The persistence and loss of hard selective sweeps amid ancient human admixture

Mariana Harris[1], Ziyi Mo[2,3], Adam Siepel[2,3] and Nandita Garud[4,5,*]

[1] Department of Computational Medicine, University of California Los Angeles, Los Angeles California, USA

[2] Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America

[3] School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America

[4] Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles California, USA

[5] Department of Human Genetics, University of California, Los Angeles, California, USA

[*] Correspondence to ngarud@ucla.edu

Abstract

The extent to which human adaptations have persisted throughout history despite strong eroding demographic events such as admixture, genetic drift, and fluctuations in selection pressures remains unknown. Understanding which loci are particularly resilient to such forces may shed light on the traits that were important for humans throughout multiple time periods. Yet, detecting ancient selection events is challenging from modern and ancient DNA due to the data and/or signal being severely degraded. Here we use a domain-adaptive neural network (DANN) trained on simulated data and applied to ancient and modern DNA for sweep detection. We show that the DANN can account for simulation misspecification, or discrepancies between the simulations and real aDNA, thereby improving the ability to detect sweeps in real data. Application of the DANN to more than 800 ancient and modern human genomes spanning the last 7000 years recovered 16 known sweeps at loci including *LCT*, *HLA*, *KITLG,* and *OCA2/HERC2*, and revealed 32 novel sweeps. All identified sweeps were classified as hard, consistent with historically low population sizes. While some sweeps were lost over time, 14 sweeps at loci involved in a range of functions including neuronal, reproductive, pigmentation, and signaling traits were found to persist from the most ancient time periods into the most recent time periods. Notably, the same top haplotype remained at high frequency across time at 9 of these 14 sweeps. Together, these results indicate that hard sweeps predominated in ancient human history and that several ancient selective events were resilient to strong admixture events and experienced sustained selective pressures.

Introduction

The growing availability of ancient DNA (aDNA) has revolutionized our ability to study how evolution has shaped human populations over the past ~12,000 years. The transition from mobile hunter-gatherer groups to sedentary, agriculture-based societies introduced profound selective pressures including shifts in diet, sustained contact with domesticated animals, and heightened pathogen exposure[1,2]. During this same period, repeated waves of migration and admixture among Western Hunter-Gatherers, Anatolian early farmers, and Steppe pastoralists continually reshaped the genetic landscape of Eurasia. These demographic events may have diluted or masked historical selective sweeps in present-day genomes[3], leaving the prevalence, persistence, and modes of selective sweeps across Eurasian populations largely unresolved.

Characterizing the targets as well as the mode and tempo of positive selection in aDNA can reveal the mechanisms and rate of human evolutionary change. However, the ability to detect adaptation in aDNA can be challenging for a number of reasons, including low read coverage, short read lengths, high levels of missing data, and the complex, often poorly characterized demographic history of human populations. Despite these challenges, several studies have shown that directional selection in humans may have been widespread[3–5]. However, these studies have largely been powered to detect classic 'hard' sweeps, in which a single adaptive variant rising to high frequency leaves behind a characteristic dip in diversity with a single dominant haplotype. In addition to hard sweeps, there may have been 'soft' sweeps, whereby multiple adaptive variants rise to high frequency simultaneously given large mutational inputs or abundant standing genetic variation (SGV) at the onset of selection[6,7]. Soft sweeps are more challenging to detect given that they leave behind more subtle signatures in the data due to there being multiple haplotypes rather than a single haplotype at high frequency[8,9] . Given the combination of data challenges, which can generate misleading signals that appear adaptive but actually stem from demographic forces or data artifacts[10–13], and the difficulty in detecting soft sweeps, it remains unknown how many historical sweeps have been missed and whether they were hard or soft.

Deep learning methods have emerged as a powerful tool in population genetics to address a wide variety of inference problems from genomic data including demographic inference[14,15] , estimating recombination rates[16,17] and detecting selection[16,18–21]. In particular, convolutional neural networks (CNNs) have proven particularly effective in detecting selective sweeps, largely

due to their ability to extract complex patterns from noisy, high-dimensional population genetic data. Notably, CNNs can natively handle multi-dimensional input, leveraging the full richness of raw genotype matrices without reducing it to a small set of summary statistics[16,18]. Despite their flexibility and strong performance on modern data and other organisms[16,19], deep learning methods have not yet been applied to aDNA to detect selection.

A key limitation to current approaches is that they rely on large amounts of labeled training data, which are typically generated through simulations based on simplified models that are restricted in their ability to fully capture the complexities of real genomic data. Discrepancies between simulated and real datasets can arise from inaccurate assumptions about demography, mutation and recombination rates, or from data artifacts. Additionally, modeling features such as large effective population sizes (N_e), heterogeneous recombination landscapes, or complex demographic histories can be computationally intensive[17,22], making such simulations impractical at the scale required for deep learning.

This mismatch between simulated training data and real genomic data, known as a simulation mis-specification[23], can reduce model accuracy and lead to inferences from data that are not robust. Several strategies have been proposed to address this issue including adaptive re-weighting of training examples[15,24,25]. Domain adaptive neural networks (DANNs)[26,27], have recently been proposed as another alternative to mitigate simulation mis-specification[26].

Domain adaptation aims to improve generalization by enabling a model trained on data from a source domain, in this case simulated data, to perform well on a target domain with different properties, such as real population genomic data[28]. This technique is widely applied in computer vision; for example, facial recognition models trained on high-quality studio images can be adapted to perform more reliably on lower-quality surveillance footage. It is also used in natural language processing, where models trained on reviews of books may require adaptation to accurately interpret sentiments in reviews of other products. In biology, domain adaptation has been used to predict transcription factor binding across distinct species[29]. Building on these applications, Mo and Siepel demonstrated that domain adaptation could also be leveraged to improve population genetic predictions, including detecting selective sweeps, inferring selection strength, and estimating recombination rates in the face of demographic misspecification[26] .

Given the unique challenges of aDNA, domain adaptation presents a powerful framework for characterizing selection across different periods of human history. Here we propose a novel

121 application of a DANN to distinguish between hard sweeps, soft sweeps, and neutrality in aDNA
122 and modern DNA. We find that hard sweeps were common throughout human history, consistent
123 with historically low population sizes, and that several sweeps have persisted over multiple time
124 periods, implying shared selective pressures over human history and resilience to major
125 demographic events.
126

127 ## Results

128 *Data*

129 In this work, we train a DANN to detect selective sweeps from 708 aDNA samples from
130 Europe that we analyzed previously[30], dated between ~7000 and 1345 years before present (BP)
131 (**Fig. 1A**,**B**). These samples are from populations that underwent major admixture events,
132 including the migration of Anatolian farmers into Europe and their admixture with local Mesolithic
133 hunter-gatherers around 8,500 BP, as well as the mixing of European farmers with steppe
134 pastoralists at the onset of the Bronze Age ~5000 BP[31] (**Fig. 1C**). This transitional period is
135 particularly important for studying adaptation, as it has been hypothesized that admixture has
136 obscured selective sweep signatures in modern humans and as a result the extent of selection has
137 likely been underestimated[3].
138

139 Based on direct radiocarbon dates and archaeological context, the samples were grouped
140 into four chronological periods with 177 samples per period as follows (**Fig. 1B**):
141

142 Neolithic (N): Individuals of European Hunter-Gatherer and Anatolian farmer ancestry dated
143 between 6500 and 5019 BP.
144 Bronze Age (BA): Individuals from the Bell Beaker cultures of Western and Central Europe, dated
145 between 4495 to 3808 BP.
146 Iron Age (IA): Individuals from Iron Age Britain and Western Europe dated between 3995 to 2350
147 BP.
148 Historic period (H): Individuals from Roman and late antique periods between 2300 to 1345 BP.
149

150 To ensure data quality, we only included samples for which a number of criteria could be
151 met (Methods), including requiring hybridization capture on at least 1.2 million positions, having

a minimum of 15,000 SNPs such that robust population genetic inferences could be performed, did not have significant contamination on the mtDNA or X chromosome (in males), were unrelated up to the third degree, and treated with the same Uracil-DNA Glycosylase process during library preparation. The last two bases were trimmed from each read to exclude the most damaged regions of aDNA. After selecting high quality samples, to have similar power to detect sweeps across different time periods, we chose the 177 hosts with highest coverage for each time period for further analysis, resulting in a total of 708 genomes (**Table S1**). These 177 hosts were subsequently down sampled per analysis window to the 150 individuals with the least amount of missing data.

Additionally, we analyzed 99 modern European individuals (CEU) from the 1000 genomes project[32]. We restricted the 1000 genomes samples to the 1.2 million positions that were in the aDNA capture array.

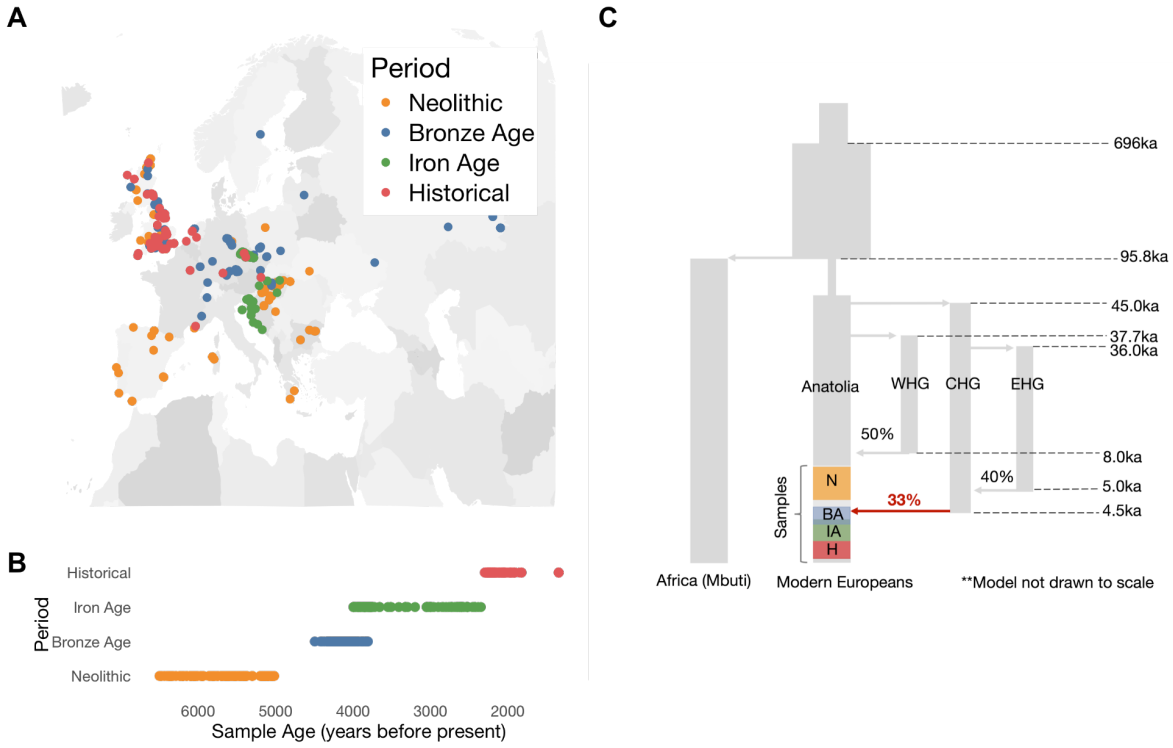

Figure 1. Location and age of samples included in study. (A) The locations of the 708 ancient human samples colored by their corresponding time periods. **(B)** Archeological or radiocarbon dates for each sample in years before present (BP). Each data point represents one sample and the colors indicate broader groupings according to four time periods. **(C)** Diagram of West Eurasian population history. Modern Europeans are composed of three main ancestries: Western Hunter-

170 Gatherers, Anatolian early farmers and Steppe pastoralists. The vertical grey segments represent
171 distinct population branches and arrows represent population splits and admixture events. Shown
172 are four main branches: the African (Mbuti) branch, Eurasian branch, West Hunter-Gatherers
173 (WHG), Caucasus Hunter-Gatherers (CHG) and Eastern Hunter-Gatherers (EHG). Highlighted are
174 the time ranges which our samples come from in the Eurasian branch. Also highlighted are the
175 admixture proportions and timing of events. The red arrow depicts a major admixture event
176 overlapping the timeframe during which our samples were collected. The full schematic of the
177 model with all associated parameters can be found in Souilmi et al. 2022 and original studies from
178 which these parameters were estimated[3,33–35] .
179
180 *Architecture of the DANN for sweep detection*
181 Before applying a DANN to aDNA, we tested its ability to (1) unlearn any differences
182 between simulated and real data including underlying demography or missing data rates and (2)
183 simultaneously distinguish between neutrality, hard sweeps, and soft sweeps. Since the goal of the
184 DANN is to both classify sweeps from neutrality and unlearn differences between domains, a
185 DANN differs from a more traditional neural network classifier by including not only a
186 classification branch, but also including a discriminator branch that distinguishes between a *source*
187 domain (e.g. simulations) and a *target* domain (e.g. real data) (**Figure 2**). One strategy for domain
188 adaptation, which we use here, is the addition of a gradient reversal layer (GRL)[27]. During
189 backpropagation, the sign of the gradient of the loss of the discriminator is reversed through the
190 GRL, penalizing features that discriminate between domains and promoting domain-invariant
191 features that are essential for sweep classification (Methods).
192 As input to the model, we provide images of haplotypes sorted by the frequency of most to
193 least common haplotype. Because of the low coverage nature of aDNA as well as ascertainment
194 bias in calling ancient SNPs, previous work has shown that heterozygous sites are not always
195 reliably determined. To address this issue, we used our previous approach[30] to 'pseudo haploidize'
196 the data by randomly selecting one of the reads mapping to a position and assigning the genotypes
197 of the read as the genotype of the sample at that site (Methods). This data was then used as input
198 to the DANN.
199 The images provided to the DANN are $n \times S$ bi-allelic genotypic matrices representing the
200 allelic states at S=201 segregating sites across n=150 pseudo haplotypes, downsampled from an

initial sample of 177 by retaining those with the least missing data (**Figure 2B**). These parameters reflect our previous findings[30] that a window size of 201 segregating sites, or approximately ~450 kb in aDNA, is adequate for capturing signals of selection in this dataset using haplotype homozygosity statistics. In addition, sorting haplotype images based on haplotype distances or frequencies has been shown to be essential for strong performance of CNN models[16,18,36]. We tested three different sorting approaches each designed to emphasize either elevated haplotype homozygosity typical of hard sweeps, or the presence of multiple high-frequency haplotypes characteristic of soft sweeps (Methods, **Fig. S1, Text S1**).

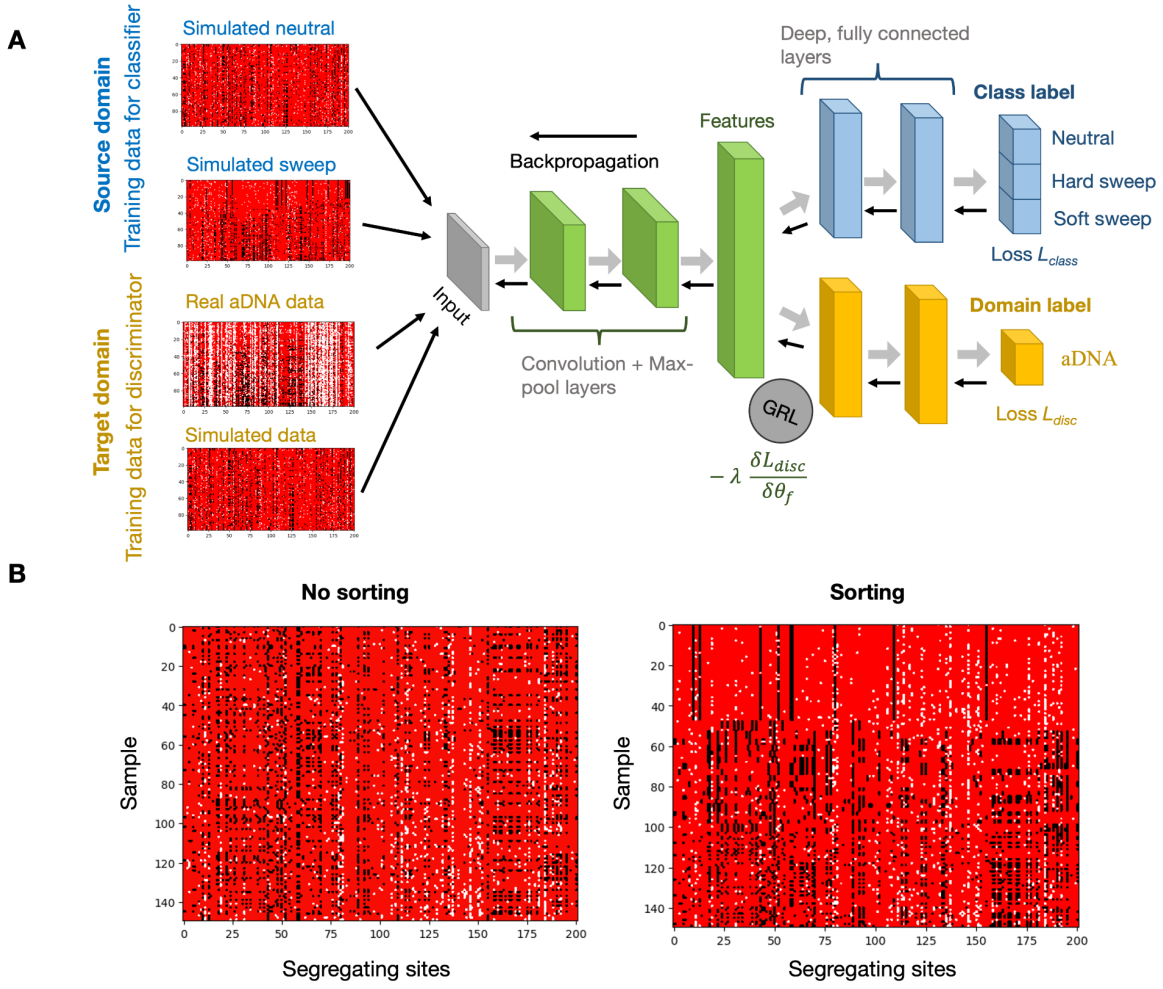

Figure 2. Domain Adaptive Neural Network (DANN) for detection of selective sweeps. (A) DANN Architecture. Haplotype images from both the source and target domains are passed through a series of convolutional layers, with dimensionality reduced by max-pooling steps

(green). The output is flattened into a feature vector, which is then processed by two branches of fully connected layers: the classifier, which predicts the sweep class (blue), and the discriminator, which distinguishes between domains (yellow). During backpropagation, the gradient reversal layer (GRL) inverts the loss from the discriminator, discouraging the model from differentiating between the domains and promoting domain-invariant features. The grey arrows indicate the forward pass and black arrows indicate backpropagation. **(B) Genomic data representation.** Genomic data is represented as images, with rows corresponding to sampled haplotypes and columns to segregating sites (201 SNPs). Each pixel represents the occurrence of a specific allele with major allele shown in red, minor allele in black, and missing data in white, coded as -1,1 and 0 respectively. On the left, we show a haplotype image of a simulated partial hard sweep where no sorting was applied. To the right we show the same image but sorting the rows (haplotypes) by frequency. Although the colors are used for visualization, the data is treated as a black and white image, where alleles are biallelic (major vs. minor).

Benchmarking DANN on simulated data

To assess the ability of the DANN to correct for simulation misspecification we evaluated the model using simulated data for both the source and target domains. We performed simulations varying the degree of mismatch between the target and source domains in terms of demography and missing data rates, as we expected these two variables to contribute the largest amount of discrepancy between real data and simulations. In all scenarios evaluated, the target domain, used as a proxy for real aDNA, consisted of simulations of a previously inferred admixture model describing ancient Europeans[3,33–35] (**Fig. 1C**) with missing data at the rate of 43% per base pair, reflecting the rates of missing data of the samples analyzed in this study (Methods, **Fig. S2**). To evaluate the ability of the DANN to correct for misspecification, we compared the performance of the DANN to that of a standard CNN lacking a discriminator branch but still having misspecified source and target domains. We evaluated the CNN on two test sets, one on a hypothetical best case scenario with matching target and source domains (hypothetical best case), and another one with mismatched domains (standard CNN, **Fig. S3**). This comparison allowed us to assess how much improvement the DANN could achieve under varying degrees of misspecification .

We found that for all simulation scenarios, the DANN outperformed the CNN, albeit modestly, when the target and source domains did not match, demonstrating its ability to mitigate

misspecification (**Fig. 3A, Fig S4**). Moreover, we found that the DANN could correct for mismatch in demography better than mismatch in missing data rates between target and source domains with the DANN performing slightly better in mitigating the most extreme misspecification in demography (constant N_e vs. Admixture; area under the precision-recall curve (AUPRC) of 0.814) compared to the most extreme misspecification in missing data rates (5% vs 43%; AUPRC =0.80) (**Fig. 3A**). This suggests that missing data introduces a more challenging form of misspecification, making it harder to correct using a domain-adaptive framework. Additionally, the DANN outperformed the haplotype homozygosity statistic H12[37], recently applied to aDNA data[30], with AUPRCs of 0.942 vs 0.898 for sweep detection, respectively (**Fig. 3B, Fig S5,6).** Finally, we found that in a model trained on a source domain of constant N_e with 43% missing data and target domain of admixture with 43% missing data, the DANN had an AUPRC of 0.956 for hard sweeps and 0.844 for soft sweeps (**Fig. S5**). We found similar performance when we trained with real aDNA for the target domain in order to explicitly unlearn the differences between simulated and real data (**Methods, Fig. S8**).

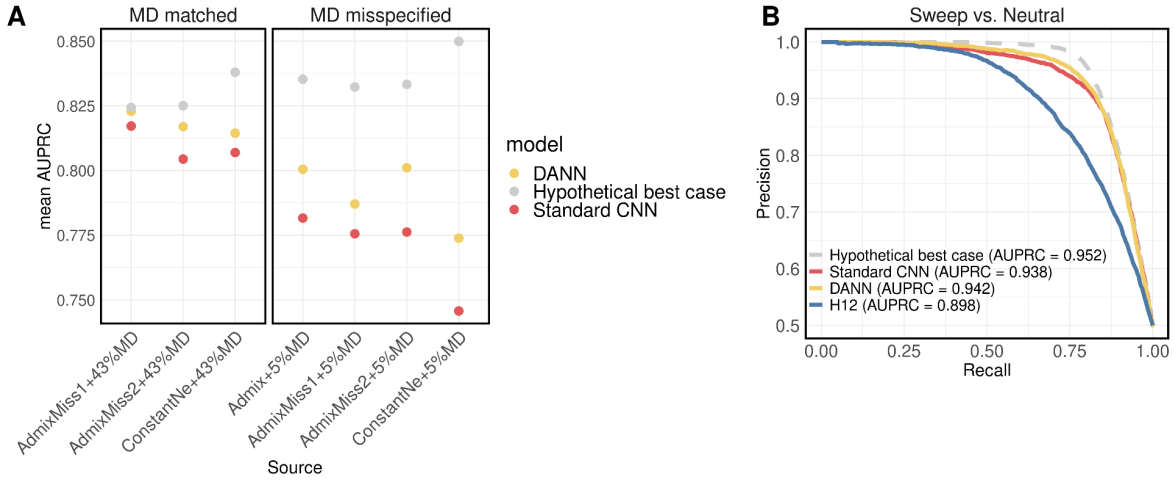

Figure 3. Domain adaptive neural network improves the detection of selective sweeps in simulated data that mimics aDNA. (A) AUPRC of the DANN (yellow). This is compared with a hypothetical best case simulation benchmark CNN representing an upper bound on performance. This benchmark was trained and tested on the source domain (grey). Additionally in red is a standard application of the CNN tested on a misspecified domain with mismatching demographic model and missing data (MD) compared to the source domain. **(B)** Precision-recall curve for

268 detection of sweeps (hard and soft) from neutrality using a constant N_e model as a source domain
269 and a human admixture model, both with an average of 43% missing data per site as a target
270 domain.
271
272 The ability to detect selective sweeps may vary with the strength of selection and the
273 softness of the sweep. To systematically evaluate the performance of the DANN across a range of
274 selection coefficients, we tested the model on weak (s~[0.005,0.05]) versus strong (s~[0.05,0.1])
275 selection, as well as on sweeps of varying softness modulated by rate of input of adaptive mutations
276 $\theta_A = 4N_e\mu_A$, where μ_A is the adaptive mutation rate. We found that when selection is strong,
277 irrespective of the softness of the sweep, we could distinguish sweeps from neutrality correctly
278 93% or more of the time (**Fig. S7**). However, when selection is weak, the ability to correctly
279 distinguish a sweep from neutrality drops to 60%, especially in the scenarios where sweeps are
280 extremely soft (θ_A=5). This result implies that the DANN has the greatest ability to detect sweeps,
281 either hard or soft, when selection is sufficiently strong.
282 We next asked how often soft sweeps are misclassified as hard sweeps and vice versa,
283 conditional on being distinguishable from neutrality (**Fig. S7**). We found that the majority of hard
284 sweeps (82%) are correctly classified as hard, with this percentage increasing to 93% for strong
285 selection. Similarly, we found that the majority of soft sweeps (~75%) are correctly classified as
286 soft, with this percentage increasing to 86% as θ_A increases to 5 when the signatures of hard vs
287 soft sweeps become most distinct.
288
289 *Application of the DANN to aDNA and modern humans*
290 Having confirmed the ability of the DANN to detect selective sweeps in simulations, we
291 trained a DANN using empirical aDNA as the target domain and simulations generated under the
292 admixture model shown in Figure 1 with a 43% missing data rate as the source domain. To identify
293 selective sweeps in aDNA, we next applied the DANN to genome-wide aDNA data across all time
294 periods. We used a sliding window of 201 SNPs, advancing each window by 10 SNPs. To ensure
295 predictions are well supported by more than one window, we averaged the predicted probabilities
296 every five consecutive, overlapping windows, generating the final class predictions (**Methods**).
297 Each window was assigned to the class with the highest predicted probability from the model's
298 multiclass output (hard sweep, soft sweep, or neutral). We show the resulting scan across the

aDNA time transect in **Fig. 4A**, where windows predicted to be hard sweeps are shown in red, soft in blue, and neutral regions in grey. In addition, In **Fig. 4B**, we show a scan on the modern European population (CEU). This scan was generated using a different model trained on modern human data as the target domain (Methods).

To avoid calling the same selective sweep multiple times, we grouped consecutive non-neutral windows into a single 'peak'. Additionally, to ensure distinct selective events, we further required peaks to be at least 1.5 Mb apart. Within each peak, we identified the representative window as the one with the highest sweep probability (the highest –log(probability of neutrality)).

We identified a total of 48 unique sweeps in ancient humans (**Fig 4A**) and 28 selective sweeps in modern humans. Among the 48 ancient human sweeps, over half the sweeps recurred across multiple periods with 5 detected in all periods, 8 in three, 12 in two, and 23 in only one (**Fig. S9**). Additionally, 18 ancient sweeps were found to persist in modern humans with 5 of these sweeps present in all ancient human time periods resulting in 58 unique peaks across all ancient and modern periods. To assess the robustness of our sweep inferences in ancient humans, we trained two additional models varying the source domain, and observed high concordance across models both in sweep detection and classification, with 44 sweeps overlapping in all three scans (**Fig. S10A**, **Fig. S11, Table S2**).

Out of the 48 sweeps detected in ancient humans (**Figure 4A)**, 16 overlap with recent studies[3–5,30,38–40]. Among these are 5 sweeps overlapping well known targets of selection highlighted in **Fig. 4**. The strongest signal across all time periods corresponds to the *HLA* region and neighboring gene *ZKSCAN3*. *HLA* encodes cell surface proteins that are involved in the adaptive immune system and has long been recognized as a target of selection[41,42] and *ZKSCAN3* is involved in transcriptional regulation of autophagy-related genes and was reported as under selection in Mathieson et al. 2015. We also detect a sweep spanning the *OCA2/HERC2* genes, which is associated with light eye color in Europeans[38]. We identify this signature across all four time periods, which was not observed in earlier scans across the same time transect[30,40]. Our scan also recovers a strong sweep signal at the *LCT* locus, which is associated with lactase persistence into adulthood[43,44]. This signal is restricted to the H period and the CEU population (**Fig. 4**), consistent with the rapid rise in frequency of the causal variant rs4988235 during this time. Notably, this allele was absent in Europe prior to the arrival of Steppe pastoralists in the Bronze Age and therefore could not have been under selection earlier[30,38,40,45–47]. We also highlight a sweep

330 overlapping the gene *KITLG*, associated with light hair and skin pigmentation and recently
331 identified as a target of selection, and *SLC22A4,* associated with Crohn's disease[4]. We summarize
332 the number of sweeps that overlap with previous work in **Table S3** and **Fig. S12,S13**.
333 The remaining 32 ancient sweeps we identify represent novel candidates that, to our
334 knowledge, have not been identified in previous studies. We do not consider peaks as novel
335 candidates if they are within 1Mb away from genes that have been previously identified as sweep
336 candidates[3–5,30,38–40]. To understand if the genes within these 32 peaks are enriched for any
337 functions, we assigned peaks to genes by annotating all protein coding genes within 300 Kb
338 distance upstream and downstream of the representative SNP of each peak using Ensembl Variant
339 Effect Predictor (VEP) (Supplementary table S5). We next performed an enrichment analysis for
340 previously identified genome-wide association study (GWAS) annotations on the set of mapped
341 genes using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)[48].
342 We observed an enrichment in many GWAS categories related to anthropometric traits as well as
343 disease and auto-immune related traits. These results are reported in **Fig. S14.**

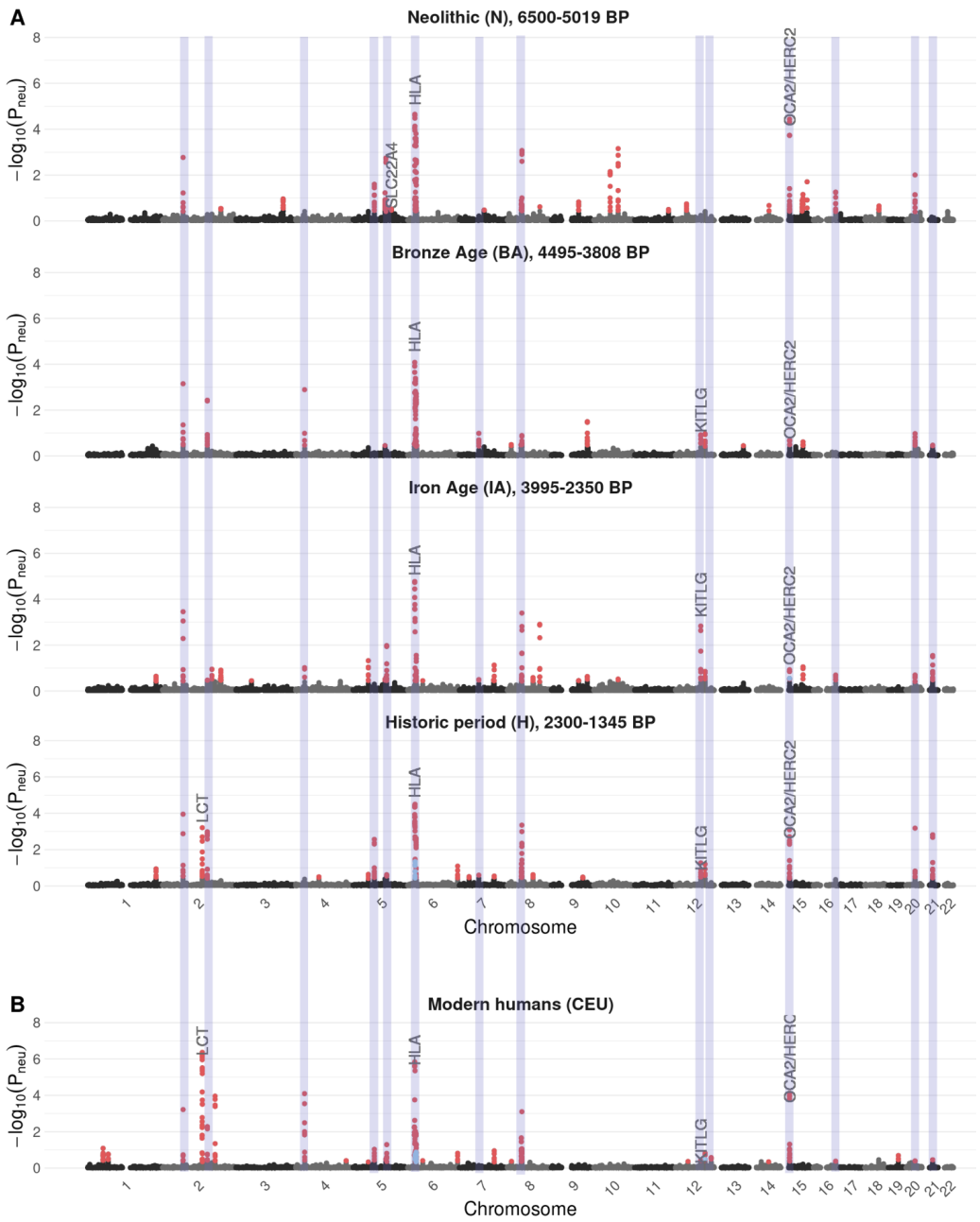

Figure 4. Genome-wide selection scan with the DANN in ancient and modern data. Results for the DANN are shown for all four ancient time periods highlighted at the top of each panel **(A)** and for modern Europeans **(B)**. The time period range is specified as years before present (BP). The y-axis shows the probability of selection -log(P_{neu}) predicted using the DANN, whereby a high value indicates the window is likely under selection. The x-axis shows the genomic position. The DANN was trained using a human admixture model with missing data as the source domain and aDNA data as target domain. Windows predicted as hard sweeps are colored in red and windows

352 predicted as soft are colored in blue. Sweeps previously reported in the literature are highlighted
353 above the peaks. We highlight with purple vertical bands 14 sweeps that are present across the two
354 ends of the major admixture event that occurred ~4.5 kya, that is sweeps that are detected in earlier
355 periods (N or BA) and are also detected in later periods (H or CEU).
356
357 *Hard sweeps were common in ancient and modern humans*
358 We next analyzed whether sweeps discovered in aDNA and modern humans were
359 predominantly hard or soft. The DANN classified all 58 sweeps (ancient and modern) in **Fig. 4** as
360 hard. In three cases, we observe that windows on the edges of the peak are classified as soft
361 (chromosomes 6 and 15, **Figs. 4 and S15)**. This pattern is consistent with the "soft shoulder
362 effect"[49], where recombination causes regions flanking a hard sweep to exhibit patterns that
363 resemble a soft sweep. However, based on our peak calling approach, these flanking regions are
364 not classified as independent soft sweeps. Instead, they are grouped with adjacent windows into a
365 single peak, which is then labeled according to the window with the highest sweep probability.
366 To evaluate whether the predicted hard sweeps may reflect a high rate of soft-to-hard sweep
367 misclassification by the model, we use simulated data from the human admixture model to estimate
368 the proportion of predicted hard sweeps that may actually correspond to soft sweeps misclassified
369 as hard. Applying these rates (**Fig. S16**) to our ancient hard sweep predictions, we estimate that,
370 conservatively, 83% of detected sweeps are hard and 15% are soft, indicating that despite any
371 potential for misclassification, hard sweeps were likely common in ancient human populations. It
372 is important to note, however, that these error rate estimates are rough approximations based on
373 simulated data, as the true labels in real aDNA are unknown.
374 Additionally, to assess robustness of our results to the underlying model of soft sweeps, we
375 trained a model in which soft sweeps were simulated from standing genetic variation (SGV) rather
376 than recurrent *de novo* mutations (**Fig. S10B**). We identified a total of 53 unique sweeps in aDNA,
377 43 of which overlap with the scan trained on *de novo* soft sweeps. Of the 53 sweeps, 38 are
378 classified as hard, 9 as soft and 6 change classification from hard to soft across time periods.
379 However, we note that the predicted probabilities for soft sweeps are low, ranging from 0.34-0.44,
380 where 0.33 is the probability of predicting either hard, soft or neutrality at random. By contrast,
381 the predicted probabilities for hard sweeps range from 0.34-0.93, suggesting that the support for
382 peaks classified as soft is weak.

To further validate whether the peaks detected by the DANN are in fact hard sweeps, we visualized the central window of all peaks detected in Figure 4. By visual inspection, we observe a single haplotype at high frequency across all windows, consistent with partial hard sweeps, as highlighted in 24 examples in **Fig. S17**. These windows are distinct from windows classified as soft on chromosome 6 of period H (**Fig. S18**), whereby multiple haplotypes are at high frequency, and distinct from windows classified as neutral, whereby no haplotypes are at high frequency (**Fig S18**).

In addition to the above, to further confirm inferences about the softness of sweeps made by the DANN, we compute the value of the haplotype homozygosity statistics H12 and H2/H1[37] for the sweep candidates detected in Figure 4A. These statistics have been previously used to detect and classify hard and soft selective sweeps, with both hard and soft sweeps showing elevated H12, and soft sweeps exhibiting higher H2/H1 values than hard sweeps[37]. Using an approximate Bayesian computation (ABC) approach to estimate whether each (H12, H2/H1) pair observed in the aDNA is more likely to arise under a hard or soft sweep model (**Fig. S19**), we find that of the 23 sweep candidates with elevated H12 values in the 95th percentile of values under neutrality, 22 peaks are better supported by a hard sweep model (BF<1) with 15 showing strong support (BF<0.5).

Finally, to assess if our DANN is able to recover known soft sweeps,we applied the DANN to a North American population of *Drosophila melanogaster* that has three well established soft sweeps that have been identified empirically at *Ace*[50–52], *CHkov1*[53,54] and *Cyp6g1*[55,56]. The sweeps at *Ace* and *Cyp6g1* arose from recurrent de novo mutations while the sweep from *CHkov1* arose from standing genetic variation. To run the scan on this data, we trained a new DANN using the *D. melanogaster* data as the target domain and simulated data from a constant N_e model with parameters relevant to this population for the source domain (Methods). Our model is able to recover all three known positive controls and classify them as soft sweeps (**Fig. S20**). Additionally we find that soft sweeps dominate across the autosomes of this population and that hard sweeps are enriched on the X chromosome relative to the autosomes, consistent with our previous findings that hemizygosity on the X results in an abundance of hard sweeps[57,58].

414 *Selective sweeps through time*

415 The massive admixture events in which large fractions of the European population
416 experienced genetic turnover may have important implications for the ability to detect selective
417 sweeps over time. Previous work has suggested that sweeps in ancient populations may no longer
418 be detectable in modern time periods due to a masking effect of admixture and drift[3,30]. Of
419 relevance to our dataset, 33% of the Anatolian population from which our samples were derived
420 was replaced by samples from the CHG population ~4.5kya, overlapping samples from the BA
421 and IA periods (**Fig. 1B, C**). We asked whether sweeps detected before this admixture event are
422 detectable subsequently, as this would imply resilience to major events and shared selective
423 pressures through time.

424 To investigate the potential impact of admixture events on sweep detection across time, we
425 quantified the overlap in selective sweep signals across the five time periods studied here. We
426 found that among the ancient time periods (N-H), 35 sweeps are detectable in only one or two time
427 periods, suggesting that admixture and drift may in fact have an impact on sweep detection.
428 However, at the same time, 14 sweeps that were identified before the admixture event in either the
429 N or BA time periods were also present after the admixture event in the H time period and modern
430 humans. This suggests that admixture events may not obscure all ancient sweeps and that some
431 sweeps are subject to sustained selection pressures over time (**Fig. S9**).

432 The genes identified in these 14 selective sweeps persisting across human epochs fall into
433 a few functional categories: These include neural and cognitive functions encoded by AUTS2,
434 ASCL1, and SEMA6A, of which AUTS2 was previously discovered to putatively be under
435 selection[59], neuronal signaling and calcium channels encoded by CACNB4, exocytosis encoded
436 by EXOC6B[60], and previously[4,38] discovered adaptations at pigmentation genes OCA2, HERC2,
437 and KITLG. Most of these genes are either found solo within the coordinates of their respective
438 selective sweeps, or with few other genes, narrowing the targets of selection. Contained in peaks
439 with more genes are metabolic and nutrient processing genes like PAH and SLC38A9,
440 reproductive and germ cell genes such as DDX4, SPAG4, and protein quality control and signaling
441 genes like LTN1, USP16, CCT8, and MAP3K7CL (**Table S4**). Together, the gene categories
442 present in the 14 sweeps persisting through history highlight functional classes, particularly
443 cognitive and pigmentation, that were potentially of great importance throughout the past 7000

444 years of history. Future work, however, is needed to fully understand the nature of positive
445 selection at these loci.

446 We next asked whether lack of overlap of sweep signals is correlated with the admixture
447 event 4.5kya. To do so, we measured the overlap in sweeps between pairs of ancestral time periods
448 (N-H) using the Jaccard similarity index (*J*), which measures the proportion of shared elements
449 between two sets relative to their union (Methods). We asked whether these observed Jaccard
450 values were larger or smaller than expected under a null where sweeps are randomly distributed
451 across time periods. We found that the farthest time periods, N and H, which have the admixture
452 event separating them, had significantly less sharing of sweeps than expected by chance (*J=0.17,*
453 p-val=0.03, permutation test, **Fig. S21,** Methods). This is consistent with some sweeps detected in
454 the earliest time period (N) being lost due to the impact of admixture and/or drift over long
455 timescales. However, the absence of sweeps in later periods may, in some cases, reflect limited
456 power rather than true biological loss. Additionally, we found that two sets of consecutive time
457 periods, (BA, IA) and (IA and H), have higher than expected sweep sharing (*J= 0.44-0.45,* p-val
458 <0.05, permutation test, **Fig. S21**, Methods), indicating that on shorter time scales sweeps can
459 persist.

460 Next, we asked whether sweeps that are shared across multiple time periods persist because
461 the same haplotype carrying the adaptive allele remained under selection, or whether the original
462 haplotype was replaced by a distinct haplotype, potentially introduced through admixture. To test
463 this, we calculated whether the most frequent haplotypes of peaks across time periods are more
464 similar than would be expected based on the average divergence between random pairs of
465 haplotypes (**Fig. S22, Fig. 5**). Out of the 14 sweeps detected on either end of the major admixture
466 event spanning our dataset (**Fig. 4**), 9 of these share the same top haplotype across at least 4 time
467 periods, including periods where our method does not detect the sweep. This implies that the
468 sweeping haplotype frequently persisted despite widespread admixture. In **Fig. 5A**, we highlight
469 6 examples of these sweeps. Among these we include a more recent sweep at the *LCT* locus where
470 the recent rise of the adaptive haplotype is particularly evident.

471 In addition to sweeps that persist across multiple time periods, we also tracked the
472 frequency of the top haplotype in cases where the sweep is only detected in the earliest period (**Fig.
473 5B**), clearly highlighting how, in some cases, a haplotype at high frequency in N is lost or masked
474 in subsequent periods. The temporal patterns highlighted in **Fig. 5A,B** are strikingly different from

those in regions classified as neutral by the DANN (**Fig. 5C**), where no haplotype reaches high frequency and the distances between top haplotypes fluctuate.

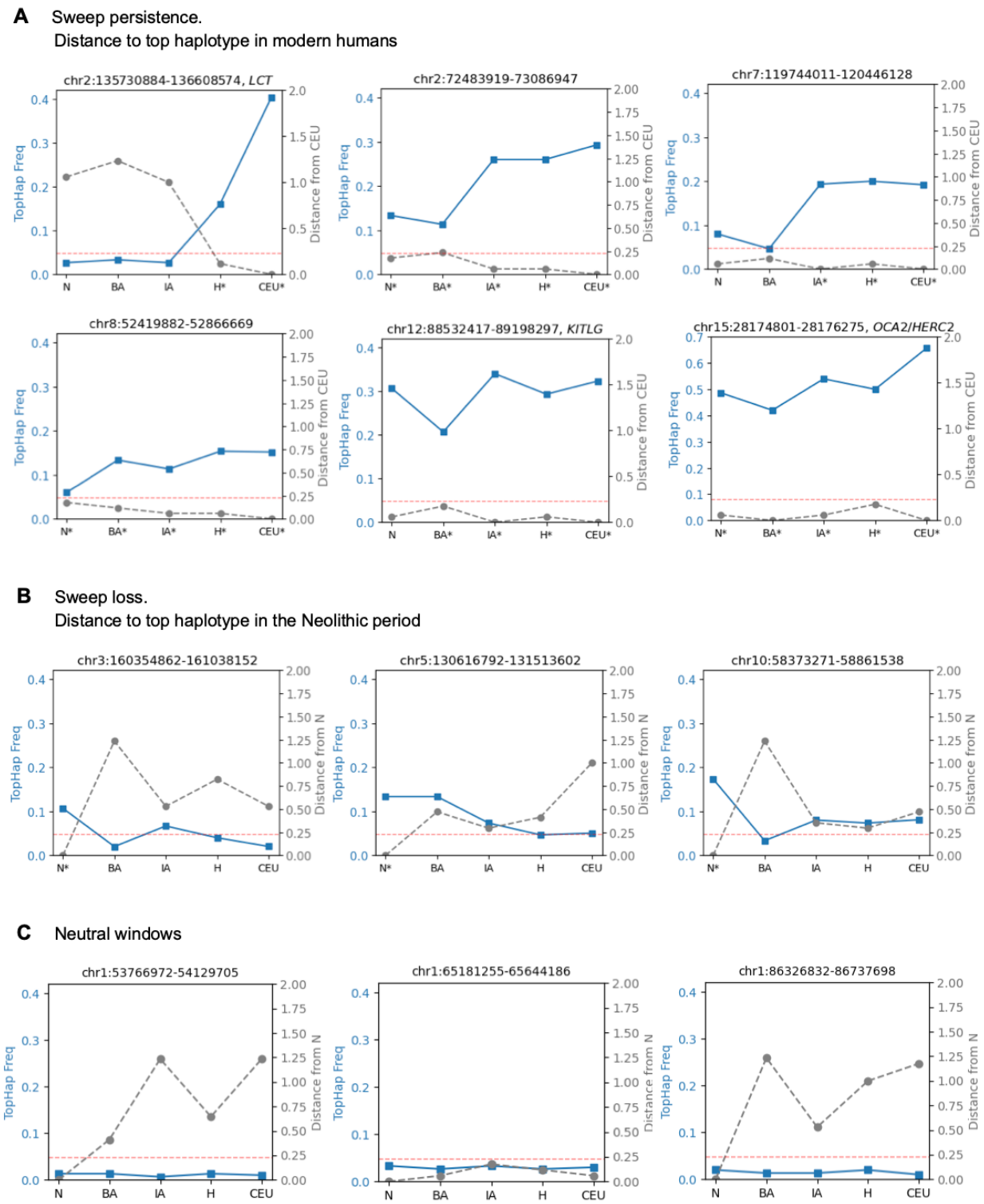

Figure 5. Persistence of top haplotype across time. (A) Six examples where the most frequent haplotype persists across multiple time periods. **(B)** Three examples of sweeps that become masked after time period N. **(C)** Three examples of neutral windows across all time periods. In all panels, the blue line (left y-axis) shows the frequency of the top haplotype in four ancient populations (N–H) and in a modern population (CEU). The gray line (right y-axis) shows the normalized Hamming distance between the top haplotype in each time period and the top haplotype in CEU in the case of **(A)** or in period N in the case of **(B)**, **(C)**. The horizontal red dashed line marks the 1st percentile of the distance between random haplotype pairs, with values above this threshold indicating distinct haplotypes. We highlight with a star (*) the epochs for which the window was classified as a sweep using the DANN.

Discussion

In this study, our goal was to characterize the targets and mode of adaptation in ancient humans. To do this, we implemented a domain adaptive neural network that is able to detect and classify selective sweeps in aDNA data and is robust to model misspecification. We applied our model to empirical aDNA from 708 individuals spanning ~7000 years in the past as well as 99 modern Europeans, and identified 48 unique ancient hard selective sweeps and 28 modern hard sweeps, recovering both previously known and novel candidates of selection. Finally, we found that while some sweeps identified in the Neolithic have been masked in more recent time periods, 14 sweeps spanning neuronal, reproductive, pigmentation, and signalling traits, can be found in the earliest and latest time periods despite the impacts of drift and admixture.

Past studies on humans have generally found few clear examples of hard sweeps in modern genomes[61,62]. Souilmi et al. proposed that this paucity of classic sweeps reflects the impact of complex population history, in particular the masking of sweep signatures by strong admixture events. Using SweepFinder2[63], which is primarily powered to detect completed hard sweeps, they identified 57 sweeps across ancient Eurasian human genomes, including many sweeps that are undetectable in modern data. Our DANN recovers several sweeps that show a similar pattern: Out of the 24 sweeps detected in the most ancient time period of our study, the Neolithic, 41% are not detected in the following periods (**Fig. S9**, **Fig. 5B**). However, our results also reveal additional dynamics not previously captured: the persistence of sweeps across time periods, even those most impacted by admixture. First, we observed sweeps in which the top haplotype is shared across

509 multiple time periods, even when the sweep is not detected by the DANN in all time intervals.
510 This pattern might reflect standing genetic variation, where the haplotype has not been driven by
511 selection to a high enough frequency to be detected by the DANN. We also detect cases where the
512 same top haplotype appears in non-consecutive periods, suggesting more complex evolutionary
513 dynamics such as fluctuating selection or a temporary replacement of the top haplotype driven by
514 admixture or migration that obscure detection of the sweep in some time intervals.

515 The finding that hard sweeps were common in aDNA suggests that adaptation in humans
516 has largely been mutation limited, or that standing genetic variation available to seed adaptation
517 was low. Even accounting for a 16% misclassification rate of soft sweeps as hard sweeps
518 (**Fig. S16**), hard sweeps remain the dominant mode of selection in aDNA. The high frequency of
519 hard sweeps in the data is consistent with the notion that hard sweeps are expected to dominate in
520 populations with small population sizes, where the input of new adaptive mutations is low[7,64]. In
521 ancient human populations, the effective population size is estimated to be relatively small[35] (
522 N_e~10^4), therefore the input of new adaptive mutations was likely moderate ($\Theta_A \sim 0.001$), making
523 adaptation through hard sweeps likely. Nonetheless, given the inherent challenges of
524 distinguishing between hard and soft sweeps, especially in regions of the parameter space where
525 their genomic signatures overlap, additional work will be needed to fully resolve the rapidity of
526 human adaptation.

527 Domain adaptation provides meaningful improvement on supervised machine learning
528 methods for analyzing genomic data. With the significant increase in available aDNA samples
529 over the past decade[65,66], domain adaptation may be especially valuable, as aDNA analyses may
530 be susceptible to false positives due to unmodeled or unknown demographic events as well as the
531 overall poor quality of the data. Domain adaptation was recently applied to SIA, a method that
532 detects selection using the ancestral recombination graph (ARG) inferred from sequence data[20,26].
533 Here, we bypass ARG inference, which is computationally intensive and can introduce an
534 additional layer of misspecification, by working directly with haplotype matrices. In other recent
535 work, site frequency spectra were used as inputs to the DANN, though this summary statistic
536 removes any linkage signal between SNPs[67]. By working with haplotypes[18], we can fully leverage
537 all available data without the need for additional inferences, making it more straightforward and
538 better suited hard vs soft sweep inference in aDNA. This framework is generalizable and in future

539 work could be extended to other organisms with low coverage data or where the demography is
540 not fully characterized.
541 Our findings highlight the power of deep learning for uncovering signatures of selection in
542 aDNA data. By applying a DANN to aDNA data, we find that hard sweeps have played an
543 important role in the evolutionary history of humans. Moreover, our approach opens the door to a
544 range of future applications of deep learning in aDNA to study adaptation, including incorporating
545 time as an explicit variable to examine how selection has fluctuated across historical periods or
546 extending our approach to detect other modes of selection. As aDNA datasets continue to grow
547 and deep learning methods for population genetics continue to improve, so will our ability to
548 disentangle the evolutionary forces that have shaped genetic diversity through human history.
549

550 # Methods

551

552 **Ancient DNA data**

553 Data was downloaded from the Allen Ancient DNA Resource (AADR version 51;
554 https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-
555 day-and-683). This includes genome-wide data from human populations from Holocene Europe
556 with samples dating from ~7,000 years before present (BP) to ~1,345 BP, covering the Neolithic
557 (N), the Bronze Age (BA), the Iron Age (IA) and the Historical periods (H) (**Fig. 1**).
558 We focused our analyses on the most reliable samples in our dataset. The criteria we used
559 to select samples are the same as in our previous work[30] and include enrichment for 1240k nuclear
560 targets with an in-solution hybridization capture reagent, removal of individuals with high
561 indication of contamination (see [30]) , and inclusion of unrelated individuals up to the third degree.
562 Additionally, like in our previous work we selected 177 individuals with the highest coverage
563 across all time periods, resulting in a total of 708 aDNA samples (Table S1). Finally, since aDNA
564 coverage is low and thus both alleles at heterozygous sites may not be sampled, and, since there is
565 ascertainment bias of alleles towards the SNPs included in the aDNA capture array[68] we pseudo-
566 haploidized the data by randomly selecting one of the reads that mapped to a given position and
567 assigned that read as the genotype of the sample at that site[30].
568
569

570 **Simulations**

571 We performed simulations of neutrality, hard sweeps and soft sweeps. We use two different
572 demographic models: a constant $N_e = 10^4$ model and a human admixture demographic model that
573 accounts for the major migratory movements contributing to the genetic diversity of contemporary
574 Europeans[3,33–35] (**Fig. 1C**).

575 We simulated a total of 400,000 neutral simulations and 400,000 selective sweep
576 simulations (200,000 hard sweeps and 200,000 soft sweeps) under each demographic model.
577 Constant mutation rates and recombination rates were applied to each simulated genome and were
578 drawn uniformly as follows: $\mu \sim U[1e\text{-}8, 1.5e\text{-}8]$ and $\rho \sim U\ [3e\text{-}9, 2e\text{-}8]$. Simulations were performed
579 using SLiM 4.1[69,70] followed by recapitation with msprime[71].

580 We simulated hard sweeps by introducing a single adaptive mutation to the center of the
581 chromosomal segment of length 450kb. We restarted the simulation and re-introduced the adaptive
582 mutation if the mutation was lost. The simulation was allowed to proceed until the adaptive
583 mutation reached a partial frequency (*PF*) drawn from a uniform distribution $PF \sim U[0.5, 0.95]$. In
584 all simulations, 177 individuals were sampled and subsequently downsampled to the 150
585 individuals with the least amount of missing data, reflecting the same procedure applied to aDNA.

586 We simulated soft sweeps from recurrent *de novo* mutations by introducing adaptive
587 mutations to the center of the chromosome at a rate determined by parameter $\Theta_A = 4N_e \mu_A$, where
588 N_e is the effective population size and μ_A the mutation rate of the adaptive mutation. The value of
589 Θ_A was drawn from a uniform distribution $\Theta_A \sim U[1,5]$ as soft sweeps are expected when $\Theta_A >=$
590 1[7,64]. Finally, the selection strength, s, for both hard and soft sweep simulations was drawn from a
591 uniform distribution $s \sim U[0.005, 0.1]$.

592 Additionally, we simulated sweeps arising from standing genetic variation (SGV) by
593 drawing the frequency of the adaptive mutation prior to the onset of selection from a uniform
594 distribution $f_{init} \sim U[0.025, 0.1]$. Given the computational constraints of simulating a sweep from
595 the SGV jointly with a complex admixture model, we only simulated SGV sweeps using the
596 constant N_e model. Additionally, to improve computational efficiency, we used a hybrid approach
597 where we first simulated a neutral process with msprime and selected a mutation m_i at frequency
598 $f_{init.}$ The resulting tree and mutation id of the variant m_i was then fed to SLiM, where m_i was
599 assigned a selective advantage and allowed to rise in frequency, producing a sweep from SGV.

Simulations were processed to mimic the missing data rate observed in aDNA data. Missing data was added to each site following a beta distribution with a mean of 0.43 per SNP and a standard deviation of 0.28, mimicking the distribution of missing data observed in aDNA (**Fig. S2**). Additionally, we pseudo-haploidized the data by randomly selecting one allele from each sampled individual at each site[30]. Finally, since we ran our scan in 201 SNP windows in aDNA, spanning roughly 450Kbs, we randomly selected 201 SNPs from the simulated chromosome.

DANN model architecture

The input to the DANN are raw genomic "images" or haplotype matrices (**Fig. 2B**), where rows represent pseudo-haplotypes from each individual and columns represent the ordinal position of each variant site in the sample. These haplotype matrices have dimensions n x L, where n=150 pseudo-haplotypes and L=201 SNPs in a given window. In these images, the color of each pixel represents the occurrence of the major or minor allele[16,18,36]. We transformed the alleles into binary values such that the major allele was coded as -1 and minor allele as 1. Missing data was coded as 0. Images were sorted by frequency of most to least frequent haplotype. In text S1 we test the effectiveness of alternative sorting approaches and window sizes.

The feature extractor of the DANN consists of two convolutional layers each containing 64 filters with kernel size of 3x3 and ReLu activation (**Fig. 2A**). Each convolutional layer is followed by 2x2 max-pooling. The last pooling layer is then flattened into a feature vector that is shared between the two subsequent branches of the network: the classifier and the discriminator. Each branch consists of two fully-connected dense layers of 128 neurons each. We use ReLu activation functions and set a dropout rate of 0.5 after each dense layer. The classifier outputs a 3-neuron softmax layer with the predicted probabilities for each of the three classes: hard sweep, soft sweep, or neutral. The discriminator outputs a single sigmoid output layer, which predicts whether the haplotype image comes from the source or target domain.

An important component of the DANN is a Gradient Reversal Layer (GRL) between the feature vector and the discriminator branch. During the feed-forwards step of training, the GRL is inactive and the data is passed along to the next layers. During backpropagation, the GRL inverts the gradient of the loss before passing it back to the feature extraction layer[27]. This operation penalizes features that discriminate between source and target domains, encouraging the model to learn domain-invariant features critical for accurate classification.

Model training

We implemented the domain-adaptive neural network model described above using TensorFlow (v2.18.0). All models were trained with the Adam optimizer and a batch size of 64. The classifier branch utilized a categorical cross-entropy loss function, while the discriminator branch used a binary cross-entropy loss function. We fed labeled data from simulations into the classifier branch to compute the class prediction loss ($L_{classifier}$). Simultaneously, we fed a mix of unlabeled data from the source domain (simulations) and target domain (aDNA data) into the discriminator branch to compute the discriminator loss ($L_{discriminator}$). During back propagation, the feature extractor's weights were updated based on a combination of the gradient from the classifier loss and the reversed gradient from the discriminator loss. We use the same simulated data for both branches, however the data is shuffled differently in each mini-batch. The real empirical aDNA data was used on the discriminator branch only. To achieve this training approach, we implement a custom data generator using the Sequence class ('tf.keras.utils.Sequence') that acts as a data generator interface for training Keras models.

The relative contribution of the model's branches can be adjusted via the hyperparameter λ, such that

$$L_{total} = L_{classifier} + \lambda L_{discriminator} .$$

If λ=0 then the GRL is effectively "off" and only the classifier branch learns to classify sweeps without attempting to unlearn domain differences. Conversely, when λ=1 the GRL is fully "on", assigning equal importance to the tasks of classifying sweeps and unlearning domain differences simultaneously. Following the approach proposed by Ganin & Lempitsky 2014[27], we gradually increased λ from 0 to 1 in order to reduce potential noisy signal from the discriminator at early stages of training. In Ganin & Lempitsky 2014 λ is defined as:

$$\lambda = \frac{2}{1+ exp(-\gamma \cdot p)} - 1,$$

Where p is the training progress changing linearly from 0 to 1 and is defined as p= $epoch_i$ / n_{epochs}, with n_{epochs} =30 total epochs and $i=1,..n_{epochs}$. In all models trained for this paper we set γ =10.

We trained the DANN for a total of 30 training epochs. Each epoch took ~282s to train on a single A100 GPU. During training, the discriminator loss gradually increased as the model "unlearned" the misspecification between the source and target. The loss eventually plateaued at ~0.693, consistent with the value expected when the domains become indistinguishable under

662 binary cross-entropy loss. Simultaneously, we observed that the classifier loss monotonically
663 decreased as it learned to distinguish between neutrality, hard sweeps, and soft sweeps (**Fig. S23**).
664 First, we trained and tested the DANN on simulated source and target domains. In this
665 scenario, we had labeled data from both domains, allowing us to evaluate the performance of the
666 DANN using a validation set from the target domain. In this setting, we selected the model weights
667 from the training epoch that achieved the lowest classifier validation loss when testing on the
668 validation dataset from the target domain.
669 Next, we trained and tested the DANN but this time using simulated data generated under
670 a human admixture model with an average of 43% missing data per site as the source domain and
671 aDNA data as the target domain. We included all aDNA data for training, where the data only
672 passed through the discriminator branch, not the classifier, since their labels are unknown. The
673 classifier only sees the aDNA data during inference (next section), avoiding any data leakage. In
674 this setting, we could not directly evaluate the performance of the model on a target validation set
675 due to the lack of labeled data. Instead, we assessed performance on simulated data by computing
676 the AUPRC for each class at every training epoch and then averaging across the three classes. We
677 restricted model selection to epochs beyond epoch four, when the GRL begins to influence training
678 ($\lambda > 0.5$), and chose the epoch with the highest average AUPRC.
679
680 **DANN genome-wide scan**
681 We performed a genome-wide scan across all autosomes and all four time periods of aDNA
682 data from this study. We also include a scan on 99 samples of modern humans from the CEU
683 population from the 1000 Genomes Project[32]. This modern data scan was generated using a
684 different DANN than the one used on ancient samples, trained on CEU data as the target domain.
685 To apply the model to all autosomes, we use a sliding window of 201 SNPs, advancing
686 each window by 10 SNPs. To refine the predictions and reduce noise, we then averaged the
687 predicted probabilities every five consecutive, overlapping windows, generating the final
688 predictions used for the scan. The DANN outputs a probability for each of the three classes
689 (neutral, hard sweep or soft sweep). Each window was assigned to the class with the highest
690 probability. To make sure we are identifying distinct selective events, we grouped consecutive
691 non-neutral windows into a single peak and required peaks to be at least 1.5 Mb apart. Each peak
692 was represented by the window with lowest probability of neutrality, in other words, the strongest

signal within the peak. The classification of this representative window (hard or soft sweep) was used to classify the whole peak.

ABC for hard/soft sweep classification with H12 and H2/H1 statistics

We assessed whether the predictions for hard vs soft sweeps in the data made by the DANN are consistent with predictions made with the haplotype homozygosity statistics, H12 and H2/H1, that can jointly discriminate between hard and soft sweeps[37]. H2/H1 is expected to be small for hard sweeps and large for soft sweeps, conditional on H12 being larger than expected under neutrality. As such, we only analyzed 23 peaks in total that had an H12 value greater than the 95th percentile of H12 values from 400,000 neutral simulations under the human admixture model as these were least likely to be neutral.

We used an approximate Bayesian computation approach to evaluate if a pair of (H12, H2/H1) values in the data are more likely under a hard vs soft sweep model. To do so, we compare observed values in the data to values measured from simulations of hard sweeps and soft sweeps under the human admixture model. We calculated Bayes factors (BFs) for the observed data by taking the ratio of the number of soft sweep and hard sweep simulations with a Euclidean distance <0.1 from each (H12, H2/H1) data point[37].

Jaccard index and permutation test

To quantify the overlap of peaks across different time periods, we calculated a Jaccard similarity index (*J*) quantifying peak overlap between pairs of time periods. *J* measures the proportion of shared elements between two sets relative to their total combined elements and its defined as

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|},$$

or in other words, the number of shared elements in sets A and B divided by the total number of elements in A and B. In this scenario A and B represent the peaks identified in two distinct time periods, such as IA and H. *J(IA,H)* is obtained by dividing the number of peaks that are shared between IA and H by the total number of peaks found across both periods.

Next, to identify whether the value of J was statistically significant, we performed a permutation test where we randomly shuffled the peaks and time periods, such that peaks were randomly distributed across time. We did this 5,000 times generating a null distribution of expected values of J. The p-value of the observed J value is given by the proportion of test statistics from the null distribution that are as extreme or more extreme than the observed value for a pair of time periods.

Code availability

All code developed and used in this study will be made publicly available via a GitHub repository upon acceptance of the manuscript.

Acknowledgements

The authors thank Vagheesh Narasimhan and Devansh Pandey for providing processed aDNA data and for helpful conversations. The authors also thank members of the Garud Lab, particularly Peter Laurin, for helpful conversations and feedback on the paper. This work was funded by an NIGMS NIH award (R35GM151023) to NRG, an NSF CAREER award (no. 2240098) to NRG, NIGMS NIH award (R35GM127070) to AS, a UCLA Dissertation Year Award to MH, and the CSHL School of Biological Sciences Gladys & Roland Harriman Fellowship ZM.

References

1. Page, A. E. *et al.* Reproductive trade-offs in extant hunter-gatherers suggest adaptive mechanism for the Neolithic expansion. *Proc. Natl. Acad. Sci.* **113**, 4694–4699 (2016).
2. Marciniak, S. *et al.* An integrative skeletal and paleogenomic analysis of stature variation suggests relatively reduced health for early European farmers. *Proc. Natl. Acad. Sci.* **119**, e2106743119 (2022).
3. Souilmi, Y. *et al.* Admixture has obscured signals of historical hard sweeps in humans. *Nat. Ecol. Evol.* **6**, 2003–2015 (2022).
4. Akbari, A. *et al.* Pervasive findings of directional selection realize the promise of ancient DNA to elucidate human adaptation. Preprint at https://doi.org/10.1101/2024.09.14.613021 (2024).

754 5. Kerner, G. *et al.* Genetic adaptation to pathogens and increased risk of
755 inflammatory disorders in post-Neolithic Europe. *Cell Genomics* **3**, 100248 (2023).
756 6. Hermisson, J. & Pennings, P. S. Soft Sweeps: molecular population genetics of
757 adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
758 7. Pennings, P. S. & Hermisson, J. Soft Sweeps II—Molecular Population Genetics
759 of Adaptation from Recurrent Mutation or Migration. *Mol. Biol. Evol.* **23**, 1076–1084 (2006).
760 8. Messer, P. W. & Petrov, D. A. Population genomics of rapid adaptation by soft
761 selective sweeps. *Trends Ecol. Evol.* **28**, 659–669 (2013).
762 9. Hermisson, J. & Pennings, P. S. Soft sweeps and beyond: understanding the
763 patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol. Evol.* **8**,
764 700–716 (2017).
765 10. Zheng, Y. & Wiehe, T. Adaptation in structured populations and fuzzy boundaries
766 between hard and soft sweeps. *PLOS Comput. Biol.* **15**, e1007426 (2019).
767 11. Harris, R. B., Sackman, A. & Jensen, J. D. On the unfounded enthusiasm for soft
768 selective sweeps II: Examining recent evidence from humans, flies, and viruses. *PLOS Genet.*
769 **14**, e1007859 (2018).
770 12. Garud, N. R., Messer, P. W. & Petrov, D. A. Detection of hard and soft selective
771 sweeps from Drosophila melanogaster population genomic data. *PLOS Genet.* **17**, e1009373
772 (2021).
773 13. Kelley, J. L., Madeoy, J., Calhoun, J. C., Swanson, W. & Akey, J. M. Genomic
774 signatures of positive selection in humans and the limits of outlier approaches. *Genome Res.* **16**,
775 980–989 (2006).
776 14. Sheehan, S. & Song, Y. S. Deep Learning for Population Genetic Inference. *PLOS*
777 *Comput. Biol.* **12**, e1004845 (2016).
778 15. Wang, Z. *et al.* Automatic inference of demographic parameters using generative
779 adversarial networks. *Mol. Ecol. Resour.* **21**, 2689–2705 (2021).
780 16. Flagel, L., Brandvain, Y. & Schrider, D. R. The Unreasonable Effectiveness of
781 Convolutional Neural Networks in Population Genetic Inference. *Mol. Biol. Evol.* **36**, 220–238
782 (2019).
783 17. Adrion, J. R., Galloway, J. G. & Kern, A. D. Predicting the Landscape of
784 Recombination Using Deep Learning. *Mol. Biol. Evol.* **37**, 1790–1808 (2020).

18. Torada, L. *et al.* ImaGene: a convolutional neural network to quantify natural selection from genomic data. *BMC Bioinformatics* **20**, 337 (2019).

19. Xue, A. T. *et al.* Discovery of Ongoing Selective Sweeps within *Anopheles* Mosquito Populations Using Deep Learning. *Mol. Biol. Evol.* **38**, 1168–1183 (2021).

20. Hejase, H. A., Mo, Z., Campagna, L. & Siepel, A. A Deep-Learning Approach for Inference of Selective Sweeps from the Ancestral Recombination Graph. *Mol. Biol. Evol.* **39**, msab332 (2022).

21. Amin, M. R., Hasan, M. & DeGiorgio, M. Digital Image Processing to Detect Adaptive Evolution. *Mol. Biol. Evol.* **41**, msae242 (2024).

22. Dabi, A. & Schrider, D. R. Population size rescaling significantly biases outcomes of forward-in-time population genetic simulations. *Genetics* **229**, iyae180 (2025).

23. Shimodaira, H. Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Stat. Plan. Inference* **90**, 227–244 (2000).

24. Burger, K. E., Pfaffelhuber, P. & Baumdicker, F. Neural networks for self-adjusting mutation rate estimation when the recombination rate is unknown. *PLOS Comput. Biol.* **18**, e1010407 (2022).

25. Riley, R., Mathieson, I. & Mathieson, S. Interpreting generative adversarial networks to infer natural selection from genetic data. *Genetics* **226**, iyae024 (2024).

26. Mo, Z. & Siepel, A. Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data. *PLOS Genet.* **19**, e1011032 (2023).

27. Ganin, Y. & Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. Preprint at https://doi.org/10.48550/arXiv.1409.7495 (2014).

28. Csurka, G. A Comprehensive Survey on Domain Adaptation for Visual Applications. in *Domain Adaptation in Computer Vision Applications* (ed. Csurka, G.) 1–35 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-58347-1_1.

29. Cochran, K. *et al.* Domain-adaptive neural networks improve cross-species prediction of transcription factor binding. *Genome Res.* **32**, 512–523 (2022).

30. Pandey, D., Harris, M., Garud, N. R. & Narasimhan, V. M. Leveraging ancient DNA to uncover signals of natural selection in Europe lost due to admixture or drift. *Nat. Commun.* **15**, 9772 (2024).

31. Skoglund, P. & Mathieson, I. Ancient Genomics of Modern Humans: The First Decade. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404 (2018).

32. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

33. Damgaard, P. D. B. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).

34. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).

35. Kamm, J., Terhorst, J., Durbin, R. & Song, Y. S. Efficiently inferring the demographic history of many populations with allele count data. *J. Am. Stat. Assoc.* **115**, (2019).

36. Zhao, H. & Alachiotis, N. Data preprocessing methods for selective sweep detection using convolutional neural networks. *Methods* **233**, 19–29 (2025).

37. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLOS Genet.* **11**, e1005004 (2015).

38. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).

39. Irving-Pease, E. K. *et al.* The selection landscape and genetic legacy of ancient Eurasians. *Nature* **625**, 312–320 (2024).

40. Le, M. K. *et al.* 1,000 ancient genomes uncover 10,000 years of natural selection in Europe. *BioRxiv* (2022) doi:10.1101/2022.08.24.505188.

41. de Bakker, P. I. W. & Raychaudhuri, S. Interrogating the major histocompatibility complex with high-throughput genomics. *Hum. Mol. Genet.* **21**, R29–R36 (2012).

42. Leffler, E. M. *et al.* Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science* **339**, 1578–1582 (2013).

43. Bersaglieri, T. *et al.* Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).

44. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).

45. Mathieson, I. & Terhorst, J. Direct detection of natural selection in Bronze Age Britain. *Genome Res.* **32**, 2057–2067 (2022).

846 46. Segurel, L. *et al.* Why and when was lactase persistence selected for? Insights from
847 Central Asian herders and ancient DNA. *PLOS Biol.* **18**, e3000742 (2020).
848 47. Evershed, R. P. *et al.* Dairying, diseases and the evolution of lactase persistence in
849 Europe. *Nature* **608**, 336–345 (2022).
850 48. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional
851 mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
852 49. Schrider, D. R., Mendes, F. K., Hahn, M. W. & Kern, A. D. Soft Shoulders Ahead:
853 Spurious Signatures of Soft and Partial Selective Sweeps Result from Linked Hard Sweeps.
854 *Genetics* **200**, 267–284 (2015).
855 50. Mutero, A., Pralavorio, M., Bride, J. M. & Fournier, D. Resistance-associated point
856 mutations in insecticide-insensitive acetylcholinesterase. *Proc. Natl. Acad. Sci.* **91**, 5922–5926
857 (1994).
858 51. Menozzi, P., Shi, M. A., Lougarre, A., Tang, Z. H. & Fournier, D. Mutations of
859 acetylcholinesterase which confer insecticide resistance in Drosophila melanogaster
860 populations. *BMC Evol. Biol.* **4**, 4 (2004).
861 52. Karasov, T., Messer, P. W. & Petrov, D. A. Evidence that Adaptation in Drosophila
862 Is Not Limited by Mutation at Single Sites. *PLOS Genet.* **6**, e1000924 (2010).
863 53. Aminetzach, Y. T., Macpherson, J. M. & Petrov, D. A. Pesticide Resistance via
864 Transposition-Mediated Adaptive Gene Truncation in Drosophila. *Science* **309**, 764–767
865 (2005).
866 54. Magwire, M. M., Bayer, F., Webster, C. L., Cao, C. & Jiggins, F. M. Successive
867 Increases in the Resistance of Drosophila to Viral Infection through a Transposon Insertion
868 Followed by a Duplication. *PLOS Genet.* **7**, e1002337 (2011).
869 55. Daborn, P., Boundy, S., Yen, J., Pittendrigh, B. & ffrench-Constant, R. DDT
870 resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance
871 to the neonicotinoid imidacloprid. *Mol. Genet. Genomics* **266**, 556–563 (2001).
872 56. Schmidt, J. M. *et al.* Copy Number Variation and Transposable Elements Feature
873 in Recent, Ongoing Adaptation at the Cyp6g1 Locus. *PLOS Genet.* **6**, e1000998 (2010).
874 57. Harris, M. & Garud, N. R. Enrichment of Hard Sweeps on the X Chromosome in
875 Drosophila melanogaster. *Mol. Biol. Evol.* **40**, msac268 (2023).

876 58. Harris, M., Kim, B. Y. & Garud, N. Enrichment of hard sweeps on the X
877 chromosome compared to autosomes in six *Drosophila* species. *GENETICS* **226**, iyae019
878 (2024).
879 59. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328**, 710–
880 722 (2010).
881 60. Evers, C. *et al.* Mosaic deletion of EXOC6B: further evidence for an important role
882 of the exocyst complex in the pathogenesis of intellectual disability. *Am. J. Med. Genet.* **164A**,
883 3088–94 (2014).
884 61. Hernandez, R. D. *et al.* Classic Selective Sweeps Were Rare in Recent Human
885 Evolution. *Science* **331**, 920–924 (2011).
886 62. Schrider, D. R. & Kern, A. D. Soft Sweeps Are the Dominant Mode of Adaptation
887 in the Human Genome. *Mol. Biol. Evol.* **34**, 1863–1877 (2017).
888 63. DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I. & Nielsen, R.
889 SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897
890 (2016).
891 64. Pennings, P. S. & Hermisson, J. Soft Sweeps III: The Signature of Positive
892 Selection from Recurrent Mutation. *PLOS Genet.* **2**, e186 (2006).
893 65. Mallick, S. & Reich, D. The Allen Ancient DNA Resource (AADR): A curated
894 compendium of ancient human genomes. Harvard Dataverse
895 https://doi.org/10.7910/DVN/FFIDCW (2023).
896 66. Mallick, S. *et al.* The Allen Ancient DNA Resource (AADR) a curated
897 compendium of ancient human genomes. *Sci. Data* **11**, 182 (2024).
898 67. Cobb, K. A. & Smith, M. L. The reasonable effectiveness of domain adaptation for
899 inference of introgression. Preprint at https://doi.org/10.1101/2025.01.17.633659 (2025).
900 68. Günther, T. & Nettelblad, C. The presence and impact of reference bias on
901 population genomic studies of prehistoric human populations. *PLOS Genet.* **15**, e1008302
902 (2019).
903 69. Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W. & Ralph, P. L. Tree-
904 sequence recording in SLiM opens new horizons for forward-time simulation of whole
905 genomes. *Mol. Ecol. Resour.* **19**, 552–566 (2019).

70. Haller, B. C. & Messer, P. W. SLiM 4: Multispecies Eco-Evolutionary Modeling. *Am. Nat.* **201**, E127–E139 (2023).

71. Baumdicker, F. *et al.* Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* **220**, iyab229 (2022).