

1 Fast and accurate estimation of selection coefficients and allele
2 histories from ancient and modern DNA

3 Andrew H. Vaughn^{1,*} and Rasmus Nielsen^{2,3}

4 ¹*Center for Computational Biology, University of California, Berkeley, Berkeley, CA
5 94720, USA*

6 ²*Departments of Integrative Biology and Statistics, University of California, Berkeley,
7 Berkeley, CA 94720, USA*

8 ³*Center for GeoGenetics, University of Copenhagen, DK-1350 Copenhagen, Denmark*

9 *Corresponding author: ahv36@berkeley.edu

10 December 16, 2023

11

Abstract

12

We here present CLUES2, a full-likelihood method to infer natural selection from sequence data that is an extension of the method CLUES. We make several substantial improvements to the CLUES method that greatly increases both its applicability and its speed. We add the ability to use ARGs on ancient data as emissions to the underlying HMM, which enables CLUES2 to use both temporal and linkage information to make estimates of selection coefficients. We also fully implement the ability to estimate distinct selection coefficients in different epochs, which allows for the analysis of changes in selective pressures through time. In addition, we greatly increase the computational efficiency of CLUES2 over CLUES using several approximations to the forward-backward algorithms and develop a new way to reconstruct historic allele frequencies by integrating over the uncertainty in the estimation of the selection coefficients. We illustrate the accuracy of CLUES2 through extensive simulations and validate the importance sampling framework for integrating over the uncertainty in the inference of gene trees. We also show that CLUES2 is well-calibrated by showing that under the null hypothesis, the distribution of log-likelihood ratios follows a chi-squared distribution with the appropriate degrees of freedom. We run CLUES2 on a set of recently published ancient human data from Western Eurasia and test for evidence of changing selection coefficients through time. We find significant evidence of changing selective pressures in several genes correlated with the introduction of agriculture to Europe and the ensuing dietary and demographic shifts of that time. In particular, our analysis supports previous hypotheses of strong selection on lactase persistence during periods of ancient famines and attenuated selection in more modern periods.

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

Introduction

32

33

34

35

36

37

38

39

40

41

42

43

44

45

One of the primary evolutionary forces that shapes the genetic variation of populations is natural selection, the causal effect of genotype on the reproductive success of an individual. While experimental evolution studies can enable a somewhat direct measurement of natural selection in certain limited cases, most methods for inferring natural selection rely on statistical frameworks to analyze a sample of sequence data. Given that sequence data lies in a very high-dimensional space, these methods for inferring natural selection have historically been based on summary statistics, that can be thought of as projections of the sequence data to a lower-dimensional subspace. These summary statistics include SNP-frequency statistics such as Tajima's D (Tajima, 1989) or Fay and Wu's H (Fay and Wu, 2000), as well as linkage disequilibrium statistics such as the extended haplotype homozygosity statistic (Sabeti et al., 2002) or the LD decay test (Wang et al., 2006). One can compute these statistics for a given sequence, or in sliding windows across the genome, and conclude that values significantly different from their null expectation are indicative of natural selection. As an extension of the basic summary statistic approach, approximate Bayesian computation (ABC) frameworks, which can use information from many different summary statistics, have also been used to infer selection coefficients (Peter et al., 2012). A variety of machine

46 learning methods to detect selection have additionally been developed (Schrider and Kern, 2018; Torada
47 et al., 2019; Hejase et al., 2021), which all rely on training models on a set of features obtained from
48 sequence data. However, these approaches ignore a large amount of the information present in the data
49 by only considering sets of summary statistics or features. Furthermore, these methods lack flexibility, for
50 example if one wishes to specify different parameters or classes of models, such as allowing for selection
51 coefficients to differ in different time periods or incorporating ancient samples that are sampled at distinct
52 time points.

53 Overall, therefore, a computationally tractable full-likelihood method is needed that does not rely on
54 summary statistics, and can thus use all the information present in the data, and which also has the
55 flexibility to be able to be applied to a variety of datasets and models. It is to this end that the method
56 CLUES was developed by (Stern et al., 2019). The original CLUES method is based on the observation
57 that if one can determine a full likelihood expression for the data conditional on the full history of the
58 frequency of an allele, then one can consider the allele history as a latent variable and integrate over the
59 full set of possible allele frequencies in order to obtain an expression for the full likelihood function. As
60 the sequence data and selection coefficient of an allele are approximately independent conditional on the
61 gene tree at that locus, the full likelihood of the data could be computed by computing the likelihood
62 of that gene tree. In reality, the true gene tree at a site is never known with certainty but instead
63 must be estimated. Estimating gene trees in recombining species is challenging as recombination causes
64 neighboring regions of the genome to have distinct, but correlated, gene trees. The full collection of these
65 correlated gene trees across a genomic region is called the *ancestral recombination graph* (ARG), the theory
66 for which was first developed by (Hudson, 1983) and (Griffiths and Marjoram, 1996). The key insight of
67 using ARGs for population genetic inference is that they contain all the information about the history of
68 each region of the genome. Therefore methods that employ ARGs are utilizing the maximum possible
69 information present in the data. The inference of ancestral recombination graphs is made difficult by the
70 large state space of possible graph topologies and the relatively small amount of information mutations
71 provide about the underlying graph. However, recent methods to infer ARGs have made significant
72 progress on this difficult computational task (Rasmussen et al., 2014; Kelleher et al., 2019; Speidel et al.,
73 2019; Mahmoudi et al., 2022; Zhang et al., 2023). Nevertheless, there will still always be uncertainty as
74 to the underlying ARG, as there will be many ARGs that are compatible with the observed sequence
75 data. For that reason, CLUES also integrates over the uncertainty in the ARG estimation, in addition
76 to integrating over the set of possible allele histories, thus giving two sets of latent variables that are
77 integrated out to obtain the full likelihood function.

78 We here present CLUES2, an extension of the CLUES method, which is able to work not only on
79 importance samples of ARGs on modern data but also on ancient genotype samples and ARGs built
80 on ancient data, thus enabling the usage of time series data and linked SNPs to give better estimates

81 of the selection coefficient of an allele. We also develop and test the capability of CLUES2 to estimate
82 different selection coefficients in different time periods. In addition, we make significant improvements to
83 runtime by making several well-justified approximations to the forward and backward algorithms. The
84 computational speedups allow more importance samples to be used, which increases accuracy and also
85 allows CLUES2 to be used in genome-wide scans of selection, where hundreds of thousands or millions of
86 SNPs may need to be analyzed. We perform rigorous testing of both selection coefficient estimation and
87 the appropriateness of the chi-squared test for hypothesis testing of selection. Finally, we improve the
88 interpretability of our reconstruction of historic allele frequencies by integrating over the uncertainty in
89 the estimation of the selection coefficients. We make CLUES2 available on GitHub as a well-documented
90 Python package at <https://github.com/avaughn271/CLUES2>.

91 Materials and Methods

92 CLUES2 Framework

We here describe the basic hidden Markov model framework of CLUES2. We wish to find the value of s ,
the selection coefficient of the derived allele, that maximizes the likelihood of the data D . What exactly
 D represents depends on the kind of data being analyzed, and we postpone a formal definition of D until
the subsequent sections. We compute this likelihood by conditioning on the historic trajectory of the
derived allele frequency. In particular, if we let \mathcal{X} be the set of all possible derived allele trajectories, we
compute the likelihood as:

$$P(D|s) = \sum_{X \in \mathcal{X}} P(D, X|s) = \sum_{X \in \mathcal{X}} P(D|X, s)P(X|s) = \sum_{X \in \mathcal{X}} P(D|X)P(X|s)$$

93 where the last equality follows from the fact that the data is independent of the selection coefficient given
94 the allele trajectory. If we discretize the derived allele frequency into K frequency bins and consider
95 the history of an allele until a maximum time point T (measured in discrete generations), computing
96 this expression would naively require summing over K^T many allele trajectories. However, we instead
97 recognize this as a hidden Markov model (HMM) where the derived allele frequency at a given time is
98 the hidden state and the data are the emissions. We can then efficiently compute this sum using the
99 forward algorithm. This modeling of allele frequencies as hidden states of an HMM has previously been
100 applied by (Williamson and Slatkin, 1999), (Bollback et al., 2008), (Steinrücken et al., 2014), (Bergman
101 et al., 2018), and (Paris et al., 2019).

102 The implementation of the forward algorithm is as follows: we compute a forward algorithm matrix
103 F , which has dimension $K \times T$. The first column is initialized to all zeros, except that a 1 is placed in the
104 row corresponding to the allele frequency bin closest to the observed modern allele frequency. Assuming

105 the first $t - 1$ columns of F have been computed, one time-step of the forward algorithm consists of
 106 computing, for each frequency bin k ,

$$F_{k,t} = e(t, k) \sum_{l=1}^K F_{l,t-1} P_{l,k}$$

107 where $P_{l,k}$ is the probability of transitioning from allele frequency bin l to bin k in one generation and
 108 $e(t, k)$ is the probability of observing the data at time t given you are in state k . Given K allele frequency
 109 bins, we consider the K numbers that are equally spaced between $1/200000$ and $1 - 1/200000$ inclusive,
 110 call them w_1, \dots, w_K . We set the allele frequency of bin k , call it x_k , to be the quantile function of a
 111 Beta($1/2, 1/2$) distribution at w_k . This creates a spacing of numbers between 0 and 1 that is denser near
 112 the boundaries. If $1 < l < K$, then $P_{k,l}$ is approximated as $\Phi_{s,N}((x_l + x_{l+1})/2) - \Phi_{s,N}((x_{l-1} + x_l)/2)$
 113 where $\Phi_{s,N}(x)$ is the CDF of a normal distribution with mean $sx(1-x)$ and variance $\frac{x(1-x)}{2N}$. $P_{k,1}$
 114 is $\Phi_{s,N}((x_1 + x_2)/2)$ and $P_{k,K} = 1 - \Phi_{s,N}((x_{K-1} + x_K)/2)$. We note that while this definition restricts
 115 population size N to be constant through time, we do indeed allow population sizes to vary through
 116 time (see section “Differing population sizes” for details). Furthermore, it is worth highlighting that
 117 throughout this manuscript, we use N to refer to haploid effective population size, denoting the number
 118 of sequences present in a population, not the number of diploid individuals in that population. After
 119 computing each entry of the matrix F , we have uniform exit probabilities from the chain, meaning
 120 that $P(D|s) = \sum_{k=1}^K F_{k,T}$. We then optimize $P(D|s)$ with respect to s to achieve our estimate \hat{s}^{MLE} .
 121 Dividing by $P(D|s=0)$ yields the log-likelihood ratio, which can then be used for hypothesis testing (see
 122 the section “Validation of chi-squared test”). In the following sections we discuss the types of data we
 123 use for the emissions $e(t, k)$.

124 Ancient Genotypes

One type of emissions considered are ancient genotypes. This is a list of genotypes of ancient individuals along with the times at which these individuals were sampled. If we denote a homozygous derived genotype as DD , a homozygous ancestral genotype as AA and a heterozygous genotype as AD we have

$$e(t, k) = \prod_{d \in D_t} (x_k^2 \mathbf{1}_{\{d=DD\}} + 2x_k(1-x_k)\mathbf{1}_{\{d=AD\}} + (1-x_k)^2 \mathbf{1}_{\{d=AA\}})$$

125 where D_t represents the set of all genotypes sampled at time t .

In practice, we allow the usage of genotype likelihoods as can be generated by the imputation of ancient genomes, so our expression becomes

$$e(t, k) = \prod_{d \in D_t} (x_k^2 P(d=DD) + 2x_k(1-x_k)P(d=AD) + (1-x_k)^2 P(d=AA))$$

126 We validate this approach on simulated data (see Figure 1a). If our implementation is correct, there
 127 should be exactly two sources of error, which both contribute to variance of our estimator:

128 1) Fluctuations in the true allele frequency trajectory due to genetic drift, which causes the true allele
 129 frequency to be less informative of the true value of s

130 2) Noisy estimation of the true allele frequency due to finite sampling of ancient genotypes.

131 Therefore, we ran simulations with very low drift and very dense sampling of ancient genotypes
 132 (see Figure 1b). We observed that the estimated values of \hat{s}^{MLE} approximately converged to the true
 133 values, which validates the correctness of our implementation. We also allow for the usage of ancient
 134 haplotype probabilities, in addition to genotype probabilities, by substituting the expression $x_k P(d =$
 135 $D) + (1 - x_k)P(d = A)$ in the product above.

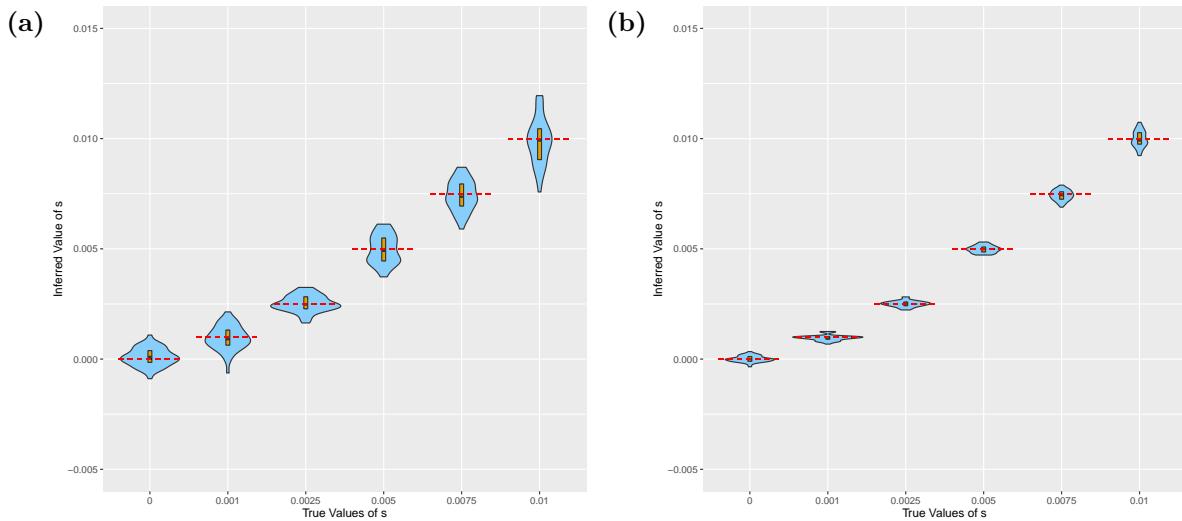


Fig 1: Violin plots showing the results of running CLUES2 on ancient genotype data. Boxplots are overlaid with the whiskers omitted. True values of s are shown as dashed red lines. 30 replicates were performed for each true value of s . Simulations were run with (a) $N = 30000$ and an individual sampled every 4 generations and with (b) $N = 600000$ and an individual sampled every 1 generation.

136 True Trees

137 The next class of emissions we consider are coalescence events of a gene tree. As in the original CLUES
 138 paper, we consider a structured coalescent model of the gene tree of a locus under selection following
 139 (Kaplan et al., 1988) and (Braverman et al., 1995). In particular, we assume we know the allele labeling
 140 at each leaf node of the tree and that the gene tree satisfies the infinite sites assumption with respect
 141 to the allele (we discuss violations to the infinite sites assumption later in the “Inferring Gene Trees on
 142 Ancient Human Data” section). We can then obtain a labeling of each branch in the tree as a derived
 143 allele branch or an ancestral branch and a labeling of each coalescence nodes in the tree as a derived allele
 144 coalescence or an ancestral branch (See Figure 2). The one exception to this is the branch on which the

145 mutation must have arisen, which we call the mixed lineage.

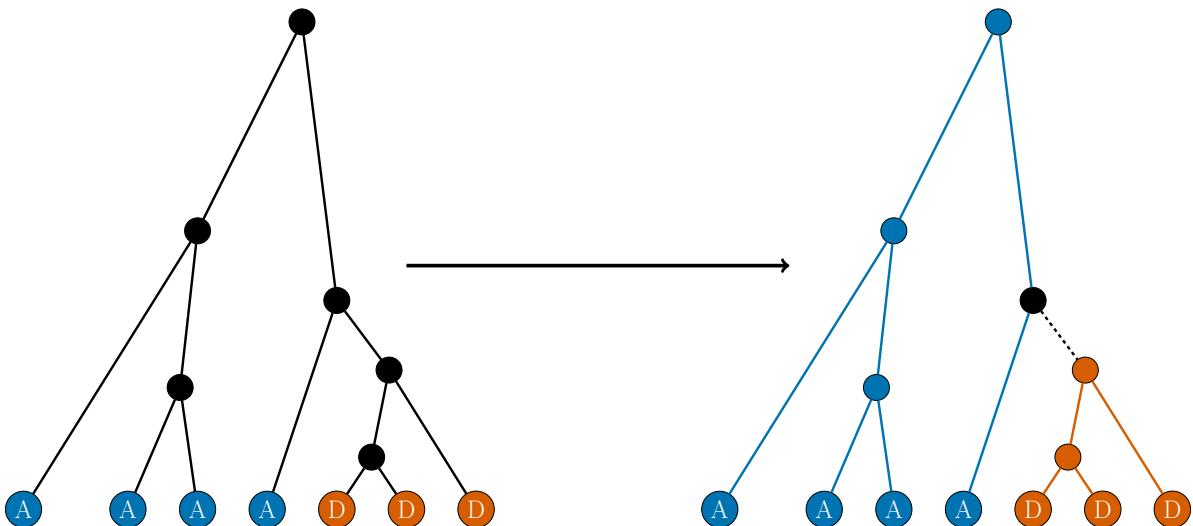


Fig 2: An outline of the labeling of branches and nodes in a tree as either ancestral (blue) or derived (orange) given a labeling of the leaf nodes. An internal node is a derived coalescence if and only if all its descendant leaves are derived. The parent node of the oldest derived coalescence is the mixed coalescence node (represented in black). All other coalescence events are ancestral coalescences. A branch represents a derived lineage if and only if it has a derived node as an ancestor. The mixed lineage is the immediate parent branch of the oldest derived coalescence (black dashed line). All other branches represent ancestral lineages.

146 Given our labeling, we begin at the present and move back in time, keeping track of the number of
 147 derived lineages n_D and ancestral lineages n_A that are present at the given time point. At time points
 148 younger than the mixed coalescence node, lineages can only coalesce with other lineages of the same
 149 allelic class. The instantaneous coalescence rate within the derived class is $X_t N$, and the instantaneous
 150 coalescence rate within the ancestral class is $s(1 - X_t)N$, where X_t is the frequency of the derived allele
 151 at time t . At time points older than the age of the mixed lineage, after which there can be no derived
 152 lineages, we only consider coalescence within the ancestral class. A similar structured coalescent approach
 153 is used by the simulation softwares **discoal** (Kern and Schrider, 2016) and **msprime** 1.0. (Baumdicker
 154 et al., 2021)

To formalize the emissions this model emits, given the model enters time t with n_D derived lineages and d derived coalescences are observed in the interval $(t, t+1)$ at times t_1 to t_d , the derived coalescence emission for frequency bin k is computed as:

$$e_D(t, k) = \left(1 - F\left(\frac{\binom{n_D-d}{2}}{x_k N}, t+1-t_d\right)\right) \prod_{i=1}^d f\left(\frac{\binom{n_D-i+1}{2}}{x_k N}, t_i - t_{i-1}\right)$$

155 where $f(\lambda, x)$ is the density function of an $\text{Exp}(\lambda)$ random variable at the point x and $F(\lambda, x)$ is the
 156 corresponding CDF. This is the probability of observing the given coalescences (possibly none) at the
 157 observed times and then not observing another coalescence before the end of the interval $(t, t+1)$. We
 158 compute the ancestral coalescence emissions $e_A(t, k)$ in the same way, replacing $x_k N$ with $(1 - x_k)N$
 159 and considering the number of ancestral lineages n_A and the number of ancestral coalescences a in the

160 interval $(t, t + 1)$, meaning in total our emission for this timestep is $e(t, k) = e_D(t, k)e_A(t, k)$.

161 The exceptions to the above expression are the following. If $n_A = 1$ and $n_D = 0$, $e(t, k) = 1$ if $k = 1$
 162 and $e(t, k) = 0$ otherwise. If $n_A > 1$ and $n_D = 0$, then $e_A(t, k)$ is calculated as usual if $k = 1$ and
 163 $e_A(t, k) = 0$ otherwise. If $n_A = n_D = 1$, then an ancestral emission with $n_A = 2$ is emitted if $k = 1$
 164 (signalling that the derived allele is no longer segregating) and $e(t, k) = 1$ otherwise. Similarly, if $n_A > 1$
 165 and $n_D = 1$, then an ancestral emission is emitted as usual if $k \neq 1$ and an ancestral emission with
 166 $n_A + 1$ is emitted if $k = 1$. We find that this framework models the mutation origin and the mixed
 167 lineage properly (Figure 3a). We note that the exact handling of the mixed lineage and the formula for
 168 the emissions produced by coalescences differs from that described in the original CLUES method.

169 There are again two sources of statistical noise in this estimation, which both contribute to the
 170 variance

- 171 1) Fluctuations in the true allele frequency trajectory due to genetic drift, which causes the true allele
 172 frequency to be less informative of the true value of s
- 173 2) Noisy estimation of the true allele history due to the finite number of lineages sampled.

174 Therefore, we ran simulations with very low drift and a very high number of sampled leaves (Figure
 175 3b). We observed that the estimated values of \hat{s}^{MLE} converged to the true values, which indicates the
 176 correctness of our implementation and that the model assumptions we make (such as discretizing time
 177 and allele frequency) are sufficiently accurate to not cause substantial biases.

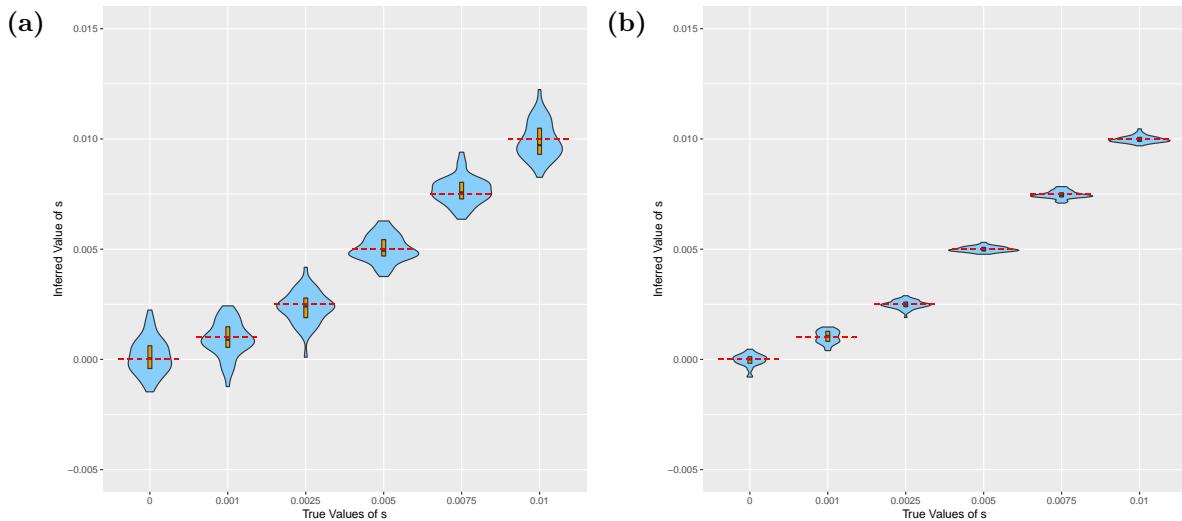


Fig 3: Violin plots showing the results of running CLUES2 on true trees. Boxplots are overlaid with the whiskers omitted. True values of s are shown as dashed red lines. 30 replicates were performed for each true value of s . Simulations were run with (a) $N = 30000$ and 80 sampled leaves and (b) $N = 600000$ and 800 samples leaves.

178 **Importance Sampling of Trees**

Of course, the true topology is never known with 100% certainty for real sequence data. Rather, the observed data consists of a set of SNPs around a locus of interest, meaning that the ancestral recombination graph (ARG) must be estimated and one must integrate over samples of the tree at the locus of interest in order to obtain an estimate of the likelihood function. In particular, (Stern et al., 2019) showed that under the assumption that the ARG G is independent of s given the marginal tree G_k at the SNP of interest, the likelihood ratio $\frac{P(D|s)}{P(D|s=0)}$ is equal to

$$E_{G|D,s=0} \left[\frac{P(G_k|s)}{P(G_k|s=0)} \right]$$

179 By sampling M graphs G^1, \dots, G^M from the distribution of ancestral recombination graphs given D and
180 given $s = 0$ and extracting the marginal tree at our SNP of interest from each graph, one can obtain the
181 following Monte Carlo estimator of the likelihood ratio

$$\frac{P(D|s)}{P(D|s=0)} \approx \frac{1}{M} \sum_{m=1}^M \left[\frac{P(G_k^m|s)}{P(G_k^m|s=0)} \right] \quad (1)$$

182 which can be computed by sampling many trees and, for each sampled tree, computing the probability
183 of that tree using the expression derived in the previous section. This reweighting of variates sampled
184 from a different distribution is a technique known as *importance sampling*. It is important to note that
185 this approach adds two additional sources of error, in addition to those described previously.

186 (1) Lack of convergence of the Monte Carlo estimator due to finite M .

187 (2) Error incurred by improper sampling of ARGs from the specified posterior distribution. This is to
188 say that the sampled graphs G^1, \dots, G^M are not *iid* samples from the distribution $P(G|D, s = 0)$. This
189 can happen due to practical considerations, such as poor mixing of the MCMC method used to generate
190 these samples, or poor calibration of the ARG-inference methods themselves.

191 In particular, with regards to the second possible source of error, we highlight the work of (Brandt
192 et al., 2022) in showing the degree of miscalibration of different ARG-inference methods and how this
193 depends on parameters such as recombination and mutation rate. Therefore, we develop our own MCMC
194 algorithm for sampling gene trees on sequence data in the absence of recombination (see the “Importance
195 Sampling” section of the “Simulation Details” Appendix for more information). This enables us to
196 validate the correctness of our importance sampling implementation without being affected by possible
197 biases induced by improper sampling of ARG-inference methods. We demonstrate the behavior of our
198 importance sampling estimator through a set of illustrative simulations. We begin by simulating genetic
199 data in `msprime` on a set of 24 haplotypes for a 1Mb region with no recombination and with a mutation
200 rate of 7×10^{-7} . We then run our purpose-built MCMC sampler to generate samples of gene trees given

the observed genetic data. We take 1 sample of a tree ($M=1$) and use it as input to CLUES2. This corresponds to not using the importance sampling framework at all but instead regarding this 1 sampled tree as the true gene tree at this locus with no uncertainty. We find that the mutation rate is so high that it overwhelms the prior centered on $s = 0$ and this 1 sample of a tree does indeed behave like the true gene tree for this data in terms of accuracy (Figure 4a). Then, we run simulations under identical parameter settings (including using $M = 1$ with no importance sampling) except that a mutation rate of $\mu = 2 \times 10^{-9}$ is used. In this case, the estimation of s is biased as there is not enough data to overwhelm the prior centered at $s = 0$. This effect is particularly pronounced for larger true values of s , which are the values that differ the most from the prior (Figure 4b). This bias can be considered an extreme case of the estimation error that can happen due to small M , and it is important to note that this results in bias, rather than increased variance. We note that this bias towards $s = 0$ for small values of M , in addition to poor calibration of existing ARG-inference methods, could explain previous analyses that showed a bias of CLUES towards $s = 0$ (see Fig. 4 of (Hejase et al., 2021), Fig. 2 of (Temple et al., 2023)). We also note that using a value of K , the number of allele frequency bins, that is too small can result in bias towards $s = 0$.

To show the utility of our importance sampling approach, we run the same simulation as before with $\mu = 2 \times 10^{-9}$ but instead take $M = 3000$ samples to be used in CLUES2. We find that by properly reweighting the samples of gene trees through our importance sampling framework, we recover approximately unbiased estimates of the selection coefficients (Figure 4c), which validates the implementation of our importance sampling approach.

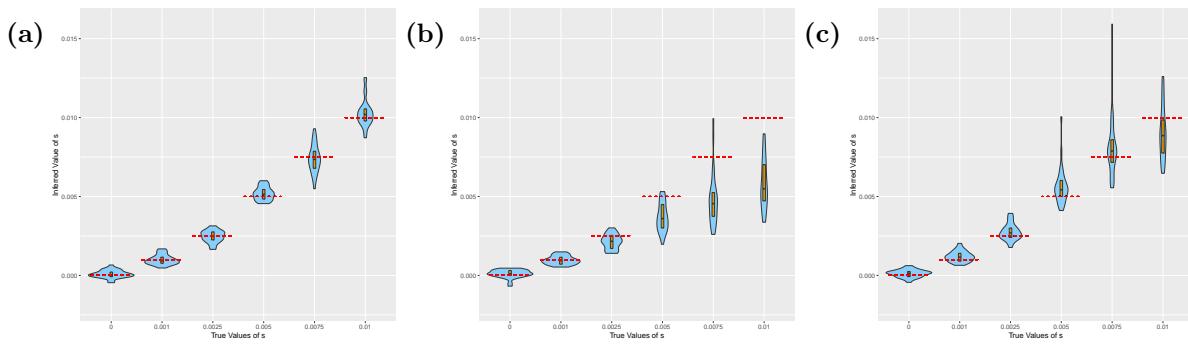


Fig 4: Violin plots showing the results of running CLUES2 on inferred topologies. Boxplots are overlaid with the whiskers omitted. True values of s are shown as dashed red lines. 30 replicates were performed for each true value of s . Simulations were run with (a) $\mu = 7 \times 10^{-7}$ and 1 sample taken without importance sampling, (b) $\mu = 2 \times 10^{-9}$ and 1 sample taken without importance sampling, and (c) $\mu = 2 \times 10^{-9}$ and 3000 samples taken and used in the importance sampling framework.

221 Ancient ARGs

222 Recently, there has been significant interest in using the increasing number of high-quality ancient
223 genomes to infer historic patterns of selection. However, while the usage of time series data has the
224 potential to greatly improve estimates of selection, most existing methods only analyze historic genotype
225 data and do not incorporate information from linked SNPs. (Malaspinas et al., 2012; Mathieson and
226 Terhorst, 2022; Le et al., 2022) In order to fully utilize all the information present in available ancient
227 data, however, it would be desirable to have a method that can incorporate linkage information from
228 SNPs around a locus of interest. To this end, we implement the usage of ancestral recombination graphs
229 on ancient genomes in CLUES2 to incorporate both temporal data and linkage data around the focal site.
230 The general HMM framework remains the same, but the possible emissions of the algorithm change. The
231 input data now consists not only of a list of derived coalescence times and ancestral coalescence times, but
232 also a list of sampling times of derived leaves and a list of sampling times of ancestral leaves. If, for a given
233 time t , there are no leaves sampled in the interval $(t, t+1)$, then our emissions are the same as for the case
234 of modern ARGs, $e(t, k) = e_D(t, k)e_A(t, k)$. However, if we sample m_D derived leaves in this interval and
235 m_A ancestral leaves in this interval, then our emission becomes $e(t, k) = e_D(t, k)e_A(t, k)x_k^{m_D}(1-x_k)^{m_A}$ to
236 account for the genotype observations of the leaves. Note that we assume that genotypes are hard-called
237 and do not allow the usage of genotype likelihoods for the leaves of the ARG. After each timestep, we
238 also increase the current number of derived lineages and ancestral lineages as necessary, corresponding
239 to any sampled leaves. To illustrate the correctness of this approach, we simulate ARGs on ancient data
240 under different selection coefficients (see the supplementary section “Simulation Details” for our simula-
241 tion methodology) and find that we obtain approximately unbiased estimates of the selection coefficient
242 (Figure 5a). Note that while these specific simulations use the true gene trees, importance sampling on
243 ancient ARGs is also possible. To show the improvements in accuracy obtained by utilizing this approach,
244 we also ran CLUES2 using only the modern ARGs obtained from these data and utilizing the ancient
245 data only as genotype emissions, rather than incorporating them into the ARG (Figure 5b). We see that
246 while the estimates of selection are still approximately unbiased, the variance in the estimation is larger
247 due to the smaller amount of linkage information being used in the analysis. Therefore, we highlight the
248 ability of this approach to utilize both time series and linkage information to generate the most accurate
249 possible estimates of selection coefficients.

250 Selection in multiple epochs

251 We also provide functionality for the joint inference of selection coefficients that differ between time
252 periods. This is done by changing the mean of the normal distribution for the transition probabilities
253 from $x_k - sx_k(1 - x_k)$ to $x_k - s_t x_k(1 - x_k)$ where s_t depends on the timestep t . In particular we allow
254 time breakpoints τ_1, \dots, τ_n to be specified, which results in $n + 1$ different selection coefficients being fit,

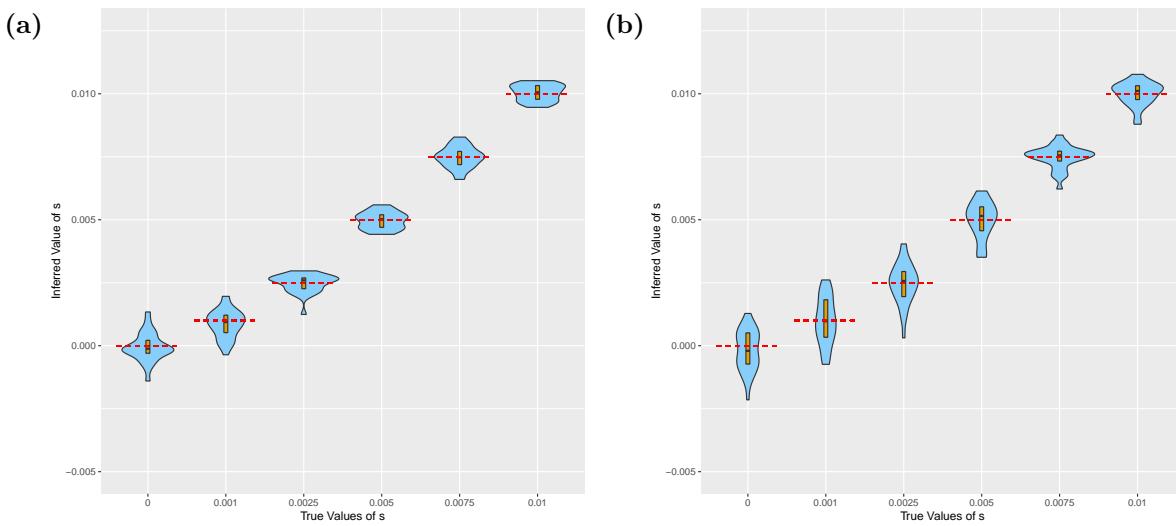


Fig 5: Violin plots showing the results of running CLUES2 on a combination of modern and ancient data. Boxplots are overlaid with the whiskers omitted. True values of s are shown as dashed red lines. 30 replicates were performed, with each replicate generating an estimate for each of the 3 selection coefficients. We run simulations (a) where the ancient data is incorporated into the tree and (b) where the ancient data is treated only as genotype emissions.

255 one for each of the epochs $[0, \tau_1], [\tau_1, \tau_2], \dots, [\tau_n, T]$. To validate our approach, we simulated ancient
 256 genotype data from a model where the selection coefficient in the epoch $[0, 200)$ was 0.01, the selection
 257 coefficient in the epoch $[200, 600)$ was 0 and the selection coefficient in the epoch $[600, \inf)$ was -0.005 .
 258 Population size is held constant at $2N = 300000$, meaning that there is very little drift in the population.
 259 We sample 1 diploid individual every generation, and their genotype is recorded. We find that we are
 260 accurately able to jointly infer these 3 different selection coefficients (Figure 6).

261 Reconstructing Historic Allele Frequencies

In addition to the inference of selection coefficients and log-likelihood ratios of selection, CLUES2 also has the capability to reconstruct historic allele frequency trajectories by calculating, for each time t , the posterior distribution over allele frequencies given the estimate of s and the data D . This is accomplished by running the backward algorithm, the complement to the forward algorithm, which consists of calculating a matrix B following the recursion

$$B_{t,k} = \sum_{l=1}^K e(t+1, l) P_{k,l} B_{t+1,l}$$

262 where the last column is initialized to all ones. Then, the desired posterior on allele frequency for a given
 263 time point t , call it π_t , has the form $\pi_t(k) \propto F_{t,k} B_{t,k}$.

264 In the original CLUES algorithm, the full forward-backward algorithm is run with the only value of s
 265 considered being the estimate of \hat{s}^{MLE} . However, we find that this does not incorporate the uncertainty

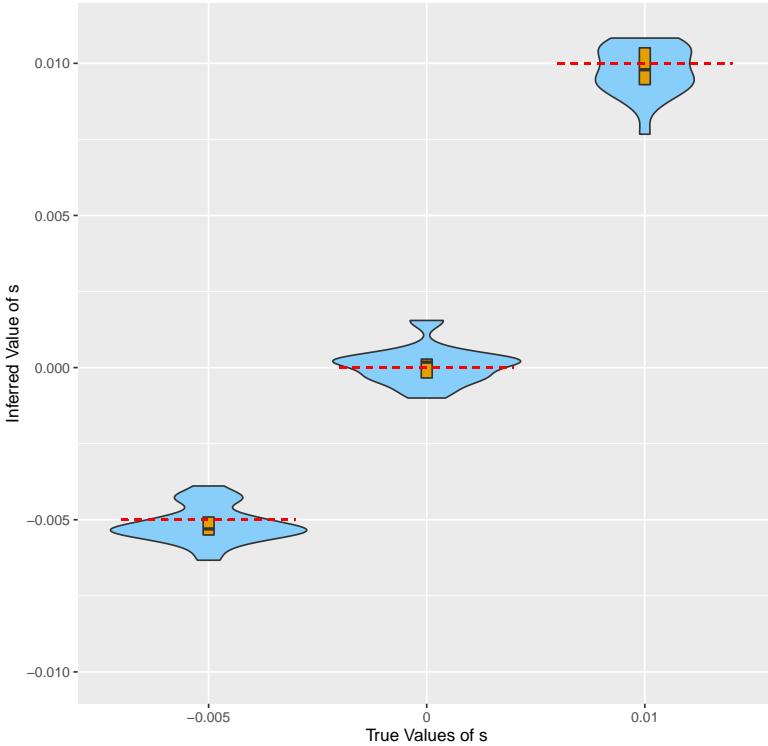


Fig 6: Violin plots showing the results of running CLUES2 on genotype data simulated with differing selection coefficients through time. Boxplots are overlaid with the whiskers omitted. True values of s are shown as dashed red lines. 30 replicates were performed, with each replicate generating an estimate for each of the 3 selection coefficients.

- 266 inherent to the estimation of s and therefore results in posterior distributions that are underdispersed.
- 267 We instead wish to compute the posterior distribution on allele frequencies $P(X_t = k|D)$ assuming a
- 268 uniform prior on s . With a uniform prior on s , this distribution is:

$$\begin{aligned}
 P(X_t = k|D) &= \int_{-\infty}^{\infty} P(X_t = k, s|D) ds \\
 &= \int_{-\infty}^{\infty} P(X_t = k|s, D) P(s|D) ds \\
 &\propto \int_{-\infty}^{\infty} P(X_t = k|s, D) P(D|s) ds \\
 &= \int_{-\infty}^{\infty} \pi_t^s(k) P(D|s) ds
 \end{aligned}$$

In theory, one could sample a set of discrete values for s , call them s_1, \dots, s_N , and approximate this integral using Simpson's rule, numerical quadrature, or any other way for deterministically approximating the area under the curve of interest. However, these methods scale poorly to higher-dimensional spaces, which will be the case when we are fitting multiple selection coefficients. For this reason, we choose to use Monte Carlo integration, as the variance in the estimate of the integral does not depend on the dimension of the state space but is instead always proportional to $1/M$ where M denotes the number of points used

in the estimation, (see e.g. (Jarosz, 2008)). Our Monte Carlo estimator is

$$\int_{-\infty}^{\infty} \pi_t^s(k) P(D|s) ds \approx \frac{1}{M} \sum_{i=1}^M \pi_t^{s_i}(k)$$

where the s_i are sampled from $P(D|s)$. To do this, we assume that $P(D|s)$ is approximately normally distributed with mean equal to \hat{s}^{MLE} as computed by our optimization algorithm. This approximation is based on the Bernstein-von Mises theorem, which states that asymptotically, the posterior is normally distributed with a variance given by the Fisher information matrix. Therefore, in the limit of large data and many Monte Carlo samples, this approximation becomes exact.

Using the pairs of selection coefficients and values of the likelihood function $P(D|s)$ computed during our optimization routine, call them $(s_j, L(s_j))$, we then fit the variance (or covariance matrix) to our data points using a least-squares regression model. Concretely, we find the value of Σ that minimizes

$$\sum_j (L(s_j) - \phi(s_j; \hat{s}^{MLE}, \Sigma))^2$$

where $\phi(x; \mu, \Sigma)$ is the density of a normal distribution with mean μ and covariance matrix Σ evaluated at the point x . We then sample a number of points M from this distribution (10d by default, where d is the dimension of s) and compute $\frac{1}{M} \sum_{i=1}^M \pi_t^{s_i}(k)$ as our estimate of the posterior distribution over allele frequencies at time t . In this way, we take uncertainty in our estimation of s into account for our estimate of the posterior distribution of allele frequencies. We validate this approach by comparing it to an exact rejection sampling approach for reconstructing historic allele frequencies. Specifically, we fix a set of genotype observations and run CLUES2 with our Monte Carlo integration approach. Then, we repeatedly sample selection coefficients uniformly from $[-0.1, 0.1]$ and simulate allele histories conditional on our sampled coefficient. We then sample genotype observations conditional on these trajectories and retain only those trajectories associated with datasets that match our fixed set of genotype observations. This is done until we have retained 1000 trajectories, from which we can reconstruct empirical posterior frequency intervals. We compare these two approaches in Figure S5. We also perform a comparison where 2 selection coefficients are independently inferred, where the coefficient in each interval is sampled uniformly from $[-0.1, 0.1]$ (Figure S6). For both simulations, we find strong concordance between our Monte Carlo integration framework and the exact rejection sampling approach.

Validation of chi-squared test

Along with reporting estimates of selection coefficients, we report the log-likelihood ratio $\ln \frac{P(D|\hat{s}=s^{MLE})}{P(D|s=0)}$ and the associated p-value for selection. By Wilks' Theorem, if data is simulated under the null hypothesis of $s = 0$, $2 \ln \frac{P(D|\hat{s}=s^{MLE})}{P(D|s=0)}$ should asymptotically follow a χ^2 distribution with degrees of freedom equal

293 to the difference in the dimensions of the parameter spaces of \hat{s}^{MLE} and $s = 0$ (Wilks, 1938). As $s = 0$
294 has dimension 0, the degrees of freedom is always equal to the number of different selection coefficients
295 that are estimated in different epochs (which is 1 by default). Furthermore, the p-value obtained from
296 this chi-squared distribution should be uniformly distributed on $[0, 1]$ under the null hypothesis.

297 We therefore perform validations on each of our above simulations by examining the distribution of
298 $2 \log(LR)$ when data is simulated under $s = 0$ to the expected chi-squared distribution with k degrees
299 of freedom where k is the number of independent selection coefficients that are estimated (3 in the
300 section “Selection in multiple epochs”, 1 in all other simulations). We compare the quantiles of the
301 empirical distribution of $2 \log(LR)$ to the theoretical chi-squared distribution using a P-P plot. We
302 also plot a histogram with 5 bins of the p-values obtained from these log-likelihood ratios by using
303 the upper tail of the appropriate chi-squared distribution and comparing it to the uniform distribution.
304 The results are shown in Supplementary Figures S7 through S12. We find that for each simulation we
305 consider, the empirical distribution of $2 \log(LR)$ follows the expected chi-squared distribution and that
306 the corresponding p-values are distributed uniformly between 0 and 1, thus indicating that CLUES2 is
307 properly calibrated and can be used in rigorous hypothesis testing frameworks.

308 Computational Improvements

309 We make modifications to the naive forward-backward algorithm that significantly improve the computa-
310 tional runtime of CLUES2. In practice, there are two steps of the basic forward algorithm implementation
311 of CLUES2 that require significant computational runtime. The first is the computation of the $K \times K$
312 per-generation allele frequency transition matrix P . Naively, this requires the computation of K^2 entries.
313 However, many of these probabilities will be close to 0 as the per-generation probability of transitioning
314 from a high allele frequency to a low allele frequency in one generation and vice-versa is very small. More
315 formally, we observe that the nature of the Gaussian transition probabilities ensures that the transition
316 matrix is a sparse banded matrix where appreciable probability only falls near the main diagonal. There-
317 fore, we instead, for each allele frequency bin k , compute only the entries in the matrix corresponding to
318 the probability of transitioning to frequencies lying in the range $[\max(0, k - \epsilon), \min(1, k + \epsilon)]$. In practice,
319 we set $\epsilon = 0.05$. For frequencies from ϵ to $1 - \epsilon$, this means that the possible frequencies in the next gen-
320 eration fall in a range representing 10% of the frequency space, meaning that the computational speedup
321 should be approximately 10x for this step. We note that this idea of approximating the full per-generation
322 transition matrix by a sparse banded matrix has been previously applied to the full Wright-Fisher model
323 (Spence et al., 2023). We call this Approximation A.

The other section that requires significant time is the execution of the forward algorithm, which must
be run every time the likelihood function is called during the optimization procedure (or M times for each
call to the likelihood function if importance sampling is used with M samples). Each call to the algorithm

integrates over all derived allele trajectories from time 0 to time T , though the dynamic programming nature of the forward algorithm means that this integration is done in polynomial time. However, some of these trajectories have vanishingly small probabilities, such as the oscillating trajectory where $x_i = 0$ for i odd and $x_i = 1$ for i even. We instead modify the forward algorithm to only integrate over likely allele trajectories. When looking at the forward algorithm equation:

$$F_{t,k} = e(t, k) \sum_{l=1}^K F_{t-1,l} P_{l,k}$$

We notice that many of these transition probabilities $P_{l,k}$ will be close to 0 if l is quite far from k (and if $|x_k - x_l| > 0.05$, the probability will be exactly 0 based on our revised computation of the transition matrix P). This means that these values of l can be left out of the summation without significantly affecting the sum. Therefore, we compute, for all K columns of P , a lower row index a_k and upper row index b_k such that the sum of the entries between a_k and b_k is at least 99.9% of the total sum of the entries in column k . We then replace the above sum by

$$F_{t,k} = e(t, k) \sum_{l=a_k}^{b_k} F_{t-1,l} P_{l,k}$$

324 which saves considerable computational time while still retaining accuracy. These lower and upper bounds
 325 can be computed at the same time as the transition matrix P is computed and do not need to be
 326 recomputed at each timestep of the forward algorithm. Conceptually, this is equivalent to replacing the
 327 Gaussian diffusion approximation with a Gaussian distribution that is truncated at extreme values. We
 328 call this Approximation B.

We also make another computational improvement based on the observation that when computing

$$F_{t,k} = e(t, k) \sum_{l=a_k}^{b_k} F_{t-1,l} P_{l,k}$$

329 even for indices, l , with fairly large values of $P_{l,k}$, the computed sum might still be near 0 if the values
 330 of $F_{t-1,l}$ are quite small. Therefore, we choose to simply not compute values of $F_{t,k}$ if we can reasonably
 331 conclude that they will be near 0. We do this by calculating a lower bound α_t and an upper β_t between
 332 which most of the probability of the column $F_{t,\cdot}$ lies (i.e. bounds for which $\frac{\sum_{k=\alpha}^{\beta} F_{t,k}}{\sum_{k=1}^K F_{t,k}} \approx 1$). We then
 333 compute the entries $F_{t+1,k}$ only for the k satisfying $\alpha_t - K/20 < k < \beta_t + K/20$, where the $K/20$ terms
 334 are chosen to correspond to the $\epsilon = 0.05$ restriction on the transition matrix. The other entries in that
 335 column of F are left as 0. α_{t+1} and β_{t+1} are then set respectively to the minimum and maximum indices
 336 k for which $\frac{F_{t+1,k}}{\max_{k'} F_{t+1,k'}} > e^{-200}$. We set α_1 and β_1 to both be equal to the index of the row corresponding
 337 to the allele frequency bin closest to the observed modern allele frequency. Informally, we are assuming
 338 that each column of F is relatively close in distribution to the previous column of F . This is a reasonable

339 assumption if the emissions at a given timestep do not significantly change the probability distribution
340 across the hidden states. The only emission that does violate this assumption is the coalescence of the
341 mixed lineage, as we enforce the fact that the derived allele frequency must be 0 at this time (as we
342 described in section “True Trees”). For this reason, if we observe only 1 derived lineage left at time t , we
343 set α_t to 1, so we always compute $F_{t,1}$ and all values $F_{t,k}$ for $k < \beta_k$ and are therefore never “surprised” by
344 observing this coalescence. We find that this approximation saves significant computational time and has
345 a negligible effect on the estimated value of the selection coefficient. This practice of running the forward
346 algorithm while only keeping track of a smaller number of “best” states is inspired by the beam search
347 approaches used to approximate the Viterbi decoding of an HMM (Deshmukh et al., 1999), although the
348 exact details differ. We call this Approximation C. The concepts underlying Approximations A, B, and
349 C are shown in Figure 7.

350 For the backward algorithm, we also perform Approximation A, and make an analogous approximation
351 to Approximation B, where a_k and b_k instead represent indices for which the sum of the entries between a_k
352 and b_k is at least 0.999 of the total sum of the entries in row k , as opposed to column k . We do not make
353 an analogous approximation for Approximation C. In practice, this set of computational improvements
354 greatly improves the runtime of CLUES2 compared with the original CLUES algorithm. We show this by
355 measuring the runtime of both CLUES and CLUES2 as a function of the number of importance samples
356 used for a given dataset. We plot the results in Figure 8.

357 We see that CLUES2 is over 10 times faster when considering only 1 importance sample and is over
358 100 times faster when 100 importance samples are used. These computational speedups are important
359 for several reasons. Firstly, they allow more importance samples to be used for the same amount of
360 computational resources, which contributes to better convergence of the Monte Carlo estimator of the
361 log-likelihood ratio (Equation 1) and thus improved accuracy. Secondly, it expands the applicability of
362 CLUES2 to large numbers of SNPs, for example in genome-wide scans of selection. We show that these
363 approximations have a negligible impact on our estimate of the log-likelihood function by comparing
364 results both with and without these approximations (see Figure 9).

365 Inferring Gene Trees on Ancient Human Data

366 We applied CLUES2 to a recently published collection of imputed and phased ancient genomes (Allentoft
367 et al., In Press). This dataset consists of 1664 diploid individuals sampled from around the world and
368 has both newly sequenced genomes as well as previously published genomes. The highest concentration
369 of these genomes is located in Western Eurasia, and for that reason, we restricted our analysis to West
370 Eurasian genomes. In addition, it has been shown that inferring ARGs on very low coverage data can
371 cause biases in the estimation of pairwise coalescent times (see Supplementary Section 2 of (Allentoft
372 et al., In Press)), so we chose to limit our analysis to a subset of 187 sampled West Eurasian genomes

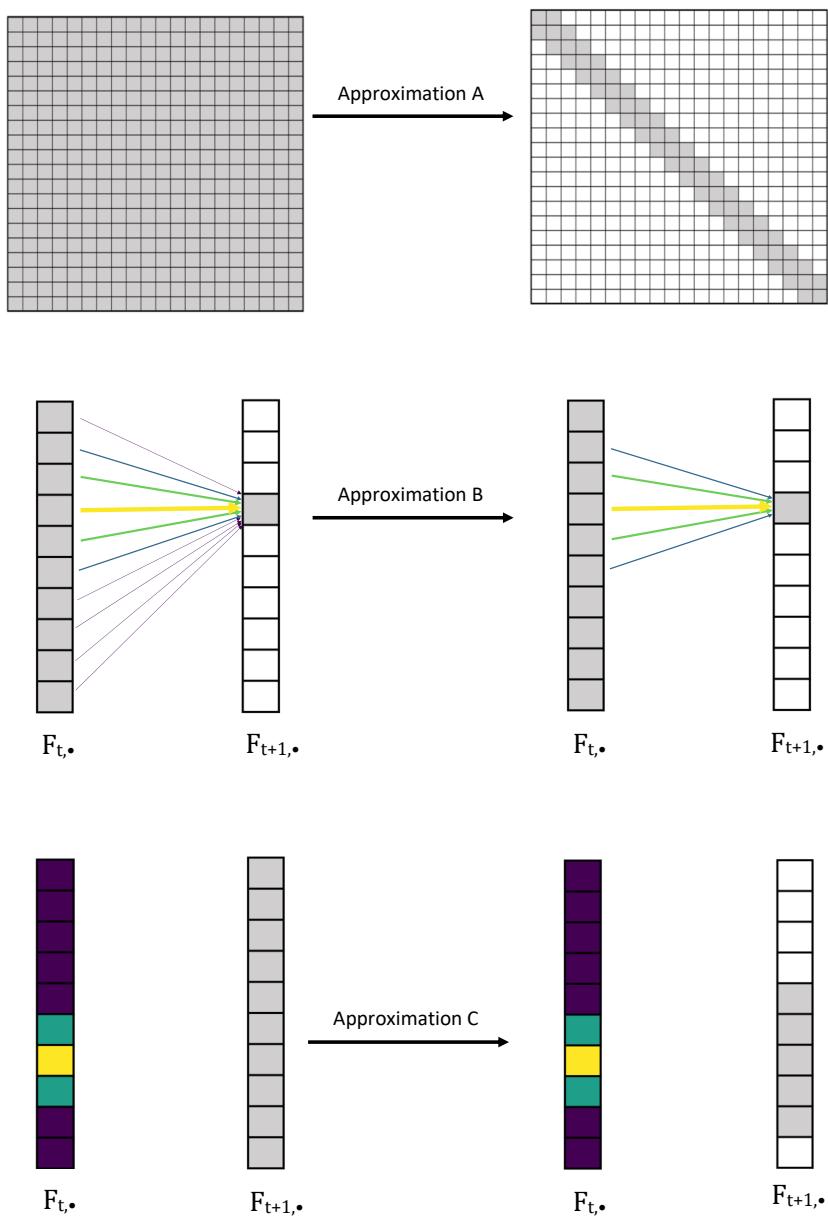


Fig 7: Illustration of Approximations A, B, and C. Approximation A approximates the transition matrix by a sparse banded matrix. Approximation B reduces the number of states in the previous column of F that are summed over to compute each entry of the forward matrix F . Approximation C reduces the number of entries that are computed in a column of the forward matrix F based on the probability density of the previous column of F . Here, the gray entries represent values that are computed, while the colored entries and arrows represent transition or forward probabilities. Lighter colors denote higher probabilities, and darker colors denote smaller probabilities.

373 with a coverage of at least 2x. We then merged this ancient data with a dataset of 100 randomly
 374 chosen individuals from the 1000 Genomes Phase 3 EUR Superpopulation, resulting in 287 total diploid
 375 individuals (Consortium, 2015).

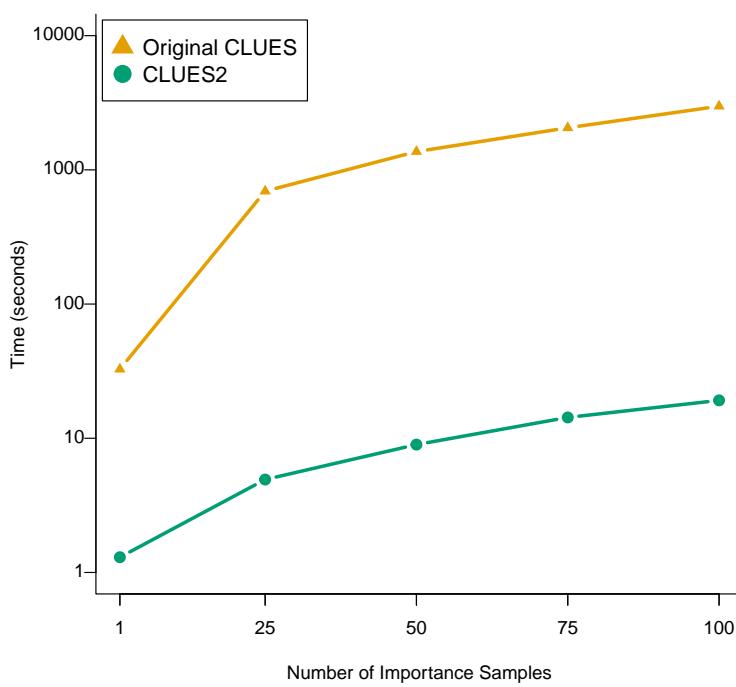


Fig 8: Comparative runtime of CLUES and CLUES2 on different numbers of importance samples.

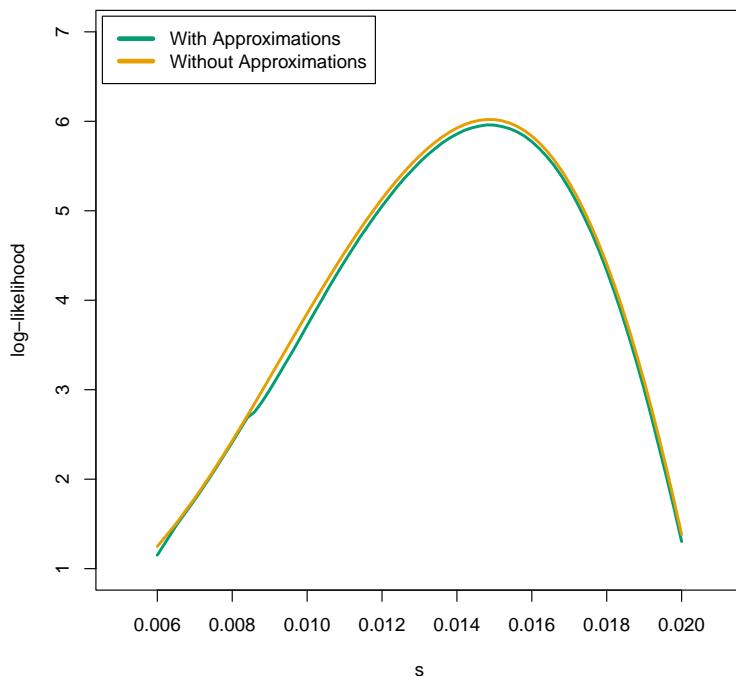


Fig 9: Comparison of the estimates of the log-likelihood function of our dataset of 100 importance samples both with and without our HMM approximations. The log-likelihood was evaluated at 60 values of s spaced equally between 0.006 and 0.02 for each case, and the plots of the functions were generated via linear interpolation.

376 To analyze a chosen set of SNPs, we ran Relate (Speidel et al., 2019) on the corresponding chromosomes
377 of our merged dataset, using recombination maps and genomic masks from the 1000 Genomes Project
378 Phase 3 in addition to the GRCh37 human ancestral sequence (Consortium, 2015) to polarize the alleles.
379 The `EstimatePopulationSize` capability was used in order to generate a .coal file representing the
380 estimated historic population sizes, which was then used as input to the main Relate algorithm. The
381 `SampleBranchLengths` function was used to generate 2000 samples of the marginal tree at each focal SNP,
382 and we converted this output to the CLUES2 input using a custom Python script `RelateToCLUES.py`.
383 For certain SNPs, the sampled topologies did not satisfy the infinite sites assumption with respect to the
384 focal SNP. When this occurs, we performed the minimum number of “leaf flips” such that the infinite
385 sites assumption is satisfied. A “leaf flip” consists of switching the allele of a leaf node from derived
386 to ancestral or vice versa. Note that while the minimum number of flips is always well-defined, the
387 exact leaves to be flipped is not necessarily unique. If this is the case, one set of leaves to flip is chosen
388 deterministically based on the ordering of leaf nodes in the given Newick representation. We list the
389 number of leaf flips performed for each SNP in Table 2. We than ran CLUES2 on each of these input
390 files with the arguments `--df 600`, `--tCutoff 536` (which, assuming a generation time of 28 years,
391 (Moorjani et al., 2016) corresponds to 15000 years) and a value of `--popFreq` that was estimated from
392 our 100 modern individuals.

393 Results

394 Comparison with Existing Methods

395 A variety of existing methods have been used to identify selection coefficients. It is not directly possible
396 to test all features of CLUES2 against these methods as to our knowledge CLUES2 is the first method
397 that can model changing selection coefficients through time in a hypothesis-testing framework and that
398 can do this on samples of ARGs on ancient data. However, we do compare the results of CLUES2
399 with other methods on our low-mutation simulated genotype data described in the section “Importance
400 Sampling of Trees” and plotted in Figure 4c. As this data consists of all modern samples and we are only
401 estimating a constant selection coefficient through time, many existing methods are applicable to this
402 data. We consider 3 other methods, all based on summary statistics of the data. The first is Tajima’s D,
403 a site-frequency-spectrum based statistic which is based on the relative difference between the number
404 of segregating sites and the nucleotide diversity from that expected under neutrality. (Tajima, 1989).
405 The second is H12, a haplotype-based statistic that calculates the imbalance in the relative frequencies
406 of the different haplotypes observed in the data. (Garud et al., 2015) The third is nSL, which is based on
407 calculating the average number of segregating sites around a focal SNP that are identical by state among
408 haplotypes with the derived allele, calculating the corresponding average among haplotypes with the

409 ancestral allele, and taking the log-ratio of the 2 quantities. (Ferrer-Admetlla et al., 2014) We measure
410 the ability of these 3 methods to infer selection coefficients through an ABC approach, detailed in the
411 following pseudocode.

Algorithm 1 Approximate Bayesian Computation Algorithm

1. Given dataset D and statistic θ , compute $\theta(D)$
 2. Initialize $m=0$ and $S = \{\}$
 3. While $m < M$:
 - (a) Sample \tilde{s} uniformly from the interval $(0, 0.025)$
 - (b) Simulate dataset \tilde{D} with selection coefficient \tilde{s}
 - (c) Compute $\theta(\tilde{D})$
 - (d) If $|\theta(\tilde{D}) - \theta(D)| < \epsilon$:
 - i. $m \leftarrow m + 1$
 - ii. $S \leftarrow S \cup \{\tilde{s}\}$
 4. Compute empirical density f of samples in S
 5. Output empirical mode of f
-

412 Notice that we use a uniform $(0, 0.025)$ prior, which allows the empirical mode to be interpreted as
413 an approximation to the maximum likelihood estimator. We run this algorithm for each dataset using
414 $M = 2000$ iterations, and ϵ values of 0.015, 0.001, and 0.05 for Tajima's D, H12, and nSL respectively.
415 The posterior is estimated, for each set of resulting ABC sampled values of s (S), using kernel density
416 estimation with the Epanechnikov (parabolic) kernel and a bandwidth specified by Silverman's 'rule of
417 thumb' (Epanechnikov, 1969; Silverman, 1986). We found that increasing M and decreasing ϵ past these
418 thresholds did not significantly change the estimates. It is worth noting that because of the relatively small
419 set of values H12 can take for 24 haplotypes (it is bounded above by the partition number $p(24)=1575$),
420 we can explicitly calculate that the value of $\epsilon = 0.001$ reduces this ABC approach to exact rejection
421 sampling. This is to say that two datasets of 24 haplotypes have H12 values that differ by less than
422 0.001 if and only if their H12 values are exactly the same. We plot the results of this analysis compared
423 with the CLUES2 results in Figure 10. Note that we only analyze the 4 largest selection coefficients
424 as `msprime` cannot simulate sweeps with negative selection coefficients, meaning that edge effects in the
425 density estimation can happen for very small selection coefficients.

426 We see that all the summary statistic based methods show much greater variance than CLUES2,
427 which is to be expected given the comparative lack of information present in summary statistics when
428 compared with the full dataset.

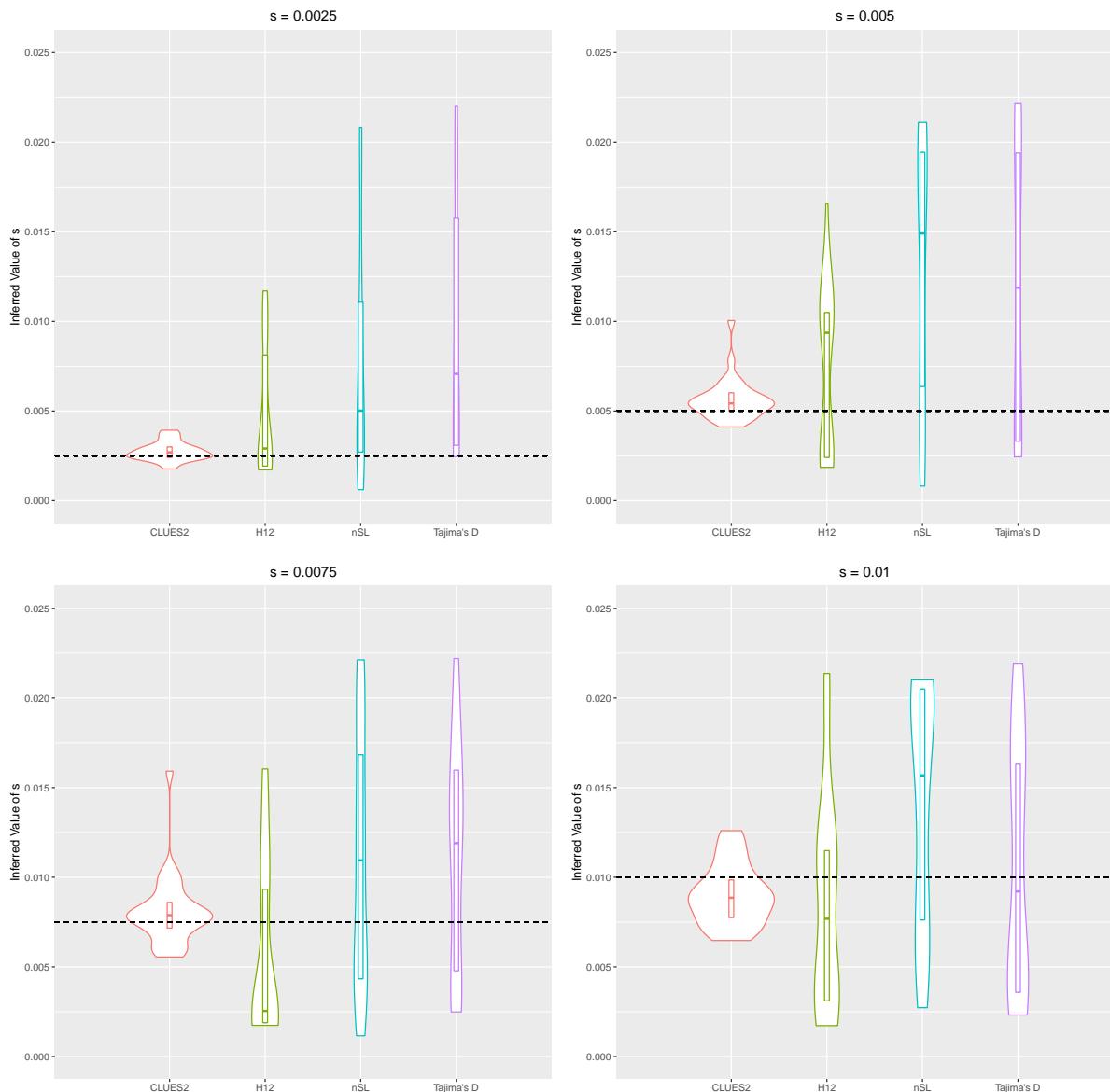


Fig 10: Comparison of the ability of CLUES2 and several summary statistic based methods to infer selection coefficients. We infer selection coefficients from the summary statistics using the ABC algorithm described above. The dataset is identical to that analyzed in Figure 4c. The true simulated values of the selection coefficients are shown as horizontal dashed lines.

429 Analysis of Ancient Human Data

430 We chose 4 SNPs to analyze in the ancient human dataset that have previously been identified as can-
 431 didates for selection in Eurasians. The first is rs4988235, located in the MCM6 gene, where the derived
 432 allele is associated with lactase persistence (Enattah et al., 2002; Bersaglieri et al., 2004; Chin et al.,
 433 2019). The second is rs35395 in the SLC45A2 gene, where the derived allele is associated with lighter
 434 skin pigmentation (Marcheco-Teruel et al., 2014; Tiosano et al., 2016; Lloyd-Jones et al., 2017; Lona-
 435 Durazo et al., 2019). The third is rs12153855 in the TNXB gene of the HLA region, where the derived
 436 allele is associated with age-related macular degeneration and atopic dermatitis (Cipriani et al., 2012;
 437 Weidinger et al., 2013; Grange et al., 2015; Ye et al., 2016). The fourth is rs75393320 in the ACP2 gene,

438 where the derived allele is associated with increased HDL cholesterol (the so-called “good cholesterol”)
 439 (Klarin et al., 2018; Ke et al., 2021). General information about each SNP is described in Table 1, where
 440 the derived allele frequency is calculated from our 100 sampled EUR individuals from the 1000 Genomes
 441 Project.

Gene	SNP rsID	GRCh37 Position	Anc/Der Allele	Derived Allele Freq
MCM6	rs4988235	chr2:136608646	G/A	0.535
SLC45A2	rs35395	chr5:33948589	T/C	0.930
TNXB	rs12153855	chr6:32074804	T/C	0.125
ACP2	rs75393320	chr11:47266471	G/C	0.150

Table 1: General information about the 4 SNPs we considered in our ancient human analysis.

442 For each SNP, we calculated the maximum likelihood estimate of the selection coefficient, the p-value
 443 as computed from the log-likelihood ratio, and the resulting Akaike Information Criterion (AIC) (Akaike,
 444 1974). We ran CLUES2 on each of these SNPs assuming a constant value of the selection coefficients
 445 through time, with a selection coefficient breakpoint at 5000 years ago, and with two breakpoints at 2500
 446 years ago and 5000 years ago. We show the results in Tables 2 to 4, where we group the results by the
 447 number of selection breakpoints. We also plot our estimates of the derived allele frequency trajectories
 448 for each SNP under the model with the lowest AIC in Supplementary Figures S1 to S4.

Gene	$s_{0-15000}^{MLE}$	$-\log_{10}(p)$	AIC	Leaf Flips
MCM6	0.00934	80.86	- 364.01	2
SLC45A2	0.00734	50.30	- 223.74	3
TNXB	0.01011	14.00	-57.88	8
ACP2	0.00461	8.14	-31.45	0

Table 2: Results of the 1-epoch model. One selection selection coefficient is estimated that is constant through time. Bolded rows indicate the genes for which this is the best model as determined by having the lowest AIC. We report the number of leaf flips necessary for each SNP. We do not repeat this information in the following tables as it depends only on the inferred tree structure and thus remains consistent across models.

Gene	s_{0-5000}^{MLE}	$s_{5000-15000}^{MLE}$	$-\log_{10}(p)$	AIC
MCM6	0.01244	0.00498	83.45	-380.30
SLC45A2	0.01434	0.00142	56.37	-255.62
TNXB	0.01804	-0.00655	16.83	-73.51
ACP2	0.00539	0.00367	7.32	-29.73

Table 3: Results of the 2-epoch model. There is one selection coefficient breakpoint at 5000 years before the present. Bolded rows indicate the genes for which this is the best model as determined by having the lowest AIC.

Gene	s_{0-2500}^{MLE}	$s_{2500-5000}^{MLE}$	$s_{5000-15000}^{MLE}$	$-\log_{10}(p)$	AIC
MCM6	0.00674	0.01988	0.00391	83.74	-385.15
SLC45A2	-0.00214	0.03292	0.00074	58.49	-268.55
TNXB	0.02016	0.01441	-0.00607	16.83	-71.86
ACP2	0.00474	0.00650	0.00349	6.65	-27.757

Table 4: Results of the 3-epoch model. There are two selection coefficient breakpoints at 2500 and 5000 years before the present. Bolded rows indicate the genes for which this is the best model as determined by having the lowest AIC.

449 Ancestry-stratified Selection Analysis

450 The selection inference framework of CLUES2 relies on interpreting significant changes in the frequency
451 of an allele as indicative of selection acting on that allele. However, population structure could confound
452 the inference of selection by contributing to the change in the frequency of an allele even in the presence
453 of selective neutrality. In particular, an observed increase in the frequency of an allele in a population
454 could occur in the absence of selection due to either admixture from a source population where that
455 allele is segregating at a higher frequency or due to a correlation between ancestry and sample age. To
456 correct for this, we utilized the local ancestry labelings available for this dataset that classify each SNP
457 in a particular haplotype as belonging to one of 4 major ancestral populations that contribute to the
458 ancestry of modern Europeans: Anatolian farmers (ANA), Caucasus hunter-gatherers (CHG), Western
459 hunter-gatherers (WHG), and Eastern hunter-gatherers (EHG) (Pearson and Durbin, 2023; Irving-Pease
460 et al., In Press). For a given SNP, we then partition each of the ancient haplotypes into these group based
461 on their labelings and compute the modern frequency of the allele in each ancestry group by measuring
462 the percentage of European 1000 Genomes haplotypes of that ancestry which have the derived allele.
463 CLUES2 was then run for each of the 4 ancestries for each SNP using the number of epochs chosen
464 by AIC in our non-stratified analysis described in the previous section. Reconstructed allele frequency
465 trajectories and p-values for the selection tests are plotted in Figures S13 to S16. We highlight that this
466 analysis, in addition to correcting for population structure, only utilizes ancient haplotypes and therefore
467 is not affected by possible biases incurred by the miscalibration of ARG-inference methods.

468 Discussion

469 In this paper, we introduced CLUES2, a flexible, full-likelihood method that is able to utilize more of
470 the information present in sequence data than summary statistics. CLUES2 has the ability to generate
471 unbiased estimates of selection coefficients and also generates well-calibrated p-values in order to run
472 statistical tests for selection. The methodology for our analysis of the West Eurasian dataset highlights
473 the ability of CLUES2 to identify distinct selection coefficients in different epochs. In particular, we
474 emphasize that different hypotheses can be tested for a given SNP without having to obtain new samples
475 of gene trees. Furthermore, the new functionality to analyze ARGs on ancient data enables the usage of
476 both time-series data and information from linked SNPs in selection analyses.

477 For the ancient human DNA analysis, we observe that the 1-selection coefficient model provides the
478 best fit for the ACP2 SNP in the pan-ancestry analysis, corresponding to a classic selective sweep in
479 progress. As increased levels of HDL have been shown to help reduce the risk of myocardial infarction
480 (heart attack) in contemporary medical patients (Remaley et al., 2008), this result indicates that a similar
481 benefit may have existed in ancient Eurasian populations as well. The ancestry-stratified analysis for this

482 SNP suggests that this sweep may have been localized to WHG ancestry. The 2-selection coefficient
483 model provides a best fit for the pan-ancestry analysis of the TNXB SNP, which we see corresponds to
484 a recent period of positive selection preceded by a period during which the derived allele was slightly
485 disfavored. We speculate that this relatively recent selection on a gene in the HLA region, which is
486 known to be associated with immune function (Gough and Simmonds, 2007), could be the result of
487 increasing population density and exposure to domesticated animals that occurred around this time.
488 (Page et al., 2016; Marciniak et al., 2022) The massive demographic shift induced by the introduction of
489 agriculture likely exposed humans to new selective pressures on genes associated with immune function,
490 explaining why we see selective sweeps on certain immunity-associated alleles beginning during this period.
491 The ancestry-stratified analysis for this SNP suggests that this sweep may have been localized to CHG
492 ancestry. Sparse sampling at very old time points contributes to high uncertainty in the allele frequency
493 of the SNP for many ancestry groups.

494 We see that the 3-selection coefficient models provide best fits for the pan-ancestry analysis of the
495 other SNPs. For the MCM6 locus, we find a significant increase in the lactase persistence allele beginning
496 approximately 6000 years before the present. We note that this is thousands of years after the consumption
497 of dairy began in Europe, as evidenced by milk fat residues discovered in potsherds dating back to at
498 least 9000 years before the present (Evershed et al., 2022). This gap, which has previously been noted in
499 several studies (Itan et al., 2009; Mathieson et al., 2015; Burger et al., 2020), suggests that the selective
500 pressure for lactase persistence was not in fact the initial domestication of animals and ensuing increase
501 in lactose consumption, an observation that has led to significant speculation as to what this pressure
502 may have been. One hypothesis is the presence of intermittent famines during this periods caused by
503 crop failure, during which the ability to digest alternative energy sources such as lactose would have
504 been more advantageous (Shennan et al., 2013; Sverrisdóttir et al., 2014). Another hypothesis is that
505 the increased population density during this time period would have caused a larger pathogen load, thus
506 exposing individuals to more common bouts of illness (Loog et al., 2017). As the consumption of lactose
507 by non-lactase persistent individuals can cause diarrhea and therefore significant fluid loss (Smith et al.,
508 2008), it is theorized that this would result in non-lactase persistent individuals having a higher mortality
509 when exposed to these frequent illnesses. While we do not independently verify any of these hypotheses,
510 our finding of an increase in selective pressure for lactase persistence that significantly postdates the first
511 appearance of sustained lactose consumption in Europe supports previous hypotheses that a separate
512 environmental event, rather than the isolated consumption of lactose itself, is what caused the strong
513 selection for lactase persistence. We note that the selection on lactase persistence appears to be continuing
514 to the present day, albeit with a slightly smaller selection coefficient than when the selective pressure first
515 began. The ancestry-stratified analysis for this SNP shows that this general pattern of strong selection
516 beginning approximately 6000 years ago followed by an attenuated selection signal in the most recent time

517 period holds for all non-Anatolian ancestries. We notice that in the WHG ancestry, selection appears
518 to disfavor the lactase persistence allele in the most recent time period, although sparse sampling of
519 ancient WHG haplotypes at this locus in this time period results in high uncertainty of the true allele
520 frequency. We therefore caution against interpreting these results as indicating modern selection against
521 lactase persistence in this ancestry, although our results do indicate that selection is weaker than when
522 the selective sweep first began.

523 Our SLC45A2 analysis appears to capture the time in which the derived allele for rs35395 was under
524 strong selection, which was roughly during the period of 2500-5000 years ago. We find that selection
525 during this period was particularly strong ($s_{2500-5000}^{MLE} = 0.03292$), but the allele appears to not be under
526 significant selective pressure in either younger or older time periods. We note that our finding of strong
527 selection on depigmentation in European populations beginning 5000 years ago is in line with the results
528 of (Wilde et al., 2014) who estimated a selection coefficient of between 2% and 10% for several skin
529 pigmentation associated alleles. The prevailing hypothesis for the evolution of lighter skin in European
530 populations is the resulting increased ability to synthesize vitamin D, a reaction catalyzed by solar UVB
531 radiation, at higher latitudes where the relative concentration of solar radiation is lower (Jablonski and
532 Chaplin, 2000). Therefore, possible explanations for the exact timing of this selective sweep include the
533 settling of environments at higher latitudes during this period and a shift away from a hunter-gatherer diet
534 that included vitamin D-rich fish and wild game to a vitamin D-poor agriculturalist diet. (Richards et al.,
535 2003; Wilde et al., 2014) The ancestry-stratified analysis for this SNP shows that while all ancestries show
536 evidence of very strong selection for skin depigmentation, the timing and strength of this selection does
537 differ between ancestries. A possible explanation for this difference in the onset and strength of selection
538 could be the differing latitudes inhabited by these ancestral groups through time and the possibly differing
539 vitamin-D content of the diets of these distinct ancestral populations.

540 Overall, the results of our ancient DNA analysis illustrate the importance of the introduction of
541 agriculture to Europe and the commensurate increase in population density to changing the selective
542 pressures associated with certain traits. With the functionality of CLUES2 to test for differing selection
543 coefficients in different time periods, it is now possible to identify these different pressures through their
544 effect on the historic frequency of an allele.

545 Despite the methodological advances of CLUES2, there are nevertheless certain limitations of this
546 approach and areas for future research. One such limitation is the effect of possible misspecification of
547 the data-generating process. For example, we do not consider the effect of background selection, which
548 is known to be pervasive in the human genome and to affect the observed patterns of linked neutral
549 mutations (Pouyet et al., 2018; Johri et al., 2021; Buffalo and Kern, 2023). However, we highlight the
550 fact that the CLUES2 framework is quite flexible, allowing for possible extensions to handle model mis-
551 specifications simply by adjusting the relative weighting of the sampled trees in the importance sampling

552 framework and without necessitating any changes to the ARG-inference methods.

553 Two possible sources of error when using CLUES2 on sequence data are improper ARG-sampling
554 (either from poor mixing of the MCMC chain or inherent problems of the algorithms themselves) and
555 using an insufficient number of importance samples in the Monte Carlo estimator. The degree to which
556 these problems are an issue will depend on the exact dataset being analyzed. For example, in regions of
557 the genome with high mutation to recombination rate ratios, ARG inference is easier and the posterior
558 distribution of marginal gene trees will be so concentrated that few samples will be necessary (See Figure
559 4a). However, in datasets with low mutation to recombination rate ratios and with large numbers of
560 samples, the state space of marginal trees may be so large and the amount of information about the
561 length of each branch may be so low that a very large number of importance samples are necessary
562 to effectively integrate over all the uncertainty in the data. With this in mind, to be able to more
563 efficiently utilize large numbers of importance samples, future extensions of CLUES2 may take advantage
564 of the fact that when evaluating the likelihood at a given selection coefficient, only a small number of
565 samples may actually contribute substantially to the sum of the Monte Carlo estimator (Equation 1),
566 a phenomenon known as *weight degeneracy*. (Vázquez and Míguez, 2017) This results in computational
567 time being wasted in computing the forward algorithm for the samples which will not have a substantial
568 contribution. Therefore, techniques from adaptive importance sampling or heuristics for choosing which
569 subset of samples to use in the sum may be used, although this approach may encounter challenges
570 when multiple selection coefficients are estimated. We note that when gene trees are correctly sampled
571 and when a sufficient number of importance samples are used, CLUES2 vastly improves on summary
572 statistic-based methods for inferring selection coefficients, which all seem to suffer from possible biases
573 and significantly larger variances (see Figure 10). Overall CLUES2 represents a significant step forward
574 in selection inference methodology through the use of a full-likelihood model and by being able to fully
575 utilize the information present in large ancient DNA datasets.

576 Acknowledgments

577 We thank the members of the Nielsen Lab at UC Berkeley for their helpful thoughts on the development
578 of this method. In particular, we thank Diana Aguilar Gómez for her extensive testing of the software
579 and recommendations for its input-output format and Yun Deng for his helpful discussions about HMM
580 algorithms and associated approximations. We also thank the researchers at the Centre for GeoGenetics
581 at the University of Copenhagen both for their feedback on CLUES2 and for hosting the authors for an
582 extended research visit, during which much of the work for this project was done.

583 References

- 584 Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic
585 Control*, 19(6):716–723, 1974. doi: 10.1109/tac.1974.1100705.
- 586 Morten E. Allentoft, Martin Sikora, Alba Refoyo-Martínez, Evan K. Irving-Pease, Anders Fischer,
587 William Barrie, Andrés Ingason, Jesper Stenderup, Karl-Göran Sjögren, and Alice Pearson et al.
588 Population genomics of postglacial western eurasia. *Nature*, In Press. doi: 10.1101/2022.05.04.490594.
589 URL <https://www.biorxiv.org/content/early/2022/10/07/2022.05.04.490594>.
- 590 Franz Baumdicker, Gertjan Bisschop, Daniel Goldstein, Graham Gower, Aaron P Ragsdale, Georgia
591 Tsambos, Sha Zhu, Bjarki Eldon, E Castedo Ellerman, and Jared G et al. Galloway. Efficient ancestry
592 and mutation simulation with msprime 1.0. *Genetics*, 220(3):iyab229, 12 2021. ISSN 1943-2631. doi:
593 10.1093/genetics/iyab229. URL <https://doi.org/10.1093/genetics/iyab229>.
- 594 Juraj Bergman, Dominik Schrempf, Carolin Kosiol, and Claus Vogl. Inference in population genetics
595 using forward and backward, discrete and continuous time processes. *Journal of Theoretical Biology*,
596 439:166–180, Feb 2018. doi: 10.1016/j.jtbi.2017.12.008.
- 597 Todd Bersaglieri, Pardis C. Sabeti, Nick Patterson, Trisha Vanderploeg, Steve F. Schaffner, Jared A.
598 Drake, Matthew Rhodes, David E. Reich, and Joel N. Hirschhorn. Genetic signatures of strong recent
599 positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6):1111–1120,
600 2004. doi: 10.1086/421051.
- 601 Jonathan P Bollback, Thomas L York, and Rasmus Nielsen. Estimation of 2Nes From Temporal Allele
602 Frequency Data. *Genetics*, 179(1):497–502, 05 2008. ISSN 1943-2631. doi: 10.1534/genetics.107.085019.
603 URL <https://doi.org/10.1534/genetics.107.085019>.
- 604 Débora Y. C. Brandt, Xinzhu Wei, Yun Deng, Andrew H Vaughn, and Rasmus Nielsen. Evaluation
605 of methods for estimating coalescence times using ancestral recombination graphs. *Genetics*, 221(1):
606 iyac044, 03 2022. ISSN 1943-2631. doi: 10.1093/genetics/iyac044. URL <https://doi.org/10.1093/genetics/iyac044>.
- 608 J M Braverman, R R Hudson, N L Kaplan, C H Langley, and W Stephan. The hitchhiking effect on
609 the site frequency spectrum of dna polymorphisms. *Genetics*, 140(2):783–796, 1995. doi: 10.1093/
610 genetics/140.2.783.
- 611 Vince Buffalo and Andrew D. Kern. A quantitative genetic model of background selection in humans.
612 *bioRxiv*, 2023. doi: 10.1101/2023.09.07.556762. URL <https://www.biorxiv.org/content/early/2023/09/08/2023.09.07.556762>.

- 614 Joachim Burger, Vivian Link, Jens Blöcher, Anna Schulz, Christian Sell, Zoé Pochon, Yoan Diekmann,
615 Aleksandra Žegarac, Zuzana Hofmanová, and Laura Winkelbach. Low prevalence of lactase persistence
616 in bronze age europe indicates ongoing strong selection over the last 3,000 years. *Current Biology*, 30
617 (21), 2020. doi: 10.1016/j.cub.2020.08.033.
- 618 Elizabeth L. Chin, Liping Huang, Yasmine Y. Bouzid, Catherine P. Kirschke, Blythe Durbin-Johnson,
619 Lacey M. Baldiviez, Ellen L. Bonnel, Nancy L. Keim, Ian Korf, and Charles B. Stephensen et al.
620 Association of lactase persistence genotypes (rs4988235) and ethnicity with dairy intake in a healthy
621 u.s. population. *Nutrients*, 11(8):1860, 2019. doi: 10.3390/nu11081860.
- 622 Valentina Cipriani, Hin-Tak Leung, Vincent Plagnol, Catey Bunce, Jane C. Khan, Humma Shahid,
623 Anthony T. Moore, Simon P. Harding, Paul N. Bishop, and Caroline Hayward et al. Genome-
624 wide association study of age-related macular degeneration identifies associated variants in the
625 tnxb–fkbpl–notch4 region of chromosome 6p21.3. *Human Molecular Genetics*, 21(18):4138–4150, 2012.
626 doi: 10.1093/hmg/dds225.
- 627 The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526:
628 68–74, 2015. doi: 10.1038/nature15393. URL <https://doi.org/10.1038/nature15393>.
- 629 Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJour-
630 nal, Complex Systems*:1695, 11 2005.
- 631 N. Deshmukh, A. Ganapathiraju, and J. Picone. Hierarchical search for large-vocabulary conversational
632 speech recognition: working toward a solution to the decoding problem. *IEEE Signal Processing
633 Magazine*, 16(5):84–107, 1999. doi: 10.1109/79.790985.
- 634 Nabil Sabri Enattah, Timo Sahi, Erkki Savilahti, Joseph D. Terwilliger, Leena Peltonen, and Irma Järvelä.
635 Identification of a variant associated with adult-type hypolactasia. *Nature Genetics*, 30(2):233–237,
636 2002. doi: 10.1038/ng826.
- 637 V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probab-
638 ity and Its Applications*, 14(1):153–158, 1969. doi: 10.1137/1114019.
- 639 Richard P. Evershed, George Davey Smith, Mélanie Roffet-Salque, Adrian Timpson, Yoan Diekmann,
640 Matthew S. Lyon, Lucy J. Cramp, Emmanuelle Casanova, Jessica Smyth, and Helen L. Whelton et al.
641 Dairying, diseases and the evolution of lactase persistence in europe. *Nature*, 608(7922):336–345, 2022.
642 doi: 10.1038/s41586-022-05010-7.
- 643 Justin C Fay and Chung-I Wu. Hitchhiking under positive darwinian selection. *Genetics*, 155(3):
644 1405–1413, 2000. doi: 10.1093/genetics/155.3.1405.

- 645 Anna Ferrer-Admetlla, Mason Liang, Thorfinn Korneliussen, and Rasmus Nielsen. On detecting incom-
646 plete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, 31(5):
647 1275–1291, 2014. doi: 10.1093/molbev/msu077.
- 648 Nandita R. Garud, Philipp W. Messer, Erkan O. Buzbas, and Dmitri A. Petrov. Recent selective sweeps in
649 north american drosophila melanogaster show signatures of soft sweeps. *PLOS Genetics*, 11(2):1–32, 02
650 2015. doi: 10.1371/journal.pgen.1005004. URL <https://doi.org/10.1371/journal.pgen.1005004>.
- 651 S. Gough and M. Simmonds. The hla region and autoimmune disease: Associations and mechanisms of
652 action. *Current Genomics*, 8(7):453–465, 2007. doi: 10.2174/138920207783591690.
- 653 Laura Grange, Jean-François Bureau, Iryna Nikolayeva, Richard Paul, Kristel Van Steen, Benno
654 Schwikowski, and Anavaj Sakuntabhai. Filter-free exhaustive odds ratio-based genome-wide inter-
655 action approach pinpoints evidence for interaction in the hla region in psoriasis. *BMC Genetics*, 16(1),
656 2015. doi: 10.1186/s12863-015-0174-3.
- 657 R. C. Griffiths and P. Marjoram. Ancestral inference from samples of dna sequences with recombination.
658 *Journal of Computational Biology*, 3(4):479–502, 1996. doi: <https://doi.org/10.1089/cmb.1996.3.479>.
- 659 Benjamin C. Haller, Jared Galloway, Jerome Kelleher, Philipp W. Messer, and Peter L. Ralph. Tree-
660 sequence recording in slim opens new horizons for forward-time simulation of whole genomes. *Molecular
661 Ecology Resources*, 19(2):552–566, 2019. doi: <https://doi.org/10.1111/1755-0998.12968>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.12968>.
- 663 Hussein A Hejase, Ziyi Mo, Leonardo Campagna, and Adam Siepel. A deep-learning approach for
664 inference of selective sweeps from the ancestral recombination graph. *Molecular Biology and Evolution*,
665 39(1), 2021. doi: 10.1093/molbev/msab332.
- 666 Richard R. Hudson. Properties of a neutral allele model with intragenic recombination. *Theoretical
667 Population Biology*, 23(2):183–201, 1983. ISSN 0040-5809. doi: [https://doi.org/10.1016/0040-5809\(83\)90013-8](https://doi.org/10.1016/0040-5809(83)
668 90013-8). URL <https://www.sciencedirect.com/science/article/pii/0040580983900138>.
- 669 Evan K. Irving-Pease, Alba Refoyo-Martínez, Andrés Ingason, Alice Pearson, Anders Fischer, William
670 Barrie, Karl-Göran Sjögren, Alma S. Halgren, Ruairidh Macleod, and et al. Fabrice Demeter.
671 The selection landscape and genetic legacy of ancient eurasians. *Nature*, In Press. doi: 10.
672 1038/s41586-023-06705-1. URL [https://www.biorxiv.org/content/early/2022/10/07/2022.09.22.509027](https://www.biorxiv.org/content/early/2022/10/07/2022.09.
673 22.509027).
- 674 Yuval Itan, Adam Powell, Mark A. Beaumont, Joachim Burger, and Mark G. Thomas. The origins of
675 lactase persistence in europe. *PLoS Computational Biology*, 5(8), 2009. doi: 10.1371/journal.pcbi.
676 1000491.

- 677 Nina G. Jablonski and George Chaplin. The evolution of human skin coloration. *Journal of Human*
678 *Evolution*, 39(1):57–106, 2000. doi: 10.1006/jhev.2000.0403.
- 679 Wojciech Jarosz. Efficient monte carlo methods for light transport in scattering media. page 149–166,
680 2008. URL <https://cs.dartmouth.edu/~wjarosz/publications/dissertation/appendixA.pdf>.
- 681 Parul Johri, Kellen Riall, Hannes Becher, Laurent Excoffier, Brian Charlesworth, and Jeffrey D. Jensen.
682 The impact of purifying and background selection on the inference of population history: Problems and
683 prospects. *Molecular Biology and Evolution*, 38(7):2986–3003, 2021. doi: 10.1093/molbev/msab050.
- 684 N.L. Kaplan, T. Darden, and R.R. Hudson. The coalescent process in models with selection. *Genetics*,
685 120(3):819–829, Nov 1988.
- 686 Wenfan Ke, Jordan N. Reed, Chenyu Yang, Noel Higgason, Leila Rayyan, Carolina Wählby, Anne E.
687 Carpenter, Mete Civelek, and Eyleen J. O'Rourke. Genes in human obesity loci are causal obesity
688 genes in *c. elegans*. *PLOS Genetics*, 17(9):1–33, 09 2021. doi: 10.1371/journal.pgen.1009736. URL
689 <https://doi.org/10.1371/journal.pgen.1009736>.
- 690 Jerome Kelleher, Yan Wong, Anthony W. Wohns, Chaimaa Fadil, Patrick K. Albers, and Gil McVean.
691 Inferring whole-genome histories in large population datasets. *Nature Genetics*, 51(9):1330–1338, 2019.
692 doi: 10.1038/s41588-019-0483-y.
- 693 Andrew D. Kern and Daniel R. Schrider. Discoal: flexible coalescent simulations with selection. *Bioin-*
694 *formatics*, 32(24):3839–3841, 08 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw556. URL
695 <https://doi.org/10.1093/bioinformatics/btw556>.
- 696 Derek Klarin, Scott M. Damrauer, Kelly Cho, Yan V. Sun, Tanya M. Teslovich, Jacqueline Honerlaw,
697 David R. Gagnon, Scott L. DuVall, Jin Li, and Gina M. Peloso et al. Genetics of blood lipids among
698 300,000 multi-ethnic participants of the million veteran program. *Nature Genetics*, 50(11):1514–1523,
699 2018. doi: 10.1038/s41588-018-0222-9.
- 700 Megan K. Le, Olivia S. Smith, Ali Akbari, Arbel Harpak, David Reich, and Vagheesh M. Narasimhan.
701 1,000 ancient genomes uncover 10,000 years of natural selection in europe. *bioRxiv*, 2022. doi:
702 10.1101/2022.08.24.505188. URL <https://www.biorxiv.org/content/early/2022/08/26/2022.08.24.505188>.
- 704 Luke R Lloyd-Jones, Matthew R Robinson, Gerhard Moser, Jian Zeng, Sandra Beleza, Gregory S Barsh,
705 Hua Tang, and Peter M Visscher. Inference on the genetic basis of eye and skin color in an admixed
706 population via bayesian linear mixed models. *Genetics*, 206(2):1113–1126, 2017. doi: 10.1534/genetics.
707 116.193383.

- 708 Frida Lona-Durazo, Natalia Hernandez-Pacheco, Shaohua Fan, Tongwu Zhang, Jiyeon Choi, Michael A.
709 Kovacs, Stacie K. Loftus, Phuong Le, Melissa Edwards, and Cesar A. Fortes-Lima et al. Meta-analysis
710 of gwa studies provides new insights on the genetic architecture of skin pigmentation in recently admixed
711 populations. *BMC Genetics*, 20(1), 2019. doi: 10.1186/s12863-019-0765-5.
- 712 Liisa Loog, Marta Mirazón Lahr, Mirna Kovacevic, Andrea Manica, Anders Eriksson, and Mark G.
713 Thomas. Estimating mobility using sparse data: Application to human genetic variation. *Proceedings
714 of the National Academy of Sciences*, 114(46):12213–12218, 2017. doi: 10.1073/pnas.1703642114.
- 715 Ali Mahmoudi, Jere Koskela, Jerome Kelleher, Yao-ban Chan, and David Balding. Bayesian inference of
716 ancestral recombination graphs. *PLOS Computational Biology*, 18(3), 2022. doi: 10.1371/journal.pcbi.
717 1009960.
- 718 Anna-Sapfo Malaspinas, Orestis Malaspinas, Steven N Evans, and Montgomery Slatkin. Estimating
719 allele age and selection coefficient from time-serial data. *Genetics*, 192(2):599–607, 2012. doi: 10.1534/
720 genetics.112.140939.
- 721 Beatriz Marcheco-Teruel, Esteban J. Parra, Evelyn Fuentes-Smith, Antonio Salas, Henriette N. But-
722 tenschøn, Ditte Demontis, María Torres-Español, Lilia C. Marín-Padrón, Enrique J. Gómez-Cabezas,
723 and Vanesa Álvarez-Iglesias et al. Cuba: Exploring the history of admixture and the genetic basis
724 of pigmentation using autosomal and uniparental markers. *PLOS Genetics*, 10(7):1–13, 07 2014. doi:
725 10.1371/journal.pgen.1004488. URL <https://doi.org/10.1371/journal.pgen.1004488>.
- 726 Stephanie Marciniak, Christina M. Bergey, Ana Maria Silva, Agata Hałuszko, Mirosław Furmanek, Bar-
727 bara Veselka, Petr Velemínský, Giuseppe Vercellotti, Joachim Wahl, and Gunita Zariņa et al. An
728 integrative skeletal and paleogenomic analysis of stature variation suggests relatively reduced health
729 for early european farmers. *Proceedings of the National Academy of Sciences*, 119(15), 2022. doi:
730 10.1073/pnas.2106743119.
- 731 Iain Mathieson and Jonathan Terhorst. Direct detection of natural selection in bronze age britain.
732 *Genome Research*, 32(11–12):2057–2067, 2022. doi: 10.1101/gr.276862.122.
- 733 Iain Mathieson, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Rood-
734 enberg, Eadaoin Harney, Kristin Stewardson, Daniel Fernandes, and Mario Novak et al. Genome-
735 wide patterns of selection in 230 ancient eurasians. *Nature*, 528(7583):499–503, 2015. doi: 10.1038/
736 nature16152.
- 737 Priya Moorjani, Sriram Sankararaman, Qiaomei Fu, Molly Przeworski, Nick Patterson, and David Reich.
738 A genetic method for dating ancient genomes provides a direct estimate of human generation interval
739 in the last 45,000 years. *Proceedings of the National Academy of Sciences*, 113(20):5652–5657, 2016.
740 doi: 10.1073/pnas.1514696113.

- 741 Abigail E. Page, Sylvain Viguier, Mark Dyble, Daniel Smith, Nikhil Chaudhary, Gul Deniz Salali, James
742 Thompson, Lucio Vinicius, Ruth Mace, and Andrea Bamberg Migliano. Reproductive trade-offs in
743 extant hunter-gatherers suggest adaptive mechanism for the neolithic expansion. *Proceedings of the*
744 *National Academy of Sciences*, 113(17):4694–4699, 2016. doi: 10.1073/pnas.1524031113.
- 745 Cyriel Paris, Bertrand Servin, and Simon Boitard. Inference of selection from genetic time series using
746 various parametric approximations to the wright-fisher model. *G3 Genes—Genomes—Genetics*, 9(12):
747 4073–4086, 2019. doi: 10.1534/g3.119.400778.
- 748 Alice Pearson and Richard Durbin. Local ancestry inference for complex population histories. *bioRxiv*,
749 2023. doi: 10.1101/2023.03.06.529121. URL <https://www.biorxiv.org/content/early/2023/03/07/2023.03.06.529121>.
- 750 Benjamin M. Peter, Emilia Huerta-Sánchez, and Rasmus Nielsen. Distinguishing between selective sweeps
751 from standing variation and from a de novo mutation. *PLoS Genetics*, 8(10), 2012. doi: 10.1371/journal.
752 pgen.1003011.
- 753 Fanny Pouyet, Simon Aeschbacher, Alexandre Thiéry, and Laurent Excoffier. Background selection and
754 biased gene conversion affect more than 95% of the human genome and bias demographic inferences.
755 *eLife*, 7:e36317, aug 2018. ISSN 2050-084X. doi: 10.7554/eLife.36317. URL <https://doi.org/10.7554/eLife.36317>.
- 756 Matthew D. Rasmussen, Melissa J. Hubisz, Ilan Gronau, and Adam Siepel. Genome-wide inference of
757 ancestral recombination graphs. *PLoS Genetics*, 10(5), 2014. doi: 10.1371/journal.pgen.1004342.
- 758 Alan T Remaley, Marcelo Amar, and Dmitri Sviridov. Hdl-replacement therapy: Mechanism of action,
759 types of agents and potential clinical indications. *Expert Review of Cardiovascular Therapy*, 6(9):
760 1203–1215, 2008. doi: 10.1586/14779072.6.9.1203.
- 761 Michael P. Richards, Rick J. Schulting, and Robert E. Hedges. Sharp shift in diet at onset of neolithic.
762 *Nature*, 425(6956):366–366, 2003. doi: 10.1038/425366a.
- 763 Pardis C. Sabeti, David E. Reich, John M. Higgins, Haninah Z. Levine, Daniel J. Richter, Stephen F.
764 Schaffner, Stacey B. Gabriel, Jill V. Platko, Nick J. Patterson, and Gavin J. McDonald et al. Detecting
765 recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832–837,
766 2002. doi: 10.1038/nature01140.
- 767 Daniel R. Schrider and Andrew D. Kern. Supervised machine learning for population genetics: A new
768 paradigm. *Trends in Genetics*, 34(4):301–312, 2018. doi: 10.1016/j.tig.2017.12.005.

- 771 Stephen Shennan, Sean S. Downey, Adrian Timpson, Kevan Edinborough, Sue Colledge, Tim Kerig,
772 Katie Manning, and Mark G. Thomas. Regional population collapse followed initial agriculture booms
773 in mid-holocene europe. *Nature Communications*, 4(1), 2013. doi: 10.1038/ncomms3486.
- 774 B.W Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.
- 775 George Davey Smith, Debbie A Lawlor, Nic J Timpson, Jamil Baban, Matt Kiessling, Ian N Day, and Shah
776 Ebrahim. Lactase persistence-related genetic variant: Population substructure and health outcomes.
777 *European Journal of Human Genetics*, 17(3):357–367, 2008. doi: 10.1038/ejhg.2008.156.
- 778 Leo Speidel, Marie Forest, Sinan Shi, and Simon R. Myers. A method for genome-wide geneal-
779 ogy estimation for thousands of samples. *Nature Genetics*, 51(9):1321–1329, 2019. doi: 10.1038/
780 s41588-019-0484-x.
- 781 Jeffrey P Spence, Tony Zeng, Hakhamanesh Mostafavi, and Jonathan K Pritchard. Scaling the discrete-
782 time Wright–Fisher model to biobank-scale datasets. *Genetics*, 225(3):iyad168, 09 2023. ISSN 1943-
783 2631. doi: 10.1093/genetics/iyad168. URL <https://doi.org/10.1093/genetics/iyad168>.
- 784 Matthias Steinrücken, Anand Bhaskar, and Yun S. Song. A novel spectral method for inferring general
785 diploid selection from time series genetic data. *The Annals of Applied Statistics*, 8(4), 2014. doi:
786 10.1214/14-aos764.
- 787 Aaron J. Stern, Peter R. Wilton, and Rasmus Nielsen. An approximate full-likelihood method for inferring
788 selection and allele frequency trajectories from dna sequence data. *PLOS Genetics*, 15(9):1–32, 09 2019.
789 doi: 10.1371/journal.pgen.1008384. URL <https://doi.org/10.1371/journal.pgen.1008384>.
- 790 Oddný Ósk Sverrisdóttir, Adrian Timpson, Jamie Toombs, Cecile Lecoeur, Philippe Froguel, Jose Miguel
791 Carretero, Juan Luis Arsuaga Ferreras, Anders Götherström, and Mark G. Thomas. Direct estimates
792 of natural selection in iberia indicate calcium absorption was not the only driver of lactase persistence
793 in europe. *Molecular Biology and Evolution*, 31(4):975–983, 2014. doi: 10.1093/molbev/msu049.
- 794 Fumio Tajima. Statistical method for testing the neutral mutation hypothesis by dna polymorphism.
795 *Genetics*, 123(3):585–595, 1989. doi: 10.1093/genetics/123.3.585.
- 796 Seth D Temple, Ryan D Waples, and Sharon R Browning. Modeling recent positive selection in americans
797 of european ancestry. *bioRxiv*, 2023. doi: 10.1101/2023.11.13.566947. URL <https://www.biorxiv.org/content/early/2023/11/16/2023.11.13.566947>.
- 799 Dov Tiosano, Laura Audi, Sharlee Climer, Weixiong Zhang, Alan R Templeton, Monica Fernández-
800 Cancio, Ruth Gershoni-Baruch, José Miguel Sánchez-Muro, Mohamed El Kholy, and Zèev
801 Hochberg. Latitudinal Clines of the Human Vitamin D Receptor and Skin Color Genes. *G3*

- 802 *Genes—Genomes—Genetics*, 6(5):1251–1266, 05 2016. ISSN 2160-1836. doi: 10.1534/g3.115.026773.
803 URL <https://doi.org/10.1534/g3.115.026773>.
- 804 Luis Torada, Lucrezia Lorenzon, Alice Beddis, Ulas Isildak, Linda Pattini, Sara Mathieson, and Matteo
805 Fumagalli. Imagene: A convolutional neural network to quantify natural selection from genomic data.
806 *BMC Bioinformatics*, 20(S9), 2019. doi: 10.1186/s12859-019-2927-x.
- 807 M.A. Vázquez and J. Míguez. Importance sampling with transformed weights. *Electronics Letters*, 53
808 (12):783–785, 2017. doi: 10.1049/el.2016.3462.
- 809 Eric T. Wang, Greg Kodama, Pierre Baldi, and Robert K. Moyzis. Global landscape of recent inferred
810 darwinian selection for homo sapiens. *Proceedings of the National Academy of Sciences*, 103(1):135–
811 140, 2006. doi: 10.1073/pnas.0509691102. URL <https://www.pnas.org/doi/abs/10.1073/pnas.0509691102>.
- 812 Stephan Weidinger, Saffron A.G. Willis-Owen, Yoichiro Kamatani, Hansjörg Baurecht, Nilesh Morar,
813 Liming Liang, Pauline Edser, Teresa Street, Elke Rodriguez, and Grainne M. O'Regan et al. A
814 genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma
815 and psoriasis. *Human Molecular Genetics*, 22(23):4841–4856, 2013. doi: 10.1093/hmg/ddt317.
- 816 Sandra Wilde, Adrian Timpson, Karola Kirsanow, Elke Kaiser, Manfred Kayser, Martina Unterländer,
817 Nina Hollfelder, Inna D. Potekhina, Wolfram Schier, Mark G. Thomas, and Joachim Burger. Direct
818 evidence for positive selection of skin, hair, and eye pigmentation in europeans during the last 5,000
819 y. *Proceedings of the National Academy of Sciences*, 111(13):4832–4837, 2014. doi: 10.1073/pnas.
820 1316513111. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1316513111>.
- 821 S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The
822 Annals of Mathematical Statistics*, 9(1):60–62, 1938. doi: 10.1214/aoms/1177732360.
- 823 Ellen G Williamson and Montgomery Slatkin. Using Maximum Likelihood to Estimate Population Size
824 From Temporal Changes in Allele Frequencies. *Genetics*, 152(2):755–761, 06 1999. ISSN 1943-2631.
825 doi: 10.1093/genetics/152.2.755. URL <https://doi.org/10.1093/genetics/152.2.755>.
- 826 Zimeng Ye, Ping Shuai, Yaru Zhai, Fang Li, Lingxi Jiang, Fang Lu, Feng Wen, Lulin Huang, Dingding
827 Zhang, and Xiaoqi Liu et al. Associations of 6p21.3 region with age-related macular degeneration and
828 polypoidal choroidal vasculopathy. *Scientific Reports*, 6(1), 2016. doi: 10.1038/srep20914.
- 829 Brian C. Zhang, Arjun Biddanda, Árni Freyr Gunnarsson, Fergus Cooper, and Pier Francesco Palamara.
830 Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex
831 traits. *Nature Genetics*, 55(5):768–776, 2023. doi: 10.1038/s41588-023-01379-x.

833 A Appendix

834 Differing population sizes

835 We also allow the specification of different population sizes in different time periods. This is accomplished
836 by replacing N by a time-specific N_t in both the variance of the normal distribution used for the transition
837 probabilities and the equation for the coalescence emissions. We run a set of ancient genotype simulations,
838 identical to those described in section “Ancient Genotypes”, except that the population size in $[0, 400)$
839 was $2N = 60000$, the population size in the epoch $[400, 600)$ was $2N = 200000$ and the population size
840 in the epoch $[600, \infty)$ was $2N = 80000$. Figure 11 shows that we can properly account for changes in
841 population size and generate accurate estimates.

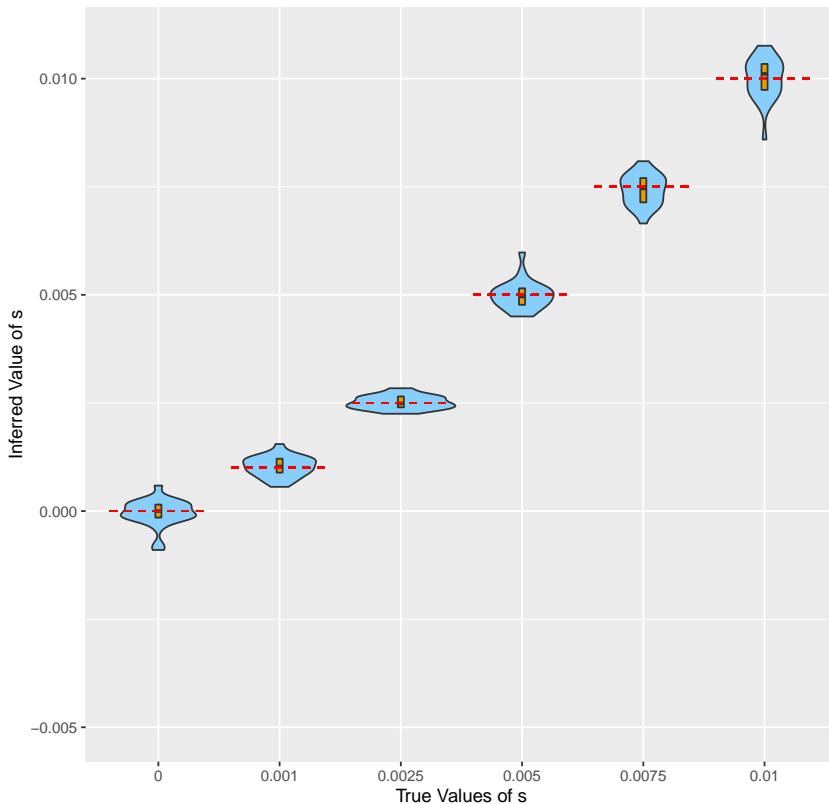


Fig 11: Violin plots showing the results of running CLUES2 on genotype data simulated with changing population sizes through time. Boxplots are overlaid with the whiskers omitted. True values of s are shown as dashed red lines. 30 replicates were performed for each selection coefficient.

842 Simulation Details

843 Ancient Genotypes

844 We simulate ancient genotype data according to a basic forward-in-time Wright-Fisher simulation with
845 rejection sampling. We begin our simulation with derived allele frequency $1/N$, corresponding to a new
846 mutation. Given the frequency at generation k , call it x_k , x_{k+1} is A_{k+1}/N where A_{k+1} is drawn from

847 $\text{Bin}(N, \frac{x_k + sx_k}{1+sx_k})$. We reject simulations where the derived allele is lost and retain simulations where the
848 derived allele is fixed. For a simulation where the allele is fixed, we then find the most recent time step
849 l for which $|x_l - f| < 0.005$, where f is our desired modern allele frequency. We reject the simulation
850 if no such time exists. If such a time exists, then we renumber time l to be the present (0 generations
851 ago) and set the ages of the allele frequencies older than this time to be relative to this timepoint. We
852 now have an allele frequency trajectory that goes from $1/N$ at some point in the past to f at the present
853 generation. Given this trajectory, we draw one individual every m generations in the past (meaning one
854 individual is sampled m generations in the past, one individual $2m$ generations in the past, etc) and their
855 genotype is recorded. The exception to our approach above is the case when $s = 0$, for which we begin
856 at time 0 with frequency f and simply run the Wright-Fisher process in reverse.

857 True Trees

858 For our simulations involving true gene trees under selection, we simulate data using the **SweepGenicSelection**
859 function in *msprime* 1.0, which allows us to quickly and easily simulate trees under genic selection.
860 (Baumdicker et al., 2021) As this feature is only available for strictly positive selection coefficients, our
861 simulation results for $s = 0$ were actually simulated with $s = 10^{-6}$, which makes no practical difference
862 to the distribution of trees sampled. These trees were “recapitated” using the **recapitate** function from
863 the *pyslim* library (Haller et al., 2019). in order to generate a complete, rooted gene tree. Note that the
864 parameterization of selection in the **SweepGenicSelection** function differs from our parameterization,
865 so we call this function using a selection coefficient of $2s$, where s is the selection coefficient under which
866 we wish to simulate.

867 Importance Sampling

868 For our importance sampling simulation study, we use *msprime* to simulate a gene tree under selection
869 as described in the previous section. We then simulate neutral mutations on this tree, using a continuous
870 genome model so that the infinite sites assumption is always satisfied. We also record the true topology
871 of the simulated tree.

872 We here describe our simple MCMC algorithm. We do not vary the topology of the tree but instead
873 use the true simulated topology. We consider this acceptable for two reasons. The first is that CLUES2
874 depends only on the coalescence times between allelic classes, resulting in the exchangeability of lineages
875 within allelic classes. Secondly, the lack of recombination and our relatively small sample size (we simulate
876 only 24 leaves) means that there will likely be at least 1 mutation per branch, resulting in only 1 possible
877 topology that is compatible with the infinite sites assumption. Then, given a fixed topology on L leaves,
878 a tree is defined by two sets of parameters:

879 (1) The vector $(t_L \dots t_2)$ of intercoalescence times, where t_k denotes the time when there are k lineages

880 remaining

881 (2) The permutation vector (p_1, \dots, p_{L-1}) , a subset of the set of permutations on the numbers 1 to
882 $L - 1$, which describes the relative ordering of the coalescence events. Note that not all permutations are
883 acceptable, as many of them violate the topological constraints of the tree. For example, a coalescence
884 node cannot occur at an older time than its parent.

We initialize the coalescence times by sampling from the prior distribution on intercoalescence times (see below). We initialize the ordering of the coalescence events in the tree using the `topo_sort` function from the *igraph* package, which generates a topological sorting of the vertices of a directed acyclic graph (Csardi and Nepusz, 2005). Our target function is:

$$\prod_{k=2}^L \frac{\binom{k}{2}}{N} \exp\left(-\frac{\binom{k}{2} t_k}{N}\right) \prod_b \frac{(\tau_b \mu L)^{m_b} e^{-\tau_b \mu L}}{m_b!}$$

885 where we use b to index the branches of the tree, μ is the mutation rate, L is the length of the genomic
886 region, m_b is the number of mutations on branch b , and τ_b is the length of branch b . We use a Metropolis-
887 Hastings algorithm that is inspired by that used by Relate to sample branch lengths (see Supplementary
888 Section 4.2 of (Speidel et al., 2019)) but differs in some minor details. With probability 1/2, we propose
889 a change to the coalescence time vector. We pick a coalescence time t_k uniformly at random and sample
890 a new time from the exponential distribution with mean t_k . With probability 1/2, we choose a pair
891 of adjacent indices in the permutation ordering and propose a switch between them. We automatically
892 reject proposals that violate the topological restraints of the tree. Acceptance probabilities are calculated
893 using the standard Metropolis-Hastings algorithm with our target function. This algorithm, while not
894 efficient, is able to obtain samples from the distribution of branch lengths in a tree with a fixed topology
895 and no recombination events without suffering from the observed problems with existing ARG-inference
896 methods.

897 Gene Trees on Ancient Data

898 Simulating and recording full gene trees on ancient data conditional on a specified selection coefficient and
899 a specified modern allele frequency is a challenging computational problem. However, we take advantage
900 of the fact that the CLUES2 model implies exchangeability of lineages within each allele class to greatly
901 simplify the simulation process. We begin by starting at the present with our specified modern allele
902 frequency, f . Then, given our allele frequency at time t before the present, call it x_t , we draw the derived
903 allele frequency at generation $t + 1$ before the present as a variate from a normal distribution with mean
904 $x_t - sx_t(1 - x_t)$ and variance $x_t(1 - x_t)/(2N)$. We continue this process until we reach T , our pre-specified
905 time cutoff. Then, we convert the frequency at each generation into an integer number of individuals
906 by multiplying by N and rounding to the nearest integer. We then initialize our fS derived lineages

and our $(1 - f)S$ ancestral lineages where S denotes the sample size of modern lineages. Going back in time, we then draw the parent of each derived lineage from the possible derived individuals in the previous generation. We do the same for the ancestral lineages. If 2 lineages have the same parent, then a coalescence is recorded for the appropriate allelic class. We continue this process until time T . In this way, we obtain a set of derived and ancestral coalescence times, as CLUES2 requires. It is straightforward to incorporate ancient data into this simulation procedure by specifying sampling times τ_1, \dots, τ_n , not necessarily unique, and adding an additional derived lineage at time τ_i with probability x_{τ_i} and adding an additional ancestral lineage otherwise. In this way, we can simulate the ancient gene trees input to CLUES without having to keep track of the entire tree topology. For the simulations described in the main text, we have 40 modern haplotypes. For the ancient gene tree simulations, we add 160 leaves to the tree at each of the times 100 and 200. As a comparison, we run identical simulations where we instead sample 80 diploid genotypes at each of the times 100 and 200.

919 **B Supplementary Figures**

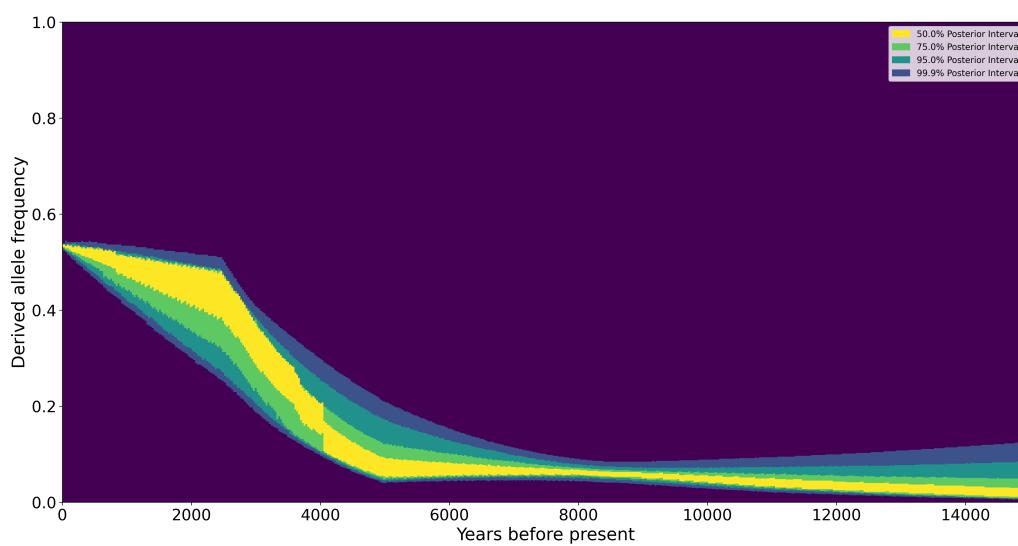


Fig S1: The reconstructed derived allele frequency trajectory for SNP rs4988235 in the MCM6 gene under the 3-coefficient model.

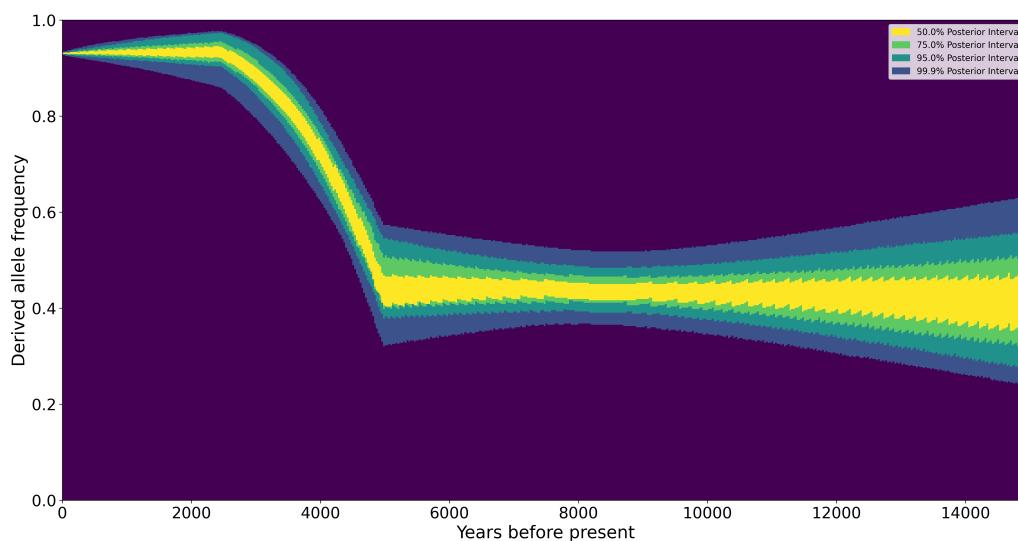


Fig S2: The reconstructed derived allele frequency trajectory for SNP rs35395 in the SLC45A2 gene under the 3-coefficient model.

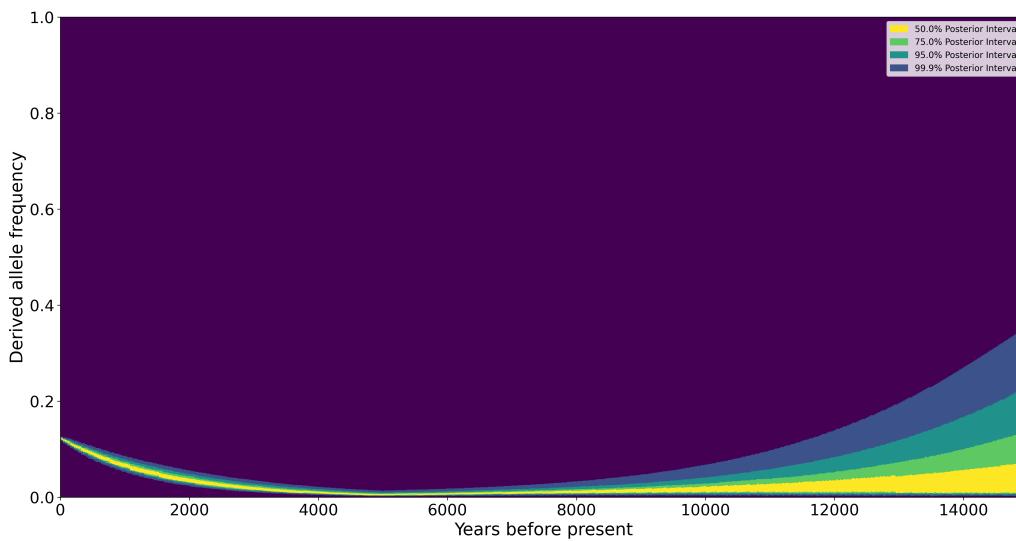


Fig S3: The reconstructed derived allele frequency trajectory for SNP rs12153855 in the TNXB gene under the 2-coefficient model.

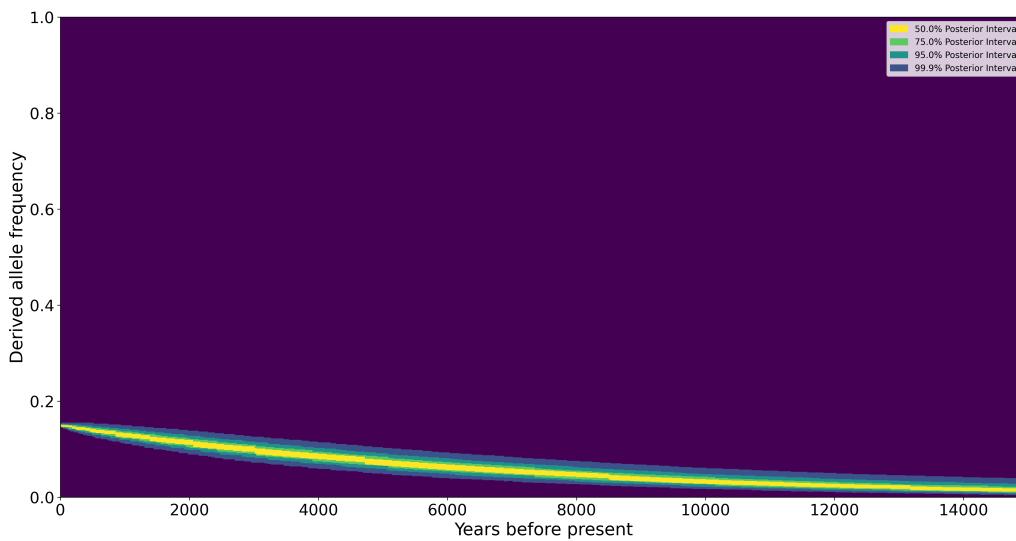


Fig S4: The reconstructed derived allele frequency trajectory for SNP rs75393320 in the ACP2 gene under the 1-coefficient model.

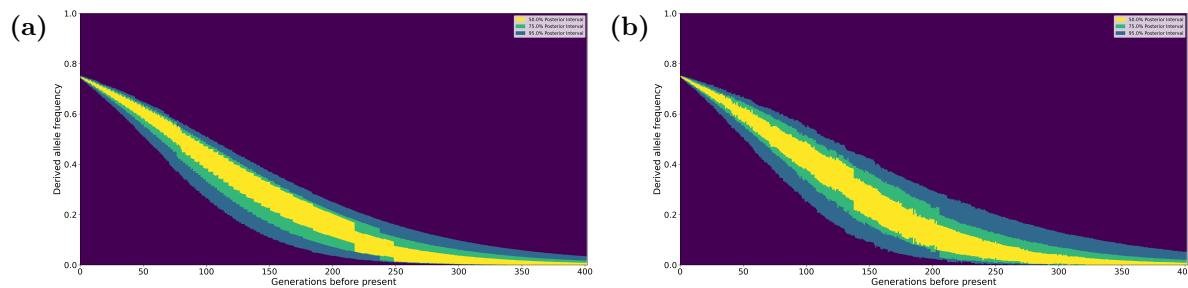


Fig S5: The results of the allele frequency trajectory reconstruction of CLUES2 with (a) the Monte Carlo integration framework and with (b) the exact rejection sampling approach for a 1-selection coefficient model.

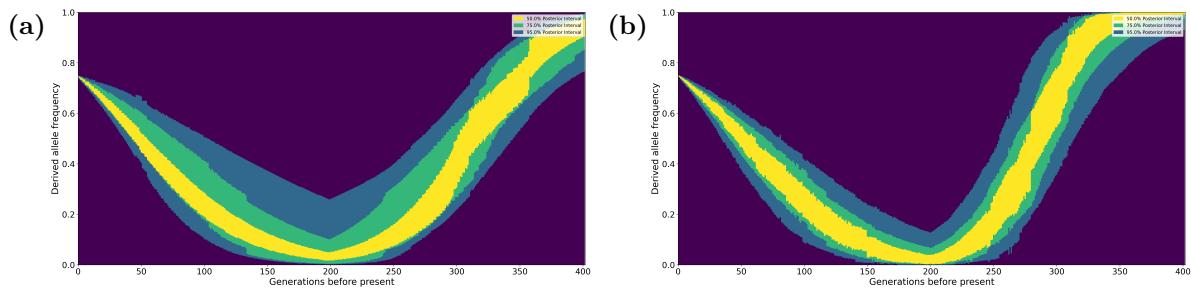


Fig S6: The results of the allele frequency trajectory reconstruction of CLUES2 with (a) the Monte Carlo integration framework and with (b) the exact rejection sampling approach for a 2-selection coefficient model.

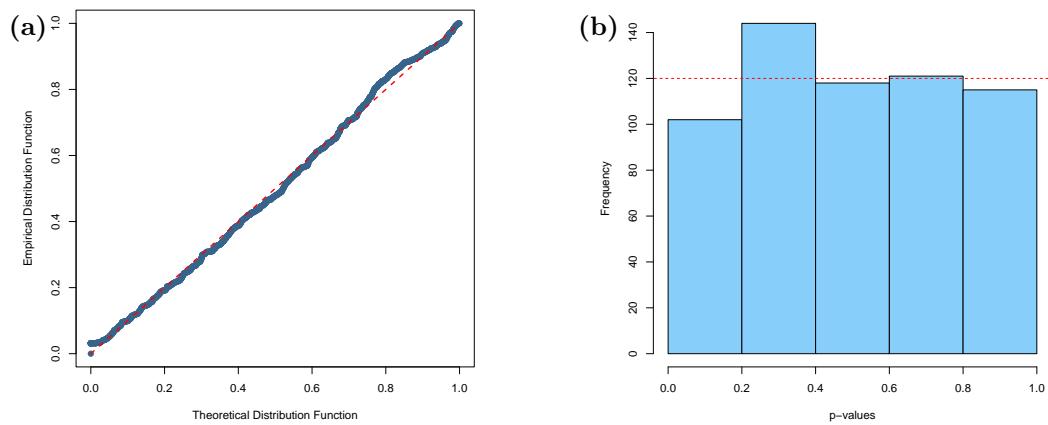


Fig S7: The (a) P-P plot and (b) p-value histogram for our ancient genotype simulations.

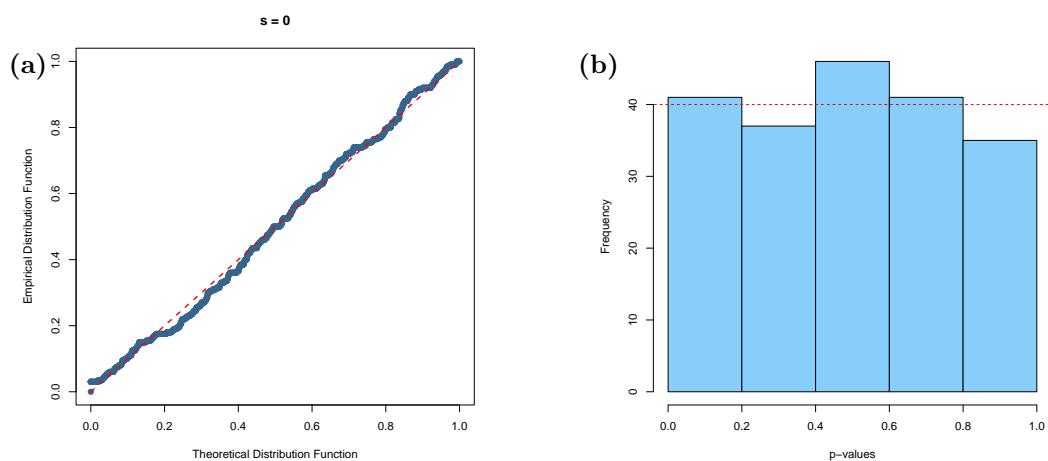


Fig S8: The (a) P-P plot and (b) p-value histogram for our true tree simulations.

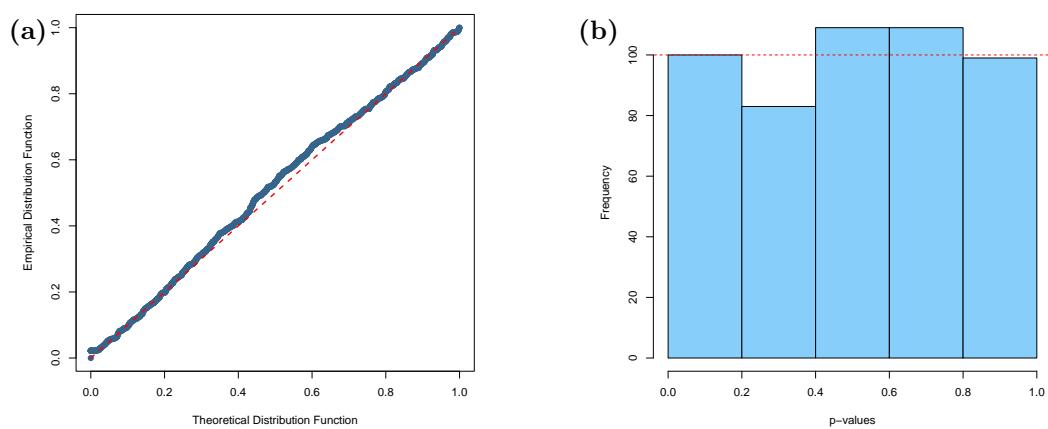


Fig S9: The (a) P-P plot and (b) p-value histogram for our importance sampling simulations.

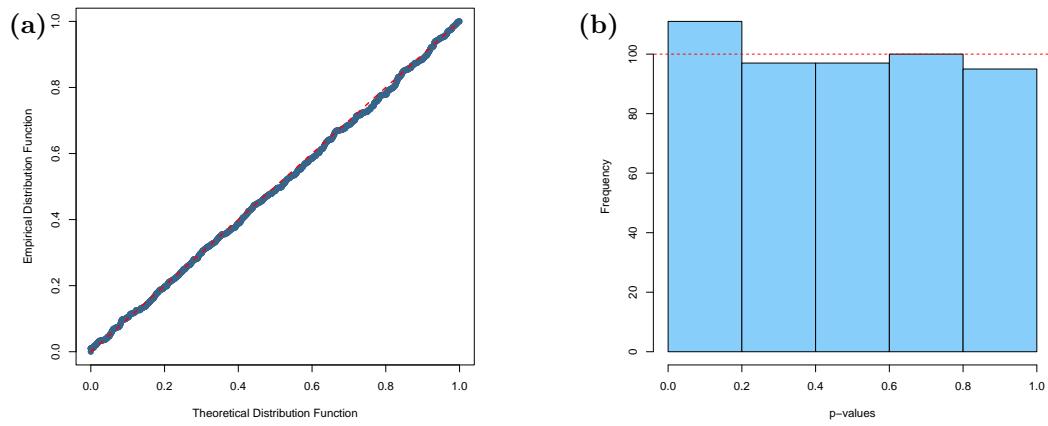


Fig S10: The (a) P-P plot and (b) p-value histogram for our simulations of multiple selection coefficients.

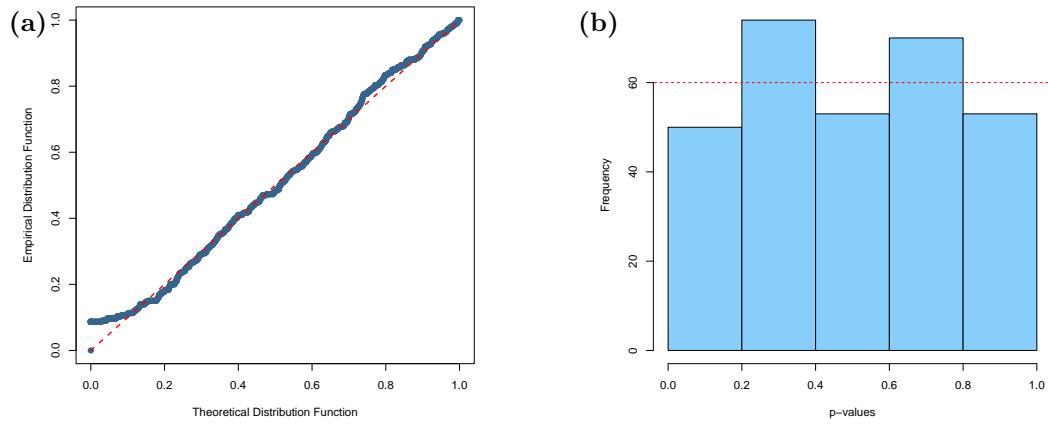


Fig S11: The (a) P-P plot and (b) p-value histogram for our simulations with varying population sizes.

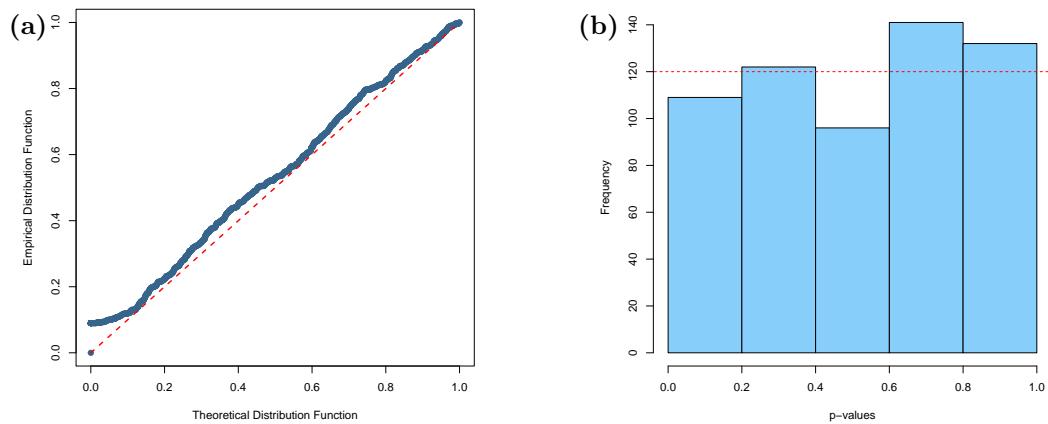


Fig S12: The (a) P-P plot and (b) p-value histogram for our simulations with ancient gene trees.

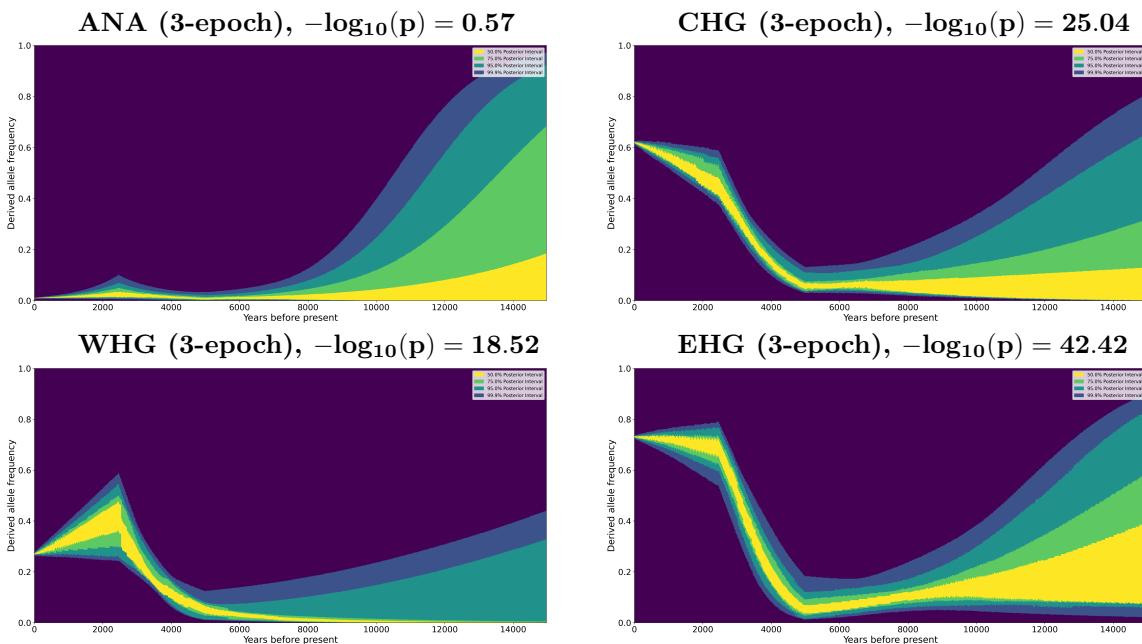


Fig S13: Ancestry-stratified allele frequency trajectories for the MCM6 SNP rs4988235, (chr2:136608646). Ancestries analyzed are Anatolian farmer (ANA), Caucasus hunter-gatherer (CHG), Western hunter-gatherer (WHG) and Eastern hunter gatherer (EHG).

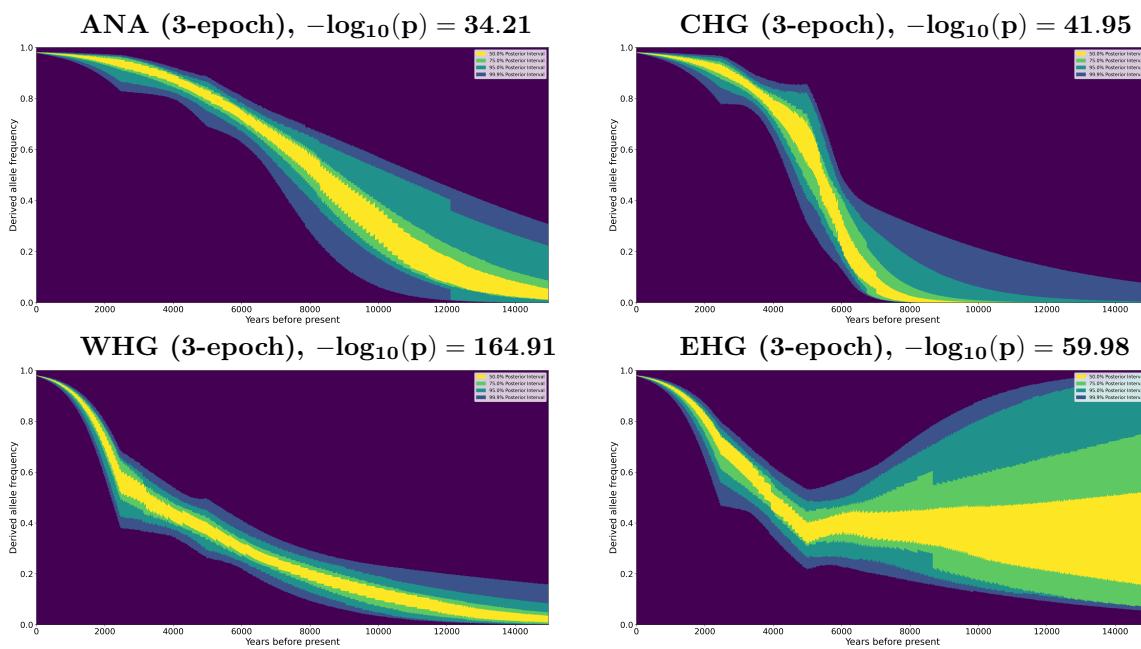


Fig S14: Ancestry-stratified allele frequency trajectories for the SLC45A2 SNP rs35395, (chr5:33948589). Ancestries analyzed are Anatolian farmer (ANA), Caucasus hunter-gatherer (CHG), Western hunter-gatherer (WHG) and Eastern hunter gatherer (EHG).

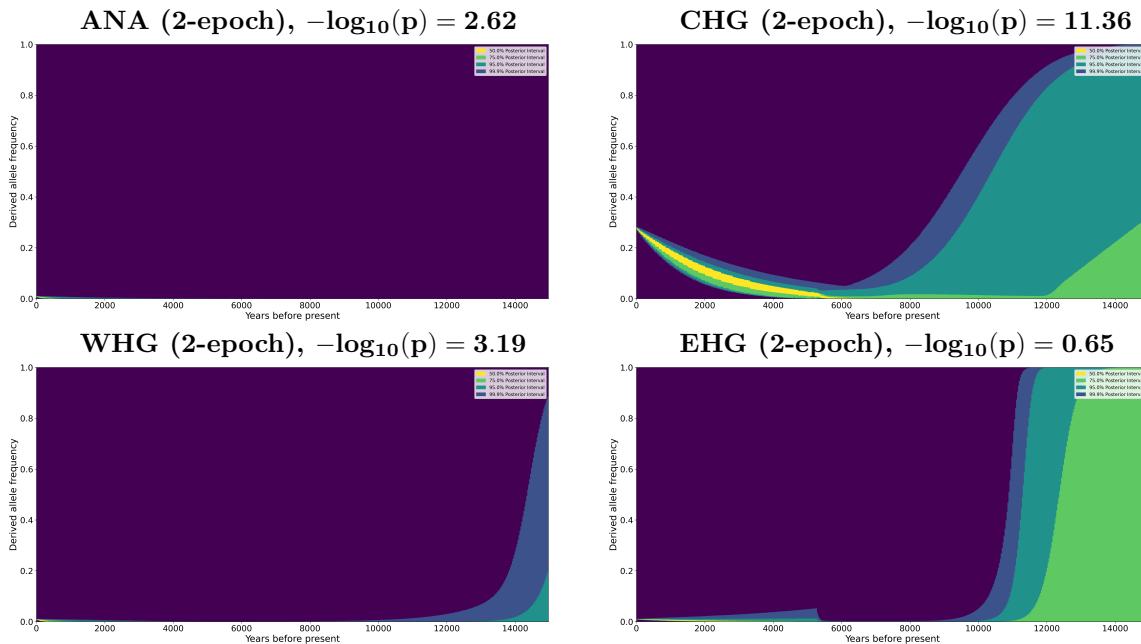


Fig S15: Ancestry-stratified allele frequency trajectories for the TNXB SNP rs12153855, (chr6:32074804). Ancestries analyzed are Anatolian farmer (ANA), Caucasus hunter-gatherer (CHG), Western hunter-gatherer (WHG) and Eastern hunter gatherer (EHG).

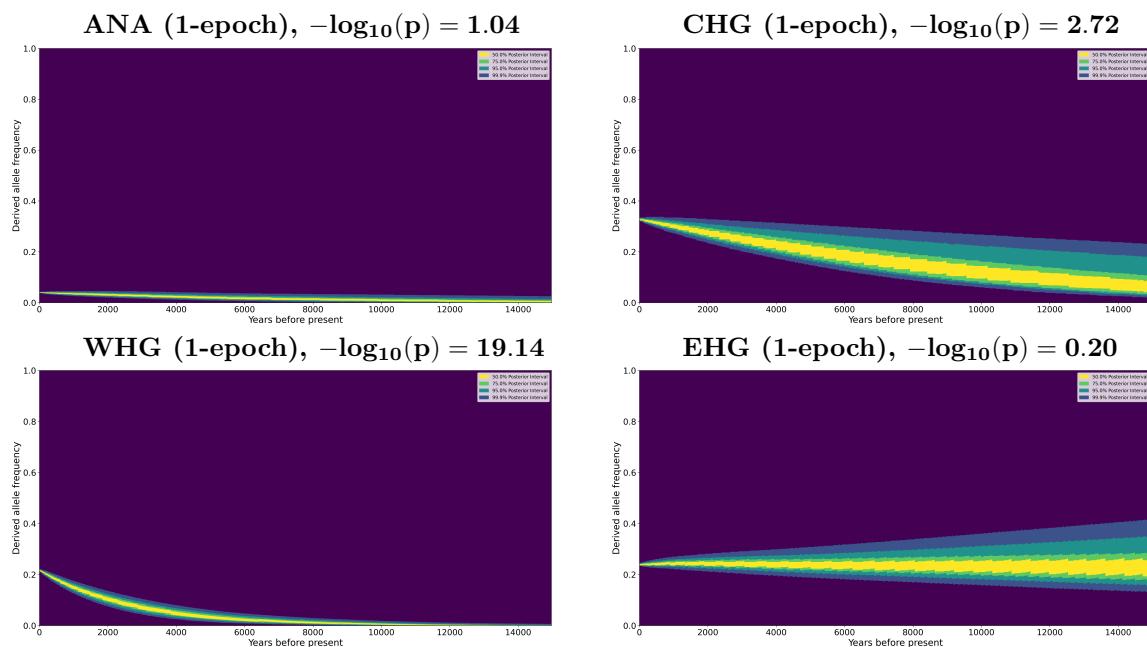


Fig S16: Ancestry-stratified allele frequency trajectories for the ACP2 SNP rs75393320, (chr11:47266471). Ancestries analyzed are Anatolian farmer (ANA), Caucasus hunter-gatherer (CHG), Western hunter-gatherer (WHD) and Eastern hunter gatherer (EHG).