA data [10, 11]. Because we were combining genetic data across data sets, technical artifacts from
271 one dataset could drive false positives in this scan. We therefore filtered all variants with FDR > 0.1
272 by requiring that at least one additional variant within 200kb and high LD ($r^2 > 0.5$) also pass that
273 FDR threshold. We calculated LD for this purpose using the 1kG populations corresponding to each
274 region. We then focused on signals passing a genome-wide Bonferroni multiple testing correction with
275 $p < 5 \times 10^{-8}$. We plotted individual significant signals using LocusZoom [42].

276 For each significant signal, we tested for replication in the other regions both of the lead variant and
277 of the locus more generally. To test if the lead variant replicated we tested for all lead variants in a
278 specific region by examining their P-values in each other region, using a Bonferroni correction for the
279 number of lead variants in that region. To test for the general locus, we identified the minimum p-value
280 within a megabase of the lead SNP, then used a Bonferroni correction for the number of SNPs tested in
281 the window to determine replication.

282 Population samples and admixture models

283 We tested for selection in 5 regions that had genetic representation of both ancient and present-day
284 populations: Africa, East Asia, South Asia, Europe and Central/South America (Fig. 1A). Most data
285 were obtained from the Allen Ancient DNA Resource (AADR) [43], with additional samples from the
286 HGDP [44] where necessary. For each region, we projected samples into PCA space constructed us-
287 ing samples genotyped on the Human Origins array from the same regions, then identified groups of
288 genetically-similar ancient samples that encompassed the present-day populations. We defined those
289 groups by drawing polygons around them in PCA space. As most ancient samples were pseudohaploid,
290 we used pseudohaploid versions of samples from the 1000 Genomes project, reported in the AADR.

291 In Africa, we included the Yoruba (YRI; $N = 101$) and Luhya (LWK; $N = 101$) populations from
292 the Thousand Genomes Project (1kG) [45], as well as Biaka from the Human Genome Diversity Project
293 (HGDP; BIAK; $N = 21$) [44]. We identified 3 distinct ancient groups corresponding to Central Africa
294 (CENT; $N = 33$), East Africa (EAST; $N = 149$), and those genetically similar to present-day Bantu
295 speakers (BANT; $N = 15$) (Supp. Fig. S8). We used $K = 3$ for ADMIXTURE modeling. The ancient
296 samples used here were originally published in [46–53]. To construct the PCA we used Human Origins
297 samples published in [52, 54–58].

298 For East Asia, we represented present-day populations with five from 1kG: Han Chinese from Beijing
299 (CHB; $N = 103$) and Southern China (CHS; $N = 106$), Chinese Dai (CDX; $N = 99$), Japanese (JPT;
300 $N = 104$), and Vietnamese (KHV; $N = 97$). We had 3 ancient groups corresponding to Northern China
301 (NOR; $N = 103$), Southern China (SOU; $N = 107$), and the Jomon in Japan (JOM; $N = 24$), and used
302 $K = 3$ for ADMIXTURE (Supp. Fig. S9). The ancient samples used here were originally published in
303 [59–72]. To construct the PCA we used Human Origins samples published in [54, 56, 73, 71].

304 For South Asia we used Gujarati (GIH; $N = 105$), Punjabi (PGL; $N = 96$), Telugu (ITU; $N = 103$),
305 Tamil (STU; $N = 99$), and Bengali (BEB; $N = 85$) populations from 1kG, and kept the ancient samples

306 in 1 group (ANCS; $N = 121$). We used $k=2$ for ADMIXTURE (Supp. Fig. S10). The ancient samples
307 used here were originally published in [74, 75]. To construct the PCA we used Human Origins samples
308 published in [54, 56, 71, 76].

309 For present-day European populations we used Italian Toscani (TSI; $N = 108$), Iberian (IBS; $N =$
310 103), British (GBR; $N = 92$) and Northwestern Europeans (CEU; $N = 99$) from 1kG, and split the
311 ancient samples into 5 groups. These were hunter-gatherers (HG; $N = 157$), early farmers (EF; $N = 605$),
312 admixed early farmers (EFAD; $N = 892$), northern Bronze Age (BAN; $N = 2113$), and southern Bronze
313 Age (BAS; $N = 1000$). We used $k=3$ for ADMIXTURE (Supp. Fig. S11). The ancient samples used
314 here were originally published in [1, 11, 26, 60, 75, 77–156]. To construct the PCA we used Human
315 Origins samples published in [54, 56, 57, 73, 157].

316 In Central/South America (shortened to 'Americas' in the manuscript), we used 1kG Peruvians (PEL;
317 $N = 69$) and HGDP Mayans (MAYA; $N = 18$) for our present-day populations, and included IBS to
318 represent the European ancestry present in many present-day Americans. We left the ancient samples in
319 one group (ANCA; $N = 216$), and used $k=2$ for ADMIXTURE (Supp. Fig. S12). The ancient samples
320 used here were originally published in [158–170]. To construct the PCA we used Human Origins samples
321 published in [54, 56].

322 Power Simulations

323 To estimate the power of our analysis, we conducted Wright-Fisher simulations of selection for each
324 of the five regions (Supp. Note A). For one variant in one population, we start the simulation at the
325 observed frequency, then draw the alleles in the next generation from a binomial distribution based on
326 a population of size N_e with a selection probability (p_{sel}) of

$$p_{sel} = \frac{f(1+s)}{1+fs}$$

327 where f is the frequency in the previous generation, and s is the strength of selection. For each population,
328 we randomly selected 1000 variants (constrained to MAF > 0.05 in that population), then simulated
329 their allele frequency trajectories for 0, 10, 20, 50, and 100 generations, under selection strengths varying
330 from $s = 0$ to $s = 0.1$. We then conducted the selection scan as for the real data, swapping in the
331 simulated frequency for that population for each SNP, and using the same genomic control and multiple
332 testing corrections as for the real scan.

333 Time series analysis

334 To analyze time series of allele frequencies while accounting for population continuity and admixture, we
335 extended our previous approach [2] to incorporate individual-specific admixture loadings. Following the
336 notation in that paper, let f_{tk} be the frequency of the derived allele in population k at time t generations
337 into the past, and similarly let s_{tk} be the strength of selection and N_{tk} the effective population size.
338 Given these, we assume as in the original paper that $f_{t-1,k} = F_{t-1,k}/[2N_{t-1,k}]$, where

$$F_{t-1,k} | f_t, s_t \sim \text{Binomial}(2N_{t-1,k}, f'_{tk}(s_{tk})) \quad (1)$$

339 and $f'_{tk}(s_{tk})$ is the allele frequency in population k immediately after mating (see equation (2) in [2].)

340 Suppose that we sample n_t individuals at time t . The genotype of sample $i \leq n_t$ at time t is denoted
341 $a_{ti} \in \{0, 1\}$, where without loss of generality we assume that each individual is haploid. (Diploid
342 individuals can be treated as two haploid individuals.) The population that this allele was inherited

343 from is denoted $z_{ti} \in \{1, \dots, K\}$. Hence,

$$a_{ti} | z_{ti} = k \sim \text{Bernoulli}(f_{tk}). \quad (2)$$

344 While we do not observe z_{ti} directly, we do assume knowledge of the overall proportion of alleles that
 345 were inherited from population k , denoted q_{tik} . Hence, $Q_t = (q_{tik}) \in [0, 1]^{n_t \times K}$ is exactly the loading
 346 matrix obtained by running ADMIXTURE [12] on the samples.

347 In the single population case, our original **bmws** model [2] worked by modeling the single-population
 348 allele frequency f_t as a mixture of beta distributions, with atoms at $f_t = 0$ and $f_t = 1$. However, it is
 349 challenging to extend this approach to the multi-population setting, because the joint distribution of f_t
 350 is a mixture of $(M + 2)^K$ components, representing all possible combinations of the M beta components
 351 and the two atoms from each population.

352 To avoid this combinatorial explosion, we instead used particle filtering [171], whereby a discrete set
 353 of atoms $\mathcal{P}_t \subset \{0, \dots, 2N\}^K, |\mathcal{P}_t| = P$ are used to approximate the distribution of f_t at each time point.
 354 (For simplicity we assumed that the effective population size N is known and constant over time and
 355 populations, but this is not essential.) Each particle $p_{tj} \in \mathcal{P}_t, j = 1, \dots, P$ represents the number of
 356 derived alleles in each of the K populations, and the particles are updated over time by incorporating
 357 the information from n_t and Q_t at each time point. Each particle has an associated weight ω_{p_t} , and the
 358 distribution of f_t is approximated by the weighted empirical distribution of the particles:

$$f_t \sim \sum_{p_t \in \mathcal{P}_t} \omega_{p_t} \delta_{p_t/(2N)}.$$

The particle filter works by iterating over time points $t = T, T - 1, \dots, 1$, transitioning the particles
 from time t to time $t - 1$ by incorporating the effects of genetic drift and selection, and then updating
 the weights $\omega_{p_{t+1}}$ to ω_{p_t} by incorporating the information from the observed data. Concretely, given
 \mathcal{P}_{t+1} , we form \mathcal{P}_t by transitioning each particle according to the model described above:

$$p_t | p_{t+1} \sim \text{Binomial}(2N, p'_{t+1}),$$

359 where p'_{t+1} is the allele frequency in each population immediately after mating and selection when the
 360 allele counts are given by p_{t+1} .

Next, we incorporate the information from the data to update the weights. The likelihood of the
 data conditional on population membership z_{ti} is given by (2) above. Therefore,

$$\omega_{p_{t+1}} \propto \omega_{p_t} \prod_{i=1}^{n_t} f_{t,z_{t,i}}.$$

361 Repeating in this way for $t = T, T - 1, \dots, 1$, we obtain a set of particles \mathcal{P}_0 and weights ω_{p_0} that
 362 approximate the distribution of f_0 . Samples from the posterior distribution $f_{0:T} | a_{1:T}, Z_{1:T}$ can be
 363 obtained by sampling a particle p_0 from \mathcal{P}_0 according to the weights ω_{p_0} , and then tracing back the
 364 ancestry of that particle through time.

365 **Gibbs sampler for selection and allele frequency trajectories**

366 To sample from the posterior distribution of selection coefficients $s_{1:T}$ we use Gibbs sampling. We
 367 iteratively sample from the following conditional distributions:

$$f_{0:T} | a_{1:T}, Z_{1:T}, s_{1:T} \quad (3)$$

$$Z_{1:T} | a_{1:T}, f_{0:T}, Q_{1:T} \quad (4)$$

$$s_{1:T} | f_{0:T} \quad (5)$$

Sampling from the distribution (3) is done using the particle filter described above. Sampling from the distribution (4) is straightforward, as the z_{ti} are independent given $f_{0:T}$ and $Q_{1:T}$. Specifically, we have

$$P(z_{ti} = k | a_{ti}, f_{0:T}, Q_{1:T}) \propto q_{tik} f_{tk}^{a_{ti}} (1 - f_{tk})^{1-a_{ti}}.$$

368 Finally, to sample $s_{1:T}$ (5), we leverage that s_t is independent of all other quantities conditional on
 369 (f_t, f_{t-1}) . The conditional likelihood of s_t is given by (1) above, multiplied by the prior on s_t (see next
 370 section). Viewed as a function of s_t , the density (1) does not correspond to any known distribution, so
 371 we approximately sample from it using the Metropolis-adjusted Langevin algorithm [172].

372 **Regularization**

373 In the original **bmws** method, allele frequency trajectories were regularized via the smoothing penalty
 374 $\alpha \sum_{t=2}^T (s_t - s_{t-1})^2$ to prevent overfitting, where α is a tuning parameter controlling the amount of
 375 regularity. As $\alpha \rightarrow \infty$, the estimate \mathbf{s} shrinks to a constant value, implying no time-varying selection.

376 Here we incorporate an additional term that encourages between-population similarity in these tra-
 377 jectories, in recognition of the fact selective pressures within and among the various ancestral populations
 378 likely covaried, at least partially. To model this, we add an additional penalty term of the form

$$\beta \sum_{t=1}^T \sum_{i \neq j} (s_{ti} - s_{tj})^2$$

379 which penalizes the pairwise differences between the selection strength at each time point. As $\beta \rightarrow \infty$
 380 the model estimates shrink to a single common trajectory (note that this may still be time-varying,
 381 depending on α), implying that selection does not vary based on ancestry background.

382 To enable the method to adapt to the level of signal in the data, we adopted a fully Bayesian approach,
 383 placing hyperpriors on α and β . Specifically, we used independent Gamma priors $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$
 384 and $\beta \sim \text{Gamma}(a_\beta, b_\beta)$, where $a_{\alpha,\beta}$ and $b_{\alpha,\beta}$ are hyperparameters. Then, α and β are sampled as part
 385 of the Gibbs sampler described above according to standard updates:

$$\begin{aligned} \alpha | \mathbf{s} &\sim \text{Gamma} \left(a_\alpha + \frac{TK}{2}, b_\alpha + \frac{1}{2} \sum_{t=1}^T \sum_{k=1}^K (s_{tk} - s_{t-1,k})^2 \right) \\ \beta | \mathbf{s} &\sim \text{Gamma} \left(a_\beta + \frac{TK(K-1)}{2}, b_\beta + \frac{1}{2} \sum_{t=1}^T \sum_{i \neq j} (s_{ti} - s_{tj})^2 \right). \end{aligned}$$

386 We chose these hyperparameters by comparing the log likelihoods across multiple different alphas for
 387 each variant in each population, and chose one alpha that maximized the average log likelihood (or was
 388 very close to the maximum) for all regions (Supp. Fig. S13). For the results presented here, we used
 389 $\alpha = 2.0$ and $\beta = \alpha/100$.

390 Time Series Simulations

391 To characterise the performance of bmws-admix we developed a simulation framework allowing us to
392 mimic observed data to varying degrees (Supplementary Note D). We simulated our populations with
393 $k = 3$ ancestry components, since that was the number we used for our best-powered selection scans. We
394 then used a Wright-Fisher framework for each ancestry component independently, allowing for variable
395 s , then sampled admixed individuals at each time point to get the overall allele frequencies.

396 Test for polygenic selection

397 To test for selection across multiple variants contributing to a polygenic trait, we adapted the same
398 pipeline used for the locus-based test. Instead of counting the number of alternate alleles for a single
399 variant across populations, we counted the proportion of trait-increasing alleles across all significant
400 variants identified in a genome-wide association study (GWAS). For population i , N_i is the total number
401 of non-missing (pseudohaploid) genotypes across all individuals and significant variants, and x_i is the
402 total number of genotypes with the trait-increasing allele. The likelihood calculation is then exactly
403 the same as for the single variant test, but we calculate P-values based on a null distribution obtained
404 by permuting the direction of effects and recalculating the likelihood, rather than based on the χ^2
405 distribution.

406 We did this for 38 biometric and disease traits analysed in European ancestry individuals: Alzheimer's
407 Disease [173], Asthma [174], Covid-19 (3 degrees of severity; Freeze 7 of the HGI) [175], Coronary Artery
408 Disease (CAD) [176], Irritable Bowel Disease, ulcerative colitis (UC), Crohn's Disease [177], HDL levels,
409 LDL levels, triglycerides [178], Rheumatoid Arthritis (RA) [179], Type II Diabetes [180], and Waist-hip
410 ratio (all, male, and female) [181], estimated Glomerular Filtration Rate [182]. For the remaining traits
411 we used summary statistics calculated in European-ancestry individuals from the UKBB by the Neale Lab
412 (<http://www.nealelab.is/uk-biobank/>): Mean corpuscular hemoglobin (MCH), Mean corpuscular
413 volume (MCV), counts of blood cells (platelets, red and white blood cells), Bone Mineral Density, Sitting
414 and Standing Height, Waist Circumference, Hip Circumference, Skin Pigmentation, Tanning, Body Mass
415 Index (BMI), Birth weight and Age at Menopause.

416 For each trait, we filtered for variants present on the 1240k array, and clumped the association signals
417 using Plink [183]. Clumped SNPs were required to be less than 200 kb apart and have $r^2 > 0.4$. We
418 then counted trait-increasing alleles across all lead 1240k variants with $p < 1 \times 10^{-8}$ for each trait. We
419 calculated P-values by permuting the effect direction of each variant included 1,000,000 times, thereby
420 changing which counts were contributing to the trait-increasing alleles. The underlying allele frequencies
421 remained unchanged. We identified traits there showed significant directional and stabilizing selection
422 by calculating 2-sided Bonferroni multiple testing thresholds for N traits as $0.025/N$ for directional,
423 and $(1 - 0.025)/N$ for stabilizing selection, where N is the number of traits. This choice tends towards
424 being conservative, as we corrected for all traits tested, regardless of their redundancy of others (e.g. we
425 included 3 different metrics of SARS-CoV-2 occurrence).

426 We also repeated this analysis using 19 traits from GWAS in East Asian-ancestry cohorts: Bone
427 Mineral Density, Body Mass Index, HDL, LDL, Triglyceride levels, Standing Height, Platelet, Red Blood
428 Cell, White Blood Cell counts, and Waist-hip ratio [184]; Asthma, Mean Corpuscular Hemoglobin, Mean
429 corpuscular Volume, Rheumatoid Arthritis, Ulcerative Colitis [185]; Coronary Artery Disease, estimated
430 Glomerular Filtration Rate, Age at Menopause [186]; Type II Diabetes [187]; skin pigmentation [188]. In
431 many cases these GWAS were less powered than the European versions, so for 4 traits with fewer than 40
432 hits at the stricter threshold (Menopause, UC, Skin Pigmentation, WHR) we lowered the genome-wide
433 significance threshold to 1×10^{-6} to include enough variants for the selection scan.

434 Comparison with other selection scans

435 To determine how our results compare to those of other selection scans, we compared our European scan
436 (with the highest power and most other scans to compare) to 4 other scans: [1, 3, 6, 8]. For each set of
437 published scores we intersected the variants tested with those in our scan. Because Kerner et al. only
438 published a list of significant tag SNPs, we also pruned the significant results of the other three, as well
439 as ours, then calculated the enrichment of the hits in each scan in our top N% of peaks. We pruned
440 using Plink's clumping function with a window of 1Mb and $r^2 > 0.4$ [183]. See Supplementary Text for
441 further details and discussion.

442 Data and Code Availability

443 Selection scores and P-values are available through Zenodo (<https://doi.org/10.5281/zenodo.18166002>),
444 and all genotype data are previously published and publicly available. Scripts for parsing data files and
445 running polygenic and AF-based analyses are available from https://github.com/colbrall/eas_selection.
446 bmws-admix analysis and simulation code are available from https://github.com/jthlab/bmws_admix.

447 Acknowledgements

448 We thank Melinda Yang, Pontus Skoglund, Mary Lauren Benton, Sarah Fong, Ziyue Gao, and members
449 of the Gao and Mathieson labs for helpful discussion and feedback.

450 Funding

451 This project was supported by the National Human Genome Research Institute training grant T32HG009495
452 to the University of Pennsylvania (L.L.C.) and the National Institute of General Medical Sciences
453 R35GM133708 (I.M) and R35GM151145 (J.T.). The content is solely the responsibility of the authors
454 and does not necessarily represent the official views of the National Institutes of Health.

455 Author Contributions

456 L.L.C., J.T. and I.M. conceived and designed the study and wrote the manuscript. J.T. implemented
457 the time series analysis, and I.M. and L.L.C. analyzed data. L.L.C conducted all other analyses. All
458 authors edited and approved the manuscript.

459 Competing Interests

460 The authors report no conflicts of interest.

461 References

- 462 [1] Iain Mathieson, Iosif Lazaridis, Nadin Rohland, et al. Genome-wide patterns of selection in 230
463 ancient eurasians. *Nature*, 528:499–503, 2015.
- 464 [2] Iain Mathieson and Jonathan Terhorst. Direct detection of natural selection in bronze age britain.
465 *Genome Research*, 32:2057–2067, 2022. 10.1101/gr.276862.122.

- 466 [3] Evan K Irving-Pease, Alba Refoyo-Martínez, William Barrie, et al. The selection landscape and
467 genetic legacy of ancient eurasians. *Nature*, 625:312–320, 2024.
- 468 [4] Y. Souilmi, R. Tobler, A. Johar, et al. Admixture has obscured signals of historical hard sweeps
469 in humans. *Nat Ecol Evol*, 6(12):2003–2015, 2022.
- 470 [5] D. Pandey, M. Harris, N. R. Garud, and V. M. Narasimhan. Leveraging ancient dna to uncover
471 signals of natural selection in europe lost due to admixture or drift. *Nat Commun*, 15(1):9772,
472 2024.
- 473 [6] Gaspard Kerner, Anna-Lena Neethus, Quentin Philippot, et al. Genetic adaptation to pathogens
474 and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell genomics*, 3(2):100248,
475 2023.
- 476 [7] Megan K Le, Olivia S Smith, Ali Akbari, et al. 1,000 ancient genomes uncover 10,000 years of
477 natural selection in Europe. *bioRxiv*, page 2022.08.24.505188, 2022.
- 478 [8] Ali Akbari, Alison R Barton, Steven Gazal, et al. Pervasive findings of directional selection realize
479 the promise of ancient dna to elucidate human adaptation. *bioRxiv*, page 2024.09.14.613021, 2024.
- 480 [9] J. Y. Cheng, A. J. Stern, F. Racimo, and R. Nielsen. Detecting selection in multiple populations
481 by modeling ancestral admixture components. *Mol Biol Evol*, 39(1), 2022.
- 482 [10] Wolfgang Haak, Iosif Lazaridis, Nick Patterson, et al. Massive migration from the steppe was a
483 source for indo-european languages in europe. *Nature*, 522:207–211, 2015.
- 484 [11] Qiaomei Fu, Mateja Hajdinjak, Oana Teodora Moldovan, et al. An early modern human from
485 Romania with a recent Neanderthal ancestor. *Nature*, 524(7564):216–219, 2015.
- 486 [12] David H Alexander, John Novembre, and Kenneth Lange. Fast model-based estimation of ancestry
487 in unrelated individuals. *Genome Research*, 19:1655–1664, 2009. 10.1101/gr.094052.109.
- 488 [13] Yi Han, Sheng Gu, Hiroki Oota, et al. Evidence of Positive Selection on a Class I α em_iADH_i/em_i
489 Locus. *The American Journal of Human Genetics*, 80(3):441–456, 2007.
- 490 [14] Yi Peng, Hong Shi, Xue-bin Qi, et al. The ADH1B Arg47His polymorphism in East Asian popu-
491 lations and expansion of rice domestication in history. *BMC Evolutionary Biology*, 10(1):15, 2010.
- 492 [15] Saori Sakaue, Masato Akiyama, Makoto Hirata, et al. Functional variants in ADH1B and ALDH2
493 are non-additively associated with all-cause mortality in Japanese population. *European Journal
494 of Human Genetics*, 28(3):378–382, 2020.
- 495 [16] Ting-Gang Chang, Ting-Ting Yen, Chia-Yi Wei, et al. Impacts of ADH1B rs1229984 and ALDH2
496 rs671 polymorphisms on risks of alcohol-related disorder and cancer. *Cancer medicine*, 12(1):
497 747–759, 2023.
- 498 [17] Kevin J. Galinsky, Gaurav Bhatia, Po-Ru Loh, et al. Fast Principal-Component Analysis Reveals
499 Convergent Evolution of ADH1B in Europe and East Asia. *The American Journal of Human
500 Genetics*, 98(3):456–472, 2016.
- 501 [18] Todd Bersaglieri, Pardis C Sabeti, Nick Patterson, et al. Genetic signatures of strong recent positive
502 selection at the lactase gene. *American journal of human genetics*, 74:1111–1120, 2004.

- 503 [19] A. Krutli, A. Bouwman, G. Akgul, et al. Ancient dna analysis reveals high frequency of european
504 lactase persistence allele (t-13910) in medieval central europe. *PLoS One*, 9(1):e86251, 2014. ISSN
505 1932-6203 (Electronic) 1932-6203 (Linking). doi: 10.1371/journal.pone.0086251.
- 506 [20] Richard P Evershed, George Davey Smith, Mélanie Roffet-Salque, et al. Dairying, diseases and
507 the evolution of lactase persistence in Europe. *Nature*, 608(7922):336–345, 2022.
- 508 [21] I. Mathieson, F. R. Day, N. Barban, et al. Genome-wide analysis identifies genetic effects on
509 reproductive success and ongoing natural selection at the fads locus. *Nat Hum Behav*, 7(5):790–
510 801, 2023.
- 511 [22] C. Suo, H. Xu, C. C. Khor, et al. Natural positive selection and north-south genetic diversity in
512 east asia. *Eur J Hum Genet*, 20(1):102–10, 2012.
- 513 [23] C. W. K. Chiang, S. Mangul, C. Robles, and S. Sankararaman. A comprehensive map of genetic
514 variation in the world’s largest ethnic group-han chinese. *Mol Biol Evol*, 35(11):2736–2750, 2018.
- 515 [24] Sara Mathieson and Iain Mathieson. FADS1 and the timing of human adaptation to agriculture.
516 *Molecular Biology and Evolution*, 35(12):2957–2970, 2018.
- 517 [25] R. A. Sturm, D. L. Duffy, Z. Z. Zhao, et al. A single snp in an evolutionary conserved region
518 within intron 86 of the herc2 gene determines human blue-brown eye color. *Am J Hum Genet*, 82
519 (2):424–31, 2008.
- 520 [26] I. Mathieson, S. Alpaslan-Roodenberg, C. Posth, et al. The genomic history of southeastern europe.
521 *Nature*, 555(7695):197–203, 2018.
- 522 [27] Nina G Jablonski and George Chaplin. Colloquium paper: human skin pigmentation as an adap-
523 tation to UV radiation. *Proceedings of the National Academy of Sciences of the United States of
524 America*, 107 Suppl 2(Suppl 2):8962–8968, 2010.
- 525 [28] V. Kuan, A. R. Martineau, C. J. Griffiths, et al. Dhcr7 mutations linked to higher vitamin d
526 status allowed early human migration to northern latitudes. *BMC Evol Biol*, 13:144, 2013. ISSN
527 1471-2148 (Electronic) 1471-2148 (Linking). doi: 10.1186/1471-2148-13-144.
- 528 [29] Diogo Meyer, Vitor R C. Aguiar, Bárbara D Bitarello, et al. A genomic perspective on HLA
529 evolution. *Immunogenetics*, 70(1):5–27, 2018.
- 530 [30] C. J. Smith, S. Strausz, FinnGen, et al. Haplotype analysis reveals pleiotropic disease associations
531 in the hla region. *medRxiv*, 2024.
- 532 [31] J. J. Berg, A. Harpak, N. Sinnott-Armstrong, et al. Reduced signal for polygenic adaptation of
533 height in uk biobank. *Elife*, 8:e39725, 2019.
- 534 [32] M. Sohail, R. M. Maier, A. Ganna, et al. Polygenic adaptation on height is overestimated due to
535 uncorrected stratification in genome-wide association studies. *Elife*, 8:e39702, 2019.
- 536 [33] M. Chen, C. Sidore, M. Akiyama, et al. Evidence of polygenic adaptation in sardinia at height-
537 associated loci ascertained from the biobank japan. *Am J Hum Genet*, 107(1):60–71, 2020.
- 538 [34] Jaleal S Sanjak, Julia Sidorenko, Matthew R Robinson, et al. Evidence of directional and stabilizing
539 selection in contemporary humans. *Proceedings of the National Academy of Sciences*, 115:151–156,
540 2018.

- 541 [35] R. D. Hernandez, J. L. Kelley, E. Elyashiv, et al. Classic selective sweeps were rare in recent human
542 evolution. *Science*, 331(6019):920–4, 2011.
- 543 [36] Dan Ju and Iain Mathieson. The evolution of skin pigmentation associated variation in west
544 eurasia. *bioRxiv*, page 2020.05.08.085274, 2020.
- 545 [37] Laura E Dedmon. The genetics of rheumatoid arthritis. *Rheumatology*, 59(10):2661–2670, 2020.
- 546 [38] Elizabeth Cashdan. Waist-to-hip ratio across cultures: Trade-offs between androgen- and estrogen-
547 dependent traits. *Current Anthropology*, 49:1099–1107, 2008.
- 548 [39] R. Lande. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolu-*
549 *tion*, 34(2):292–305, 1980. ISSN 1558-5646 (Electronic) 0014-3820 (Linking). doi: 10.1111/j.1558-
550 5646.1980.tb04817.x. Lande, Russell eng 1980/03/01 Evolution. 1980 Mar;34(2):292-305. doi:
551 10.1111/j.1558-5646.1980.tb04817.x.
- 552 [40] Carrie Zhu, Matthew J Ming, Jared M Cole, et al. Amplification is the primary mode of gene-by-sex
553 interaction in complex human traits. *Cell Genomics*, 3:100297, 2023.
- 554 [41] B Devlin and Kathryn Roeder. Genomic control for association studies. *Biometrics*, 55:997–1004,
555 1999. <https://doi.org/10.1111/j.0006-341X.1999.00997.x>.
- 556 [42] Andrew P Boughton, Ryan P Welch, Matthew Flickinger, et al. LocusZoom.js: interactive and
557 embeddable visualization of genetic association study results. *Bioinformatics*, 37(18):3017–3018,
558 2021.
- 559 [43] Swapan Mallick, Adam Micco, Matthew Mah, et al. The allen ancient dna resource (aadr) a curated
560 compendium of ancient human genomes. *Scientific Data*, 11:182, 2024.
- 561 [44] Anders Bergström, Shane A McCarthy, Ruoyun Hui, et al. Insights into human genetic variation
562 and population history from 929 diverse genomes. *Science*, 367:eaay5012, 2020.
- 563 [45] The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*,
564 526:68–74, 2015.
- 565 [46] M Gallego Llorente, E R Jones, A Eriksson, et al. Ancient Ethiopian genome reveals extensive
566 Eurasian admixture throughout the African continent. *Science (New York, N.Y.)*, 350(6262):
567 820–822, 2015.
- 568 [47] Mark Lipson, Isabelle Ribot, Swapan Mallick, et al. Ancient West African foragers in the context
569 of African population history. *Nature*, 577(7792):665–670, 2020.
- 570 [48] Mark Lipson, Elizabeth A Sawchuk, Jessica C Thompson, et al. Ancient DNA and deep population
571 structure in sub-Saharan African foragers. *Nature*, 603(7900):290–296, 2022.
- 572 [49] Mary E. Prendergast, Mark Lipson, Elizabeth A. Sawchuk, et al. Ancient dna reveals a multistep
573 spread of the first herders into sub-saharan africa. *Science*, 6275(May):1–19, 2019.
- 574 [50] Carina M Schlebusch, Helena Malmström, Torsten Günther, et al. Southern African ancient
575 genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 2017.
- 576 [51] Kendra A Sirak, Daniel M Fernandes, Mark Lipson, et al. Social stratification without genetic
577 differentiation at the site of Kulubnarti in Christian Period Nubia. *bioRxiv*, 48109, 2021.

- 578 [52] Pontus Skoglund, Jessica C Thompson, Mary E Prendergast, et al. Reconstructing Prehistoric
579 African Population Structure. *Cell*, 171(1):59–71.e21, 2017.
- 580 [53] Ke Wang, Steven Goldstein, Madeleine Bleasdale, et al. Ancient genomes reveal complex patterns
581 of population movement, interaction, and replacement in sub-Saharan Africa. *Science Advances*, 6
582 (24):eaaz0183, 2020.
- 583 [54] Nick Patterson, Priya Moorjani, Yontao Luo, et al. Ancient Admixture in Human History. *Genetics*,
584 192(3):1065–1093, 2012.
- 585 [55] Joseph K Pickrell, Nick Patterson, Chiara Barbieri, et al. The genetic prehistory of southern Africa.
586 *Nature Communications*, 3(1):1143, 2012.
- 587 [56] Iosif Lazaridis, Nick Patterson, Alissa Mittnik, et al. Ancient human genomes suggest three ances-
588 tral populations for present-day Europeans. *Nature*, 513(7518):409–413, 2014.
- 589 [57] Iosif Lazaridis, Dani Nadel, Gary Rollefson, et al. Genomic insights into the origin of farming in
590 the ancient Near East. *Nature*, 536(7617):419–424, 2016.
- 591 [58] Deven N Vyas, Ali Al-Meeri, and Connie J Mulligan. Testing support for the northern and south-
592 ern dispersal routes out of Africa: an analysis of Levantine and southern Arabian populations.
593 *American Journal of Physical Anthropology*, 164(4):736–749, 2017.
- 594 [59] Niall P Cooke, Valeria Mattiangeli, Lara M Cassidy, et al. Ancient genomics reveals tripartite
595 origins of Japanese populations. *Science Advances*, 7(38):eabkh2419, 2025.
- 596 [60] Peter de Barros Damgaard, Nina Marchi, Simon Rasmussen, et al. 137 ancient human genomes
597 from across the Eurasian steppes. *Nature*, 2018.
- 598 [61] Pere Gelabert, Asta Blazyte, Yongjoon Chang, et al. Northeastern Asian and Jomon-related genetic
599 structure in the Three Kingdoms period of Gimhae, Korea. *Current Biology*, 32(15):3232–3244.e6,
600 2022.
- 601 [62] Choongwon Jeong, Ke Wang, Shevan Wilkin, et al. A Dynamic 6,000-Year Genetic History of
602 Eurasia’s Eastern Steppe. *Cell*, 183(4):890–904.e29, 2020.
- 603 [63] Hideaki Kanzawa-Kiriyama, Kirill Kryukov, Timothy A Jinam, et al. A partial nuclear genome of
604 the Jomons who lived 3000 years ago in Fukushima, Japan. *Journal Of Human Genetics*, 62:213,
605 2016.
- 606 [64] Maximilian Larena, Federico Sanchez-Quinto, Per Sjödin, et al. Multiple migrations to the Philip-
607 pines during the last 50,000 years. *Proceedings of the National Academy of Sciences*, 118(13):
608 e2026132118, 2021.
- 609 [65] Mark Lipson, Olivia Cheronet, Swapan Mallick, and Nadin Rohland. Ancient genomes document
610 multiple waves of migration in Southeast Asian prehistory. *Science*, 361(July):92–95, 2018.
- 611 [66] Xiaowei Mao, Hucai Zhang, Shiyu Qiao, et al. The deep population history of northern East Asia
612 from the Late Pleistocene to the Holocene. *Cell*, 184(12):3256–3266.e13, 2021.
- 613 [67] Hugh McColl, Fernando Racimo, Lasse Vinner, et al. The prehistoric peopling of Southeast Asia.
614 *Science*, 361(July):88–92, 2018.
- 615 [68] Chao Ning, Tianjiao Li, Ke Wang, et al. Ancient genomes from northern China suggest links
616 between subsistence changes and human migration. *Nature Communications*, 11(1):2700, 2020.

- 617 [69] Martine Robbeets, Remco Bouckaert, Matthew Conte, et al. Triangulation supports agricultural
618 spread of the Transeurasian languages. *Nature*, 599(7886):616–621, 2021.
- 619 [70] Tianyi Wang, Wei Wang, Guangmao Xie, et al. Human population history at the crossroads of
620 East and Southeast Asia since 11,000 years ago. *Cell*, 184(14):3829–3841.e21, 2021.
- 621 [71] Chuan-Chao Wang, Hui-Yuan Yeh, Alexander N Popov, et al. Genomic insights into the formation
622 of human populations in East Asia. *Nature*, 591(7850):413–419, 2021.
- 623 [72] Melinda A Yang, Xuechun Fan, Bo Sun, et al. Ancient DNA indicates human population shifts
624 and admixture in northern and southern China. *Science*, 369(6501):282–288, 2020.
- 625 [73] Choongwon Jeong, Oleg Balanovsky, Elena Lukianova, et al. The genetic history of admixture
626 across inner Eurasia. *Nature Ecology and Evolution*, 2019.
- 627 [74] Éadaoin Harney, Ayushi Nayak, Nick Patterson, et al. Ancient DNA from the skeletons of Roop-
628 kund Lake reveals Mediterranean migrants in India. *Nature Communications*, 10(1):3670, 2019.
- 629 [75] Vagheesh M Narasimhan, Nick Patterson, Priya Moorjani, et al. The formation of human popula-
630 tions in South and Central Asia. *Science*, 365(6457):eaat7487, 2019.
- 631 [76] Pontus Skoglund, Cosimo Posth, Kendra Sirak, et al. Genomic insights into the peopling of the
632 Southwest Pacific. *Nature*, 538(7626):510–513, 2016.
- 633 [77] Serena Aneli, Tina Saupe, Francesco Montinaro, et al. The Genetic Origin of Daunians and the
634 Pan-Mediterranean Southern Italian Iron Age Context. *Molecular Biology and Evolution*, 39(2):
635 msac014, 2022.
- 636 [78] Margaret L Antonio, Ziyue Gao, Hannah M Moots, et al. Ancient Rome: A genetic crossroads of
637 Europe and the Mediterranean. *Science*, 714(November):708–714, 2019.
- 638 [79] Margaret L Antonio, Clemens L Weiß, Ziyue Gao, et al. Stable population structure in Europe
639 since the Iron Age, despite high mobility. *eLife*, 13:e79714, 2024.
- 640 [80] Morten E Allentoft, Martin Sikora, Karl-Goran Sjogren, et al. Population genomics of Bronze Age
641 Eurasia. *Nature*, 522(7555):167–172, 2015.
- 642 [81] Carlos Eduardo G Amorim, Stefania Vai, Cosimo Posth, et al. Understanding 6th-century barbar-
643 ian social organization and migration through paleogenomics. *Nature Communications*, 9(1):3547,
644 2018.
- 645 [82] Selina Brace, Yoan Diekmann, Thomas J. Booth, et al. Ancient genomes indicate population
646 replacement in Early Neolithic Britain. *Nature Ecology and Evolution*, 3(5):765–771, 2019.
- 647 [83] Samantha Brunel, E Andrew Bennett, Laurent Cardin, et al. Ancient genomes from present-day
648 France unveil 7,000 years of its demographic history. *Proceedings of the National Academy of
649 Sciences*, 117(23):12791 – 12798, 2020.
- 650 [84] Joachim Burger, Vivian Link, Jens Blöcher, et al. Low Prevalence of Lactase Persistence in Bronze
651 Age Europe Indicates Ongoing Strong Selection over the Last 3,000 Years. *Current Biology*, 30
652 (21):4307–4315.e13, 2020.
- 653 [85] Lara M Cassidy, Rui Martiniano, Eileen M Murphy, et al. Neolithic and Bronze Age migration to
654 Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of
655 Sciences*, 113(2):368–373, 2016.

- 656 [86] Lara M. Cassidy, Ros Maoldúin, Thomas Kador, et al. A dynastic elite in monumental Neolithic
657 society. *Nature*, 582(7812):384–388, 2020.
- 658 [87] Ainash Childebayeva, Adam Benjamin Rohrlach, Rodrigo Barquera, et al. Population Genetics
659 and Signatures of Selection in Early Neolithic European Farmers. *Molecular Biology and Evolution*,
660 39(6):msac108, 2022.
- 661 [88] Florian Clemente, Martina Unterländer, Olga Dolgova, et al. The genomic history of the Aegean
662 palatial civilizations. *Cell*, 184(10):2565–2586.e21, 2021.
- 663 [89] Flavio De Angelis, Marco Romboni, Virginia Veltre, et al. First Glimpse into the Genomic Char-
664 acterization of People from the Imperial Roman Community of Casal Bertone (Rome, First-Third
665 Centuries AD). *Genes*, 13(1), 2022.
- 666 [90] Katharina Dulias, M George B Foody, Pierre Justeau, et al. Ancient DNA at the edge of the
667 world: Continental immigration and the persistence of Neolithic male lineages in Bronze Age
668 Orkney. *Proceedings of the National Academy of Sciences*, 119(8):e2108001119, 2022.
- 669 [91] S Sunna Ebenesersdóttir, Marcela Sandoval-Velasco, Ellen D Gunnarsdóttir, et al. Ancient genomes
670 from Iceland reveal the making of a human population. *Science*, 360(6392):1028 LP – 1032, 2018.
- 671 [92] Anne Friis-Holm Egfjord, Ashot Margaryan, Anders Fischer, et al. Genomic Steppe ancestry in
672 skeletons from the Neolithic Single Grave Culture in Denmark. *PLOS ONE*, 16(1):e0244872, 2021.
- 673 [93] D M Fernandes, D Strapagiel, P Borówka, et al. A genomic Neolithic time transect of hunter-farmer
674 admixture in central Poland. *Scientific Reports*, 8(1):14879, 2018.
- 675 [94] Daniel M Fernandes, Alissa Mittnik, Iñigo Olalde, et al. The spread of steppe and Iranian-related
676 ancestry in the islands of the western Mediterranean. *Nature Ecology & Evolution*, 4(3):334–345,
677 2020.
- 678 [95] Chris Fowler, Iñigo Olalde, Vicki Cummings, et al. A high-resolution picture of kinship practices
679 in an Early Neolithic tomb. *Nature*, 601(7894):584–587, 2022.
- 680 [96] Suzanne Freilich, Harald Ringbauer, Dženi Los, et al. Reconstructing genetic histories and social
681 organisation in Neolithic and Bronze Age Croatia. *Scientific Reports*, 11(1):16729, 2021.
- 682 [97] Rosa Fregel, Fernando L Méndez, Youssef Bokbot, et al. Ancient genomes from North Africa
683 evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proceedings of
684 the National Academy of Sciences*, 2018.
- 685 [98] Anja Furtwängler, A B Rohrlach, Thisseas C Lamnidis, et al. Ancient genomes reveal social and
686 genetic structure of Late Neolithic Switzerland. *Nature Communications*, 11(1):1915, 2020.
- 687 [99] Cristina Gamba, Eppie R. Jones, Matthew D. Teasdale, et al. Genome flux and stasis in a five
688 millennium transect of European prehistory. *Nature Communications*, 5:5257, 2014.
- 689 [100] Pere Gelabert, Ryan W Schmidt, Daniel M Fernandes, et al. Genomes from Verteba cave suggest
690 diversity within the Trypillians in Ukraine. *Scientific Reports*, 12(1):7242, 2022.
- 691 [101] Guido Alberto Gnechi-Ruscone, Anna Szécsényi-Nagy, István Koncz, et al. Ancient genomes
692 reveal origin and rapid trans-Eurasian migration of 7(th) century Avar elites. *Cell*, 185(8):1402–
693 1413.e21, 2022.

- 694 [102] Joscha Gretzinger, Duncan Sayer, Pierre Justeau, et al. The Anglo-Saxon migration and the
695 formation of the early English gene pool. *Nature*, 610(7930):112–119, 2022.
- 696 [103] Gloria González-Fortes, Eppie R. Jones, Emma Lightfoot, et al. Paleogenomic Evidence for Multi-
697 generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower
698 Danube Basin. *Current Biology*, 27(12):1801–1810.e10, 2017.
- 699 [104] G. González-Fortes, F. Tassi, E. Trucchi, et al. A western route of prehistoric human migration
700 from Africa into the Iberian Peninsula. *Proceedings of the Royal Society B: Biological Sciences*,
701 286(1895), 2019.
- 702 [105] Torsten Günther, Cristina Valdiosera, Helena Malmström, et al. Ancient genomes link early farmers
703 from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences*,
704 112(38):11917–11922, 2015.
- 705 [106] Torsten Günther, Helena Malmström, Emma M Svensson, et al. Population genomics of Mesolithic
706 Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS
707 Biology*, 16(1):e2003703, 2018.
- 708 [107] Éadaoin Harney, Olivia Cheronet, Daniel M Fernandes, et al. A minimally destructive protocol for
709 DNA extraction from ancient teeth. *Genome research*, 31(3):472–483, 2021.
- 710 [108] Zuzana Hofmanová, Susanne Kreutzer, Garrett Hellenthal, et al. Early farmers from across Europe
711 directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences*, 113
712 (25):6886–6891, 2016.
- 713 [109] Mari Järve, Lehti Saag, Christiana Lyn Scheib, et al. Shifts in the Genetic Landscape of the
714 Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance.
715 *Current Biology*, 29(14):2430–2441.e10, 2019.
- 716 [110] Theis Z.T. Jensen, Jonas Niemann, Katrine Højholt Iversen, et al. A 5700 year-old human genome
717 and oral microbiome from chewed birch pitch. *Nature Communications*, 10(1):1–10, 2019.
- 718 [111] Eppie R Jones, Gunita Zarina, Vyacheslav Moiseyev, et al. The Neolithic Transition in the Baltic
719 Was Not Driven by Admixture with Early European Farmers. *Current Biology*, 27(4):576–582,
720 2017.
- 721 [112] Maja Krzewińska, Gülsah Merve Kılınç, Anna Juras, et al. Ancient genomes suggest the eastern
722 Pontic-Caspian steppe as the source of western Iron Age nomads. *Science Advances*, 4(10):eaat4457,
723 2018.
- 724 [113] Maja Krzewińska, Anna Kjellström, Torsten Günther, et al. Genomic and Strontium Isotope
725 Variation Reveal Immigration Patterns in a Viking Age Town. *Current Biology*, 28(17):2730–
726 2738.e10, 2018.
- 727 [114] Thisseas C. Lamnidis, Kerttu Majander, Choongwon Jeong, et al. Ancient Fennoscandian genomes
728 reveal origin and spread of Siberian ancestry in Europe. *Nature Communications*, 9(1):1–12, 2018.
- 729 [115] Iosif Lazaridis, Alissa Mittnik, Nick Patterson, et al. Genetic origins of the Minoans and Myce-
730 naeans. *Nature*, 548(7666):214–218, 2017.
- 731 [116] Iosif Lazaridis, Songül Alpaslan-Roodenberg, Ayşe Acar, et al. The genetic history of the Southern
732 Arc: A bridge between West Asia and Europe. *Science*, 377(6609):eabm4247, 2022.

- 733 [117] Anna Linderholm, Gülsah Merve Kılıç, Anita Szczepanek, et al. Corded Ware cultural complexity
734 uncovered using genomic and isotopic analysis from south-eastern Poland. *Scientific Reports*, 10
735 (1):6885, 2020.
- 736 [118] Mark Lipson, Anna Szécsényi-Nagy, Swapan Mallick, et al. Parallel palaeogenomic transects reveal
737 complex genetic history of early European farmers. *Nature*, 551(7680):368–372, 2017.
- 738 [119] Helena Malmström, Torsten Günther, Emma M Svensson, et al. The genomic ancestry of the
739 Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon.
740 *Proceedings. Biological sciences*, 286(1912):20191528, 2019.
- 741 [120] Joseph H Marcus, Cosimo Posth, Harald Ringbauer, et al. Genetic history from the Middle
742 Neolithic to present on the Mediterranean island of Sardinia. *Nature Communications*, 11(1):
743 939, 2020.
- 744 [121] Nina Marchi, Laura Winkelbach, Ilektra Schulz, et al. The genomic origins of the world's first
745 farmers. *Cell*, 185(11):1842–1859.e18, 2022.
- 746 [122] Zoltán Maróti, Endre Néparáczki, Oszkár Schütz, et al. The genetic origin of Huns, Avars, and
747 conquering Hungarians. *Current Biology*, 32(13):2858–2870.e7, 2022.
- 748 [123] Hannah M Moots, Margaret Antonio, Susanna Sawyer, et al. A genetic history of continuity and
749 mobility in the Iron Age central Mediterranean. *Nature Ecology & Evolution*, 7(9):1515–1524, 2023.
- 750 [124] Ashot Margaryan, Daniel J Lawson, Martin Sikora, et al. Population genomics of the Viking world.
751 *Nature*, 585(7825):390–396, 2020.
- 752 [125] Rui Martiniano, Anwen Caffell, Malin Holst, et al. Genomic signals of migration and continuity in
753 Britain before the Anglo-Saxons. *Nature Communications*, 7:10326, 2016.
- 754 [126] Rui Martiniano, Lara M. Cassidy, Ros Ó'Maoldúin, et al. The population genomics of archaeological
755 transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-
756 based methods. *PLoS Genetics*, 13(7):1–24, 2017.
- 757 [127] Alissa Mittnik, Chuan-Chao Wang, Saskia Pfengle, et al. The genetic prehistory of the Baltic Sea
758 region. *Nature Communications*, 9(1):442, 2018.
- 759 [128] Alissa Mittnik, Ken Massy, Corina Knipper, et al. Kinship-based social inequality in Bronze Age
760 Europe. *Science*, 366(6466):731 LP – 734, 2019.
- 761 [129] Alexey G Nikitin, Peter Stadler, Nadezhda Kotova, et al. Interactions between earliest Linearband-
762 keramik farmers and central European hunter gatherers at the dawn of European Neolithization.
763 *Scientific Reports*, 9(1):19544, 2019.
- 764 [130] Mario Novak, Iñigo Olalde, Harald Ringbauer, et al. Genome-wide analysis of nearly all the victims
765 of a 6200 year old massacre. *PLOS ONE*, 16(3):e0247332, 2021.
- 766 [131] Iñigo Olalde, Hannes Schroeder, Marcela Sandoval-Velasco, et al. A Common Genetic Origin
767 for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Molecular
768 Biology and Evolution*, 32(12):3132–3142, 2015.
- 769 [132] Iñigo Olalde, Selina Brace, Morten E. Allentoft, et al. The Beaker phenomenon and the genomic
770 transformation of northwest Europe. *Nature*, 2, 2018.

- 771 [133] Iñigo Olalde, Swapan Mallick, Nick Patterson, et al. The genomic history of the Iberian Peninsula
772 over the past 8000 years. *Science*, 363(6432):1230 LP – 1234, 2019.
- 773 [134] Niall O’Sullivan, Cosimo Posth, Valentina Coia, et al. Ancient genome-wide analyses infer kinship
774 structure in an Early Medieval Alemannic graveyard. *Science Advances*, 4(9):eaao1262, 2018.
- 775 [135] Luka Papac, Michal Ernée, Miroslav Dobeš, et al. Dynamic changes in genomic and social structures
776 in third millennium BCE central Europe. *Science Advances*, 7(35):eabi6941, 2025.
- 777 [136] Nick Patterson, Michael Isakov, Thomas Booth, et al. Large-scale migration into Britain during
778 the Middle to Late Bronze Age. *Nature*, 601(7894):588–594, 2022.
- 779 [137] Cosimo Posth, Valentina Zaro, Maria A. Spyrou, et al. The origin and legacy of the Etruscans
780 through a 2000-year archeogenomic time transect. *Science Advances*, 7(39):1–16, 2021.
- 781 [138] Maïté Rivollat, Aline Thomas, Emmanuel Ghesquière, et al. Ancient DNA gives new insights into
782 a Norman Neolithic monumental cemetery dedicated to male elites. *Proceedings of the National
783 Academy of Sciences of the United States of America*, 119(18):e2120786119, 2022.
- 784 [139] Maïté Rivollat, Choongwon Jeong, Stephan Schiffels, et al. Ancient genome-wide DNA from
785 France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic
786 farmers. *Science Advances*, 6(22):eaaz5344, 2020.
- 787 [140] Lehti Saag, Liivi Varul, Christiana Lyn Scheib, et al. Extensive Farming in Estonia Started through
788 a Sex-Biased Migration from the Steppe. *Current Biology*, 27(14):2185–2193.e6, 2017.
- 789 [141] Lehti Saag, Margot Laneman, Liivi Varul, et al. The Arrival of Siberian Ancestry Connecting the
790 Eastern Baltic to Uralic Speakers further East. *Current Biology*, 29(10):1701–1711.e16, 2019.
- 791 [142] Tina Saupe, Francesco Montinaro, Cinzia Scaggion, et al. Ancient genomes reveal structural shifts
792 after the arrival of Steppe-related ancestry in the Italian Peninsula. *Current Biology*, 31(12):
793 2576–2591.e12, 2021.
- 794 [143] Federico Sánchez-Quinto, Helena Malmström, Magdalena Fraser, et al. Megalithic tombs in western
795 and northern Neolithic Europe were linked to a kindred society. *Proceedings of the National
796 Academy of Sciences*, 116(19):9469 LP – 9474, 2019.
- 797 [144] Stephan Schiffels, Wolfgang Haak, Pirita Paajanen, et al. Iron Age and Anglo-Saxon genomes from
798 East England reveal British migration history. *Nature Communications*, 7:10408, 2016.
- 799 [145] Christiana L. Scheib, Ruoyun Hui, Eugenia D’Atanasio, et al. East Anglian early Neolithic mon-
800 ument burial linked to contemporary Megaliths. *Annals of Human Biology*, 46(2):145–149, 2019.
- 801 [146] Hannes Schroeder, Ashot Margaryan, Marzena Szmyt, et al. Unraveling ancestry, kinship, and
802 violence in a Late Neolithic mass grave. *Proceedings of the National Academy of Sciences of the
803 United States of America*, pages 1–6, 2019.
- 804 [147] Andaine Seguin-Orlando, Thorfinn S Korneliussen, Martin Sikora, et al. Genomic structure in
805 Europeans dating back at least 36,200 years. *Science*, 346(6213):1113 LP – 1118, 2014.
- 806 [148] Marina Silva, Gonzalo Oteo-García, Rui Martiniano, et al. Biomolecular insights into North
807 African-related ancestry, mobility and diet in eleventh-century Al-Andalus. *Scientific Reports*,
808 11(1):18121, 2021.

- 809 [149] Pontus Skoglund, Helena Malmström, Ayça Omrak, et al. Genomic Diversity and Admixture
810 Differs for Stone-Age Scandinavian Foragers and Farmers. *Science*, 344(6185):747 – 750, 2014.
- 811 [150] Cristina Valdiosera, Torsten Günther, Juan Carlos Vera-Rodríguez, et al. Four millennia of Iberian
812 biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia.
813 *Proceedings of the National Academy of Sciences*, 115(13):3428 LP – 3433, 2018.
- 814 [151] Krishna R Veeramah, Andreas Rott, Melanie Groß, et al. Population genomic analysis of elongated
815 skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proceedings of the*
816 *National Academy of Sciences*, 115(13):3494 LP – 3499, 2018.
- 817 [152] Vanessa Villalba-Mouco, Marieke S van de Loosdrecht, Cosimo Posth, et al. Survival of Late
818 Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Current Biology*, 29(7):1169–
819 1177.e7, 2019.
- 820 [153] Vanessa Villalba-Mouco, Camila Oliart, Cristina Rihuete-Herrada, et al. Genomic transformation
821 and social organization during the Copper Age-Bronze Age transition in southern Iberia. *Science*
822 *Advances*, 7(47):eabi7038, 2021.
- 823 [154] He Yu, Marieke S. van de Loosdrecht, Marcello A. Mannino, et al. Genomic and dietary discontinuities
824 during the Mesolithic and Neolithic in Sicily. *iScience*, 25(5):104244, 2022.
- 825 [155] A. Žegarac, L. Winkelbach, J. Blöcher, et al. Ancient genomes provide insights into family structure
826 and the heredity of social status in the early Bronze Age of southeastern Europe. *Scientific Reports*,
827 11(1):10072, 2021.
- 828 [156] Pierre Zalloua, Catherine J Collins, Anna Gosling, et al. Ancient DNA of Phoenician remains
829 indicates discontinuity in the settlement history of Ibiza. *Scientific Reports*, 8(1):17567, 2018.
- 830 [157] Simone Andrea Biagini, Neus Solé-Morata, Elizabeth Matisoo-Smith, et al. People from Ibiza: an
831 unexpected isolate in the Western Mediterranean. *European journal of human genetics : EJHG*,
832 27(6):941–951, 2019.
- 833 [158] Jacob L Bongers, Nathan Nakatsuka, Colleen O’Shea, et al. Integration of ancient DNA with
834 transdisciplinary dataset finds strong support for Inca resettlement in the south Peruvian coast.
835 *Proceedings of the National Academy of Sciences*, 117(31):18359–18368, 2020.
- 836 [159] Marco Rosario Capodiferro, Bethany Aram, Alessandro Raveane, et al. Archaeogenomic distinctiveness
837 of the Isthmo-Colombian area. *Cell*, 184(7):1706–1723.e24, 2021.
- 838 [160] Constanza de la Fuente, María C Ávila-Arcos, Jacqueline Galimany, et al. Genomic insights
839 into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia.
840 *Proceedings of the National Academy of Sciences*, 115(17):E4006 LP – E4012, 2018.
- 841 [161] Douglas J Kennett, Stephen Plog, Richard J George, et al. Archaeogenomic evidence reveals
842 prehistoric matrilineal dynasty. *Nature Communications*, 8:14115, 2017.
- 843 [162] John Lindo, Randall Haas, Courtney Hofman, et al. The genetic prehistory of the Andean highlands
844 7000 years BP through European contact. *Science Advances*, 4(11):eaau4921, 2018.
- 845 [163] J Víctor Moreno-Mayar, Ben A Potter, Lasse Vinner, et al. Terminal Pleistocene Alaskan genome
846 reveals first founding population of Native Americans. *Nature*, 2018.

- 847 [164] Nathan Nakatsuka, Pierre Luisi, Josefina M B Motti, et al. Ancient genomes in South Patagonia
848 reveal population movements associated with technological shifts and geography. *Nature Communications*, 11(1):3868, 2020.
- 850 [165] Nathan Nakatsuka, Iosif Lazaridis, Chiara Barbieri, et al. A Paleogenomic Reconstruction of the
851 Deep Population History of the Andes. *Cell*, 181(5):1131–1145.e21, 2020.
- 852 [166] Cosimo Posth, Nathan Nakatsuka, Iosif Lazaridis, et al. Reconstructing the Deep Population
853 History of Central and South America. *Cell*, 175(5):1185–1197.e22, 2018.
- 854 [167] Danijela Popović, Martyna Molak, Mariusz Ziolkowski, et al. Ancient genomes reveal long-range
855 influence of the pre-Columbian culture and site of Tiwanaku. *Science Advances*, 7(39):eabg7261,
856 2021.
- 857 [168] Maanasa Raghavan, Matthias Steinrücken, Kelley Harris, et al. Genomic evidence for the Pleis-
858 tocene and recent population history of Native Americans. *Science*, 349(6250), 2015.
- 859 [169] C L Scheib, Hongjie Li, Tariq Desai, et al. Ancient human parallel lineages within North America
860 contributed to a coastal expansion. *Science*, 360(6392):1024 LP – 1027, 2018.
- 861 [170] Vera Tiesler, Jakob Sedig, Nathan Nakatsuka, et al. Life and death in early colonial Campeche:
862 new insights from ancient DNA. *Antiquity*, 96(388):937–954, 2022.
- 863 [171] Arnaud Doucet, Nando De Freitas, Neil James Gordon, et al. *Sequential Monte Carlo methods in
864 practice*, volume 1. Springer, 2001.
- 865 [172] Peter J Rossky, Jimmie D Doll, and Harold L Friedman. Brownian dynamics as smart monte carlo
866 simulation. *The Journal of Chemical Physics*, 69(10):4628–4633, 1978.
- 867 [173] Brian W Kunkle, Benjamin Grenier-Boley, Rebecca Sims, et al. Genetic meta-analysis of diagnosed
868 Alzheimer’s disease identifies new risk loci and implicates A β , tau, immunity and lipid processing.
869 *Nature Genetics*, 51(3):414–430, 2019.
- 870 [174] Yi Han, Qiong Jia, Pedram Shafiei Jahani, et al. Genome-wide analysis highlights contribution of
871 immune system pathways to the genetic architecture of asthma. *Nature Communications*, 11(1):
872 1776, 2020.
- 873 [175] Mari E K Niemi, Juha Karjalainen, Rachel G Liao, et al. Mapping the human genetic architecture
874 of COVID-19. *Nature*, 600(7889):472–477, 2021.
- 875 [176] Krishna G Aragam, Tao Jiang, Anuj Goel, et al. Discovery and systematic characterization of risk
876 variants and genes for coronary artery disease in over a million participants. *Nature Genetics*, 54
877 (12):1803–1815, 2022.
- 878 [177] Katrina M de Lange, Loukas Moutsianas, James C Lee, et al. Genome-wide association study
879 implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature
880 genetics*, 49(2):256–261, 2017.
- 881 [178] Sarah E Graham, Shoa L Clarke, Kuan-Han H Wu, et al. The power of genetic diversity in
882 genome-wide association studies of lipids. *Nature*, 600(7890):675–679, 2021.
- 883 [179] Kazuyoshi Ishigaki, Saori Sakaue, Chikashi Terao, et al. Multi-ancestry genome-wide association
884 analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nature Genetics*, 54(11):1640–
885 1651, 2022.

- 886 [180] Anubha Mahajan, Cassandra N Spracklen, Weihua Zhang, et al. Multi-ancestry genetic study of
887 type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nature*
888 *Genetics*, 54(5):560–572, 2022.
- 889 [181] Sara L Pulit, Charli Stoneman, Andrew P Morris, et al. Meta-analysis of genome-wide association
890 studies for body fat distribution in 694 649 individuals of European ancestry. *Human molecular*
891 *genetics*, 28(1):166–174, 2019.
- 892 [182] Kira J Stanzick, Yong Li, Pascal Schlosser, et al. Discovery and prioritization of variants and genes
893 for kidney function in ~1.2 million individuals. *Nature Communications*, 12(1):4350, 2021.
- 894 [183] Christopher C Chang, Carson C Chow, Laurent C A M Tellier, et al. Second-generation PLINK:
895 rising to the challenge of larger and richer datasets. *GigaScience*, 4(1):s13742–015–0047–8, 2015.
- 896 [184] Chia-Yen Chen, Tzu-Ting Chen, Yen-Chen Anne Feng, et al. Analysis across Taiwan Biobank,
897 Biobank Japan, and UK Biobank identifies hundreds of novel loci for 36 quantitative traits. *Cell*
898 *genomics*, 3(12):100436, 2023.
- 899 [185] Saori Sakaue, Masahiro Kanai, Yosuke Tanigawa, et al. A cross-population atlas of genetic associa-
900 tions for 220 human phenotypes. *Nature genetics*, 53(10):1415–1424, 2021.
- 901 [186] Masahiro Kanai, Jacob C Ulirsch, Juha Karjalainen, et al. Insights from complex trait fine-mapping
902 across diverse populations. *medRxiv*, page 2021.09.03.21262975, 2021.
- 903 [187] Cassandra N Spracklen, Momoko Horikoshi, Young Jin Kim, et al. Identification of type 2 diabetes
904 loci in 433,540 East Asian individuals. *Nature*, 582(7811):240–245, 2020.
- 905 [188] Beomsu Kim, Dan Say Kim, Joong-Gon Shin, et al. Mapping and annotating genomic loci to pri-
906 oritize genes and implicate distinct polygenic adaptations for skin color. *Nature Communications*,
907 15(1):4874, 2024.