Estimating temporally variable selection intensity from ancient DNA data with the flexibility of modelling linkage and epistasis

Zhangyi He[a,1,*], Xiaoyang Dai[b,1], Wenyang Lyu[c], Mark Beaumont[d], Feng Yu[c]

[a] *Cancer Research UK Beatson Institute, Glasgow G61 1BD, United Kingdom*
[b] *The Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, United Kingdom*
[c] *School of Mathematics, University of Bristol, Bristol BS8 1UG, United Kingdom*
[d] *School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom*

Abstract

Innovations in ancient DNA (aDNA) preparation and sequencing technologies have exponentially increased the quality and quantity of aDNA data extracted from ancient biological materials. The additional temporal component from the incoming aDNA data can provide improved power to address fundamental evolutionary questions like characterising selection processes that shape the phenotypes and genotypes of contemporary populations or species. However, utilising aDNA to study past selection processes still involves considerable hurdles like how to eliminate the confounding factor of genetic interactions in the inference of selection. To address this issue, we extend the approach of He et al. (2022) to infer temporally variable selection from the aDNA data in the form of genotype likelihoods with the flexibility of modelling linkage and epistasis in this work. Our posterior computation is carried out by a robust adaptive version of the particle marginal Metropolis-Hastings algorithm with a coerced acceptance rate. Our extension inherits the desirable features of He et al. (2022) such as modelling sample uncertainty resulting from the damage and fragmentation of aDNA molecules and reconstructing underlying gamete frequency trajectories of the population. We evaluate its performance through extensive simulations and show its utility with an application to the aDNA data from pigmentation loci in horses.

Keywords: Ancient DNA, Natural selection, Genetic linkage, Epistatic interaction, Two-layer hidden Markov model, Adaptive particle marginal Metropolis-Hastings

*Corresponding author.
Email address: z.he@beatson.gla.ac.uk (Zhangyi He)
[1] These authors contributed equally to this work.

1. Introduction

Natural selection is one of the primary mechanisms of evolutionary changes and is responsible for the evolution of adaptive features (Darwin, 1859). A full understanding of the role of selection in driving evolutionary changes needs accurate estimates of the underlying timing and strength of selection. With recent advances in sequencing technologies and molecular techniques tailored to ultra-damaged templates, high-quality time serial samples of segregating alleles have become increasingly common in ancestral populations, (*e.g.*, Mathieson et al., 2015; Loog et al., 2017; Fages et al., 2019; Alves et al., 2019). The additional temporal dimension of the ancient DNA (aDNA) data has the promise of boosting power of estimating population genetic parameters, in particular for the pace of adaptation, as the allele frequency trajectory through time itself gives us valuable information collected before, during and after genetic changes driven by selection. See Dehasque et al. (2020) for a review of the inference of selection from aDNA.

The temporal component provided by the incoming aDNA data spurred the development of statistical approaches for the inference of selection from time series data of allele frequencies in the last 15 years (see Malaspinas, 2016, for a review). Most existing methods are built upon the hidden Markov model (HMM) framework of Williamson & Slatkin (1999), where the population allele frequency is modelled as a hidden state evolving under the Wright-Fisher model (Fisher, 1922; Wright, 1931), and the sample allele frequency drawn from the underlying population at each given time point is modelled as a noisy observation of the population allele frequency (see Tataru et al., 2017, for a review of statistical inference in the Wright-Fisher model based on time series data of allele frequencies). However, such an HMM framework can be computationally infeasible for large population sizes and evolutionary timescales owing to a prohibitively large amount of computation and storage required in its likelihood calculations.

To our knowledge, most existing methods tailored to aDNA are based on the diffusion limit of the Wright-Fisher model. By working with the diffusion approximation, their HMM framework permits efficient integration over the probability distribution of the underlying population allele frequencies, and therefore, the calculation of the likelihood based on the observed sample allele frequencies can be completed within a reasonable amount of time (*e.g.*, Bollback et al., 2008; Malaspinas et al., 2012; Steinrücken et al., 2014; Schraiber et al., 2016; Ferrer-Admetlla et al., 2016; He et al., 2020b,c, 2022; Lyu et al., 2022). These methods have been successfully applied

in aDNA studies, *e.g.*, Ludwig et al. (2009) used the method of Bollback et al. (2008) to analyse the aDNA data from pigmentation loci in horses and found that positive selection acted on the derived *ASIP* and *MC1R* alleles, showing that domestication and selective breeding contributed to changes in horse coat colouration.

Despite the availability of a certain number of statistical methods for the inference of selection from genetic time series, their application to aDNA data from natural populations remains limited. Most existing methods are developed in the absence of genetic interactions like linkage and epistasis, with the exception of He et al. (2020b). Linkage and recombination are explicitly modelled in He et al. (2020b), which has been illustrated to significantly improve the inference of selection, in particular for tightly linked loci. Ignoring epistasis can also cause severe issues in the study of selection since the combined effects of mutant alleles may be impossible to predict according to the measured individual effects of a given mutant allele (Bank et al., 2014). As an example, the base coat colours of horses are determined by *ASIP* and *MC1R*. The derived *ASIP* and *MC1R* alleles have been shown to be positively selected through existing approaches (*e.g.*, Bollback et al., 2008; Malaspinas et al., 2012; Steinrücken et al., 2014; Schraiber et al., 2016; He et al., 2020c). However, this is not sufficient to conclude that black horses are favoured by selection since the alleles at *MC1R* interact epistatically with those at *ASIP*, *i.e.*, the presence of at least one copy of the dominant ancestral allele at *MC1R*, and the resulting production of black pigment, is required to check the action of the alleles at *ASIP* (Corbin et al., 2020).

To circumvent this issue, in this work we introduce a novel Bayesian method for the inference of selection acting on the phenotypic trait, allowing the intensity to vary over time, from data on aDNA sequences, with the flexibility of modelling genetic linkage and epistatic interaction. Our method is built upon the two-layer HMM framework of He et al. (2022), and our key innovation is to introduce a Wright-Fisher diffusion that can model the dynamics of two linked genes under phenotypic selection over time to be the underlying Markov process, which permits linkage and epistasis. To remain computationally feasible, our posterior computation is carried out with the particle marginal Metropolis-Hastings (PMMH) algorithm developed by Andrieu et al. (2010), where we adopt the adaptation strategy introduced by Vihola (2012) to tune the covariance structure of the proposal to achieve a given acceptance rate. Also, our approach inherits certain desirable features from He et al. (2022) such as modelling sample uncertainty due to the damage

and fragmentation of aDNA molecules and reconstructing underlying frequency trajectories of the gametes in the population. We use our method to analyse the aDNA data associated with horse base coat colours (determined by *ASIP* and *MC1R* with epistatic interaction) and horse pinto coat patterns (controlled by *KIT13* and *KIT16* with tight linkage) from Wutke et al. (2016) to show it applicability, and its performance is evaluated with various simulated data.

2. Materials and Methods

In this section, we construct a Wright-Fisher model to characterise two linked genes evolving under phenotypic selection over time first and then derive its diffusion limit. Working with the diffusion approximation, we extend the approach of He et al. (2022) to infer temporally variable selection from the data on aDNA sequences while modelling linkage and epistasis.

2.1. Wright-Fisher diffusion

We consider a population of randomly mating diploid individuals represented by alleles at loci $\mathcal{A}$ and $\mathcal{B}$ evolving under selection with discrete non-overlapping generations. At each locus, there are two possible allele types, labelled $\mathcal{A}_0$, $\mathcal{A}_1$ and $\mathcal{B}_0$, $\mathcal{B}_1$, respectively, resulting in four possible haplotypes on both loci, $\mathcal{A}_0\mathcal{B}_0$, $\mathcal{A}_0\mathcal{B}_1$, $\mathcal{A}_1\mathcal{B}_0$ and $\mathcal{A}_1\mathcal{B}_1$, labelled haplotypes 00, 01, 10 and 11, respectively. We attach the symbols $\mathcal{A}_0$ and $\mathcal{B}_0$ to the ancestral alleles, which we assume originally exist in the population, and we attach the symbols $\mathcal{A}_1$ and $\mathcal{B}_1$ to the mutant alleles, which we assume arise only once in the population. Considering the absence of sex effects, this setup gives rise to 10 possible (unordered) genotypes $\mathcal{A}_i\mathcal{B}_j/\mathcal{A}_{i'}\mathcal{B}_{j'}$, corresponding to at most 10 distinct phenotypes $\mathcal{P}_{ij,i'j'}$. Phenotypes $\mathcal{P}_{ij,i'j'}$ and $\mathcal{P}_{i'j',ij}$ are identical in our notation.

We incorporate viability selection into the population dynamics and assume that the viability is only determined by the phenotype. Viabilities of all phenotypes at loci $\mathcal{A}$ and $\mathcal{B}$ per generation are assigned $1 + s_{ij,i'j'}$, where $s_{ij,i'j'}$ is the selection coefficient of the $\mathcal{P}_{ij,i'j'}$ phenotype with $s_{ij,i'j'} \in [-1, +\infty)$ and $s_{ij,i'j'} = s_{i'j',ij}$. In what follows, we let the selection coefficient $s_{00,00} = 0$ unless otherwise noted, and then $s_{ij,i'j'}$ denotes the selection coefficient of the $\mathcal{P}_{ij,i'j'}$ phenotype against the $\mathcal{P}_{00,00}$ phenotype.

2.1.1. Wright-Fisher model

Let $X_{ij}^{(N)}(k)$ denote the gamete frequency of haplotype ij at generation $k \in \mathbb{N}$ and $\boldsymbol{X}^{(N)}(k)$ be the vector of the four gamete frequencies. To incorporate non-constant demographic histories,

we assume that the population size changes deterministically, with $N(k)$ denoting the number of diploid individuals in the population at generation k. In the Wright-Fisher model, we assume that gametes are randomly chosen from an effectively infinite gamete pool reflecting the parental gamete frequencies at each generation. We therefore have

$$\boldsymbol{X}^{(N)}(k+1) \mid \boldsymbol{X}^{(N)}(k) = \boldsymbol{x} \sim \frac{1}{2N(k)} \text{Multinomial}(2N(k), \boldsymbol{p}), \tag{1}$$

where $\boldsymbol{p}$ is the vector of parental gamete frequencies. Under the assumption of random mating, we can further express the vector of parental gamete frequencies as

$$p_{ij} = (1-r)x'_{ij} + r(x'_{i0} + x'_{i1})(x'_{0j} + x'_{1j}) \tag{2}$$

for $i, j \in \{0, 1\}$, where

$$x'_{ij} = \frac{\sum_{i',j'=0}^{1}(1+s_{ij,i'j'})x_{i'j'}x_{ij}}{\sum_{i,j=0}^{1}\sum_{i',j'=0}^{1}(1+s_{ij,i'j'})x_{i'j'}x_{ij}},$$

and r denotes the recombination rate of the $\mathcal{A}$ and $\mathcal{B}$ loci located on the same chromosome, *i.e.*, the fraction of recombinant offspring showing a crossover between the two loci per generation. If the $\mathcal{A}$ and $\mathcal{B}$ loci are located on separate chromosomes, we let the (artificial) recombination rate $r = 0.5$ (*i.e.*, free recombination). The two-locus Wright-Fisher model with selection is defined as the Markov process $\boldsymbol{X}^{(N)}$ evolving with transition probabilities in Eq. (1) in the state space $\Omega_{\boldsymbol{X}^{(N)}} = \{\boldsymbol{x} \in \{0, 1/(2N), \ldots, 1\}^4 : \sum_{i,j=0}^{1} x_{ij} = 1\}$.

2.1.2. Diffusion approximation

We study the two-locus Wright-Fisher model with selection through its diffusion limit due to the complicated nature of its transition probability matrix, in particular for large population sizes or evolutionary timescales. More specifically, we measure time in a unit of $2N_0$ generations, denoted by t, where N_0 is an arbitrary reference population size fixed through time, and assume that the selection coefficients and recombination rate are all of order $1/(2N_0)$. As the reference population size N_0 approaches infinity, the scaled selection coefficients $\alpha_{ij,i'j'} = 2N_0 s_{ij,i'j'}$ and the scaled recombination rate $\rho = 4N_0 r$ are kept constant, and the ratio of the population size to the reference population size $N(t)/N_0$ converges to a function, denoted by $\beta(t)$. Notice that the assumption will be violated if the $\mathcal{A}$ and $\mathcal{B}$ loci are located on separate chromosomes, *i.e.*,

$r = 0.5$, but we shall nevertheless use this scaling to find the drift term in the diffusion limit. We will plug the unscaled recombination rate r into the resulting system of stochastic differential equations (SDE's) and use that as our diffusion approximation.

Let $\Delta X_{ij}^{(N)}(k)$ denote the change in the gamete frequency of haplotype ij over generation k. With standard techniques of diffusion theory (see, *e.g.*, Karlin & Taylor, 1981), we can formulate the infinitesimal mean vector $\boldsymbol{\mu}(t, \boldsymbol{x})$ and the infinitesimal (co)variance matrix $\boldsymbol{\Sigma}(t, \boldsymbol{x})$ as

$$\begin{aligned}
\mu_{ij}(t, \boldsymbol{x}) &= \lim_{N_0 \to \infty} 2N_0 \, \mathrm{E}[\Delta X_{ij}^{(N)}([2N_0 t]) \mid \boldsymbol{X}^{(N)}([2N_0 t]) = \boldsymbol{x}] \\
&= \lim_{N_0 \to \infty} 2N_0 (p_{ij} - x_{ij}) \\
\Sigma_{ij,i'j'}(t, \boldsymbol{x}) &= \lim_{N_0 \to \infty} 2N_0 \, \mathrm{E}[\Delta X_{ij}^{(N)}([2N_0 t]) \Delta X_{i'j'}^{(N)}([2N_0 t]) \mid \boldsymbol{X}^{(N)}([2N_0 t]) = \boldsymbol{x}] \\
&= \lim_{N_0 \to \infty} \frac{2N_0}{2N([2N_0 t])} p_{ij}(\delta_{ii'}\delta_{jj'} - p_{i'j'}) + 2N_0 (p_{ij} - x_{ij})(p_{i'j'} - x_{i'j'})
\end{aligned}$$

for $i, j, i', j' \in \{0, 1\}$, where δ denotes the Kronecker delta function and $[\,\cdot\,]$ is used to represent the integer part of the value in the brackets.

To obtain the expression for the infinitesimal mean vector $\boldsymbol{\mu}(t, \boldsymbol{x})$, we compute the limit of the expected change in the gamete frequency of haplotype ij within a single generation as the reference population size N_0 goes to infinity. The only terms that survive after taking the limit are the first order terms in the Taylor expansion of the sampling probability p_{ij} in Eq. (2) with respect to the selection coefficients $s_{ij,i'j'}$ and recombination rate r. We have

$$\mu_{ij}(t, \boldsymbol{x}) = x_{ij} \sum_{i',j'=0}^{1} \alpha_{ij,i'j'} x_{i'j'} - x_{ij} \sum_{i',j'=0}^{1} \sum_{i,j=0}^{1} x_{ij} \alpha_{ij,i'j'} x_{i'j'} - (-1)^{\delta_{ij}} \frac{\rho}{2} (x_{00}x_{11} - x_{01}x_{10}) \quad (3)$$

for $i, j \in \{0, 1\}$. Note that we take the scaled recombination rate to be $\rho = 2N_0$ (*i.e.*, the (artificial) recombination rate $r = 0.5$) if the $\mathcal{A}$ and $\mathcal{B}$ loci are located on separate chromosomes. Such a strong recombination term serves to uncouple the two genes located on separate chromosomes. The infinitesimal (co)variance matrix $\boldsymbol{\Sigma}(t, \boldsymbol{x})$ corresponds to the standard Wright-Fisher diffusion on four haplotypes (see, *e.g.*, He et al., 2020a). We have

$$\Sigma_{ij,i'j'}(t, \boldsymbol{x}) = \frac{x_{ij}(\delta_{ii'}\delta_{jj'} - x_{i'j'})}{\beta(t)} \quad (4)$$

for $i, j, i', j' \in \{0, 1\}$.

Combining the Wright-Fisher diffusion with the infinitesimal mean vector $\boldsymbol{\mu}(t, \boldsymbol{x})$ in Eq. (3) and the infinitesimal (co)variance matrix $\boldsymbol{\Sigma}(t, \boldsymbol{x})$ in Eq. (4), we achieve the following system of SDE's as our diffusion approximation of the Wright-Fisher model in Eq. (1)

$$dX_{ij}(t) = \mu_{ij}(t, \boldsymbol{X}(t))dt + \sum_{i',j'=0}^{1} \sqrt{\frac{X_{ij}(t)X_{i'j'}(t)}{\beta(t)}}\, dW_{ij,i'j'}(t) \tag{5}$$

for $i, j \in \{0, 1\}$, where $W_{ij,i'j'}$ denotes an independent standard Wiener process with $W_{ij,i'j'}(t) = -W_{i'j',ij}(t)$. This anti-symmetry requirement implies $W_{ij,ij}(t) = 0$, and the (co)variance matrix for the X_{ij}'s is exactly the infinitesimal (co)variance matrix $\boldsymbol{\Sigma}(t, \boldsymbol{x})$ in Eq. (4). We refer to the diffusion process $\boldsymbol{X}$ evolving in the state space $\Omega_{\boldsymbol{X}} = \{\boldsymbol{x} \in [0, 1]^4 : \sum_{i,j=0}^{1} x_{ij} = 1\}$ that solves the system of SDE's in Eq. (5) as the two-locus Wright-Fisher diffusion with selection.

2.2. Bayesian inference of selection

Suppose that the available data are always sampled from the underlying population at a finite number of distinct time points, say $t_1 < t_2 < \ldots < t_K$, measured in units of $2N_0$ generations. We assume that N_k individuals are drawn from the underlying population at the k-th sampling time point, and for individual n, let $\boldsymbol{r}_{l,n,k}$ be, in this generic notation, all of the reads at locus l for $l \in \{1, 2\}$. The population genetic quantities of our interest are the selection coefficients $s_{ij,i'j'}$ for $i, j, i', j' \in \{0, 1\}$. Recall that our setup gives rise to at most 10 distinct phenotypes (*i.e.*, at most 9 distinct selection coefficients). For simplicity, we use $\boldsymbol{\vartheta}$ to represent all distinct selection coefficients to estimate.

2.2.1. Hidden Markov model

We extend the two-layer HMM framework of He et al. (2022) to model linkage and epistasis, where the first hidden layer $\boldsymbol{X}(t)$ characterises the gamete frequency trajectories of the underlying population through time by the Wright-Fisher diffusion in Eq. (5), the second hidden layer $\boldsymbol{G}(t)$ represents the genotype of the individual in the sample, and the third observed layer $\boldsymbol{R}(t)$ denotes the data on ancient DNA sequences (see Figure 1).

We let $\boldsymbol{x}_{1:K} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_K\}$ be the frequency trajectories of the gametes in the underlying population at the sampling time points $\boldsymbol{t}_{1:K}$ and $\boldsymbol{g}_{1:K} = \{\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_K\}$ be the genotypes of the individuals drawn from the underlying population at the sampling time points $\boldsymbol{t}_{1:K}$, where

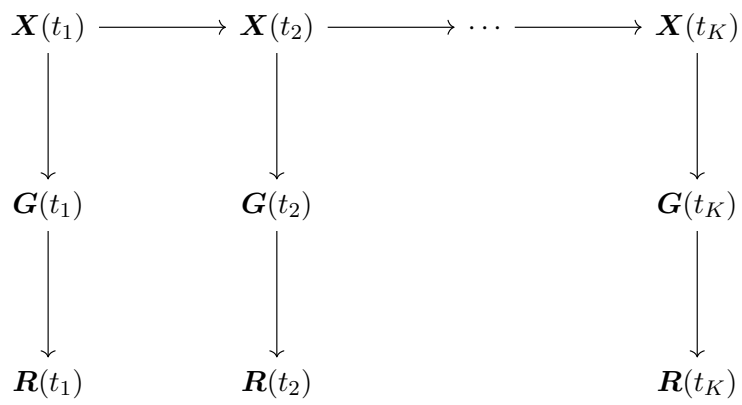

Figure 1: Graphical representation of the two-layer HMM framework extended from He et al. (2022) for the data on ancient DNA sequences.

$\boldsymbol{g}_k = \{\boldsymbol{g}_{1,k}, \boldsymbol{g}_{2,k}, \ldots, \boldsymbol{g}_{N_k,k}\}$ with $\boldsymbol{g}_{n,k} = \{g_{1,n,k}, g_{2,n,k}\}$ and $g_{l,n,k} \in \{0, 1, 2\}$ being the number of mutant alleles at locus l in individual n at sampling time point t_k. Based on the HMM framework illustrated in Figure 1, the posterior probability distribution for the selection coefficients and population gamete frequency trajectories can be expressed as

$$p(\boldsymbol{\vartheta}, \boldsymbol{x}_{1:K} \mid \boldsymbol{r}_{1:K}) = \sum_{\boldsymbol{g}_{1:K}} p(\boldsymbol{\vartheta}, \boldsymbol{x}_{1:K}, \boldsymbol{g}_{1:K} \mid \boldsymbol{r}_{1:K}),$$

where

$$p(\boldsymbol{\vartheta}, \boldsymbol{x}_{1:K}, \boldsymbol{g}_{1:K} \mid \boldsymbol{r}_{1:K}) \propto p(\boldsymbol{\vartheta}) p(\boldsymbol{x}_{1:K} \mid \boldsymbol{\vartheta}) p(\boldsymbol{g}_{1:K} \mid \boldsymbol{x}_{1:K}) p(\boldsymbol{r}_{1:K} \mid \boldsymbol{g}_{1:K}) \tag{6}$$

and $\boldsymbol{r}_{1:K} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_K\}$ with $\boldsymbol{r}_k = \{\boldsymbol{r}_{1,k}, \boldsymbol{r}_{2,k}, \ldots, \boldsymbol{r}_{N_k,k}\}$ and $\boldsymbol{r}_{n,k} = \{\boldsymbol{r}_{1,n,k}, \boldsymbol{r}_{2,n,k}\}$.

The first term of the product in Eq. (6), $p(\boldsymbol{\vartheta})$, is the prior probability distribution for the selection coefficients. We can adopt a uniform prior over the interval $[-1, +\infty)$ for each selection coefficient if our prior knowledge is poor.

The second term of the product in Eq. (6), $p(\boldsymbol{x}_{1:K} \mid \boldsymbol{\vartheta})$, is the probability distribution for the population gamete frequency trajectories at all sampling time points. As the Wright-Fisher diffusion is a Markov process, we can decompose the probability distribution $p(\boldsymbol{x}_{1:K} \mid \boldsymbol{\vartheta})$ as

$$p(\boldsymbol{x}_{1:K} \mid \boldsymbol{\vartheta}) = p(\boldsymbol{x}_1 \mid \boldsymbol{\vartheta}) \prod_{k=1}^{K-1} p(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k; \boldsymbol{\vartheta}),$$

where $p(x_1 \mid \boldsymbol{\vartheta})$ is the prior probability distribution for the population gamete frequencies at the initial sampling time point, set to be a flat Dirichlet distribution over the state space $\Omega_{\boldsymbol{X}}$ if our

prior knowledge is poor, and $p(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k; \boldsymbol{\vartheta})$ is the transition probability density function of the Wright-Fisher diffusion $\boldsymbol{X}$ between two consecutive sampling time points for $k = 1, 2, \ldots, K-1$, solving the Kolmogorov backward equation (or its adjoint) associated with the Wright-Fisher diffusion in Eq. (5).

The third term of the product in Eq. (6), $p(\boldsymbol{g}_{1:K} \mid \boldsymbol{x}_{1:K})$, is the probability distribution for the genotypes of all individuals in the sample given the population gamete frequency trajectories at all sampling time points. With the conditional independence from our HMM framework (see Figure 1), we can decompose the probability distribution $p(\boldsymbol{g}_{1:K} \mid \boldsymbol{x}_{1:K})$ as

$$p(\boldsymbol{g}_{1:K} \mid \boldsymbol{x}_{1:K}) = \prod_{k=1}^{K} p(\boldsymbol{g}_k \mid \boldsymbol{x}_k) = \prod_{k=1}^{K} \prod_{n=1}^{N_k} p(\boldsymbol{g}_{n,k} \mid \boldsymbol{x}_k),$$

where $p(\boldsymbol{g}_{n,k} \mid \boldsymbol{x}_k)$ is the probability distribution for the genotypes $\boldsymbol{g}_{n,k}$ of sampled individual n given the gamete frequencies $\boldsymbol{x}_k$ of the population. Under the assumption that all individuals in the sample are drawn from the population in their adulthood (*i.e.*, the stage after selection but before recombination in the life cycle, see He et al. (2017)), the probability of observing the sampled individual genotypes $\boldsymbol{g}_{n,k} = (i + i', j + j')$ given the population gamete frequencies $\boldsymbol{x}_k$ can be calculated with

$$p(\boldsymbol{g}_{n,k} \mid \boldsymbol{x}_k) = \begin{cases} \dfrac{(1 + s_{ij,i'j'}) x_{i'j',k} x_{ij,k}}{\sum_{i,j=0}^{1} \sum_{i',j'=0}^{1} (1 + s_{ij,i'j'}) x_{i'j',k} x_{ij,k}}, & \text{if } i + i' \neq 1 \text{ and } j + j' \neq 1 \\ \dfrac{(1 + s_{00,11}) 2x_{11,k} x_{00,k} + (1 + s_{01,10}) 2x_{10,k} x_{01,k}}{\sum_{i,j=0}^{1} \sum_{i',j'=0}^{1} (1 + s_{ij,i'j'}) x_{i'j',k} x_{ij,k}}, & \text{if } i + i' = 1 \text{ and } j + j' = 1 \\ \dfrac{(1 + s_{ij,i'j'}) 2x_{i'j',k} x_{ij,k}}{\sum_{i,j=0}^{1} \sum_{i',j'=0}^{1} (1 + s_{ij,i'j'}) x_{i'j',k} x_{ij,k}}, & \text{otherwise} \end{cases} \tag{7}$$

for $i, j, i', j' = 0, 1$.

The fourth term of the product in Eq. (6), $p(\boldsymbol{r}_{1:K} \mid \boldsymbol{g}_{1:K})$, is the probability of observing the reads of all sampled individuals given their corresponding genotypes. Using the conditional independence from our HMM framework, as shown in Figure 1, we can decompose the probability $p(\boldsymbol{r}_{1:K} \mid \boldsymbol{g}_{1:K})$ as

$$p(\boldsymbol{r}_{1:K} \mid \boldsymbol{g}_{1:K}) = \prod_{k=1}^{K} p(\boldsymbol{r}_k \mid \boldsymbol{g}_k) = \prod_{k=1}^{K} \prod_{n=1}^{N_k} p(\boldsymbol{r}_{n,k} \mid \boldsymbol{g}_{n,k}) = \prod_{k=1}^{K} \prod_{n=1}^{N_k} \prod_{l=1}^{2} p(\boldsymbol{r}_{l,n,k} \mid g_{l,n,k}),$$

where $p(\boldsymbol{r}_{l,n,k} \mid g_{l,n,k})$ is the probability of observing the reads $\boldsymbol{r}_{l,n,k}$ of sampled individual n at locus l given its genotype $g_{l,n,k}$, known as the genotype likelihood, which is commonly available with aDNA data.

2.2.2. Adaptive particle marginal Metropolis-Hastings

Similar to He et al. (2022), we carry out our posterior computation by the PMMH algorithm (Andrieu et al., 2010) that enables us to jointly update the selection coefficients and population gamete frequency trajectories. More specifically, we estimate the marginal likelihood

$$p(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta}) = \int_{\Omega_{\mathbf{X}}^K} p(\boldsymbol{\vartheta}, \boldsymbol{x}_{1:K} \mid \boldsymbol{r}_{1:K})\, d\boldsymbol{x}_{1:K}$$

through the bootstrap particle filter (Gordon et al., 1993), where we generate the particles from the Wright-Fisher SDE's in Eq. (5) by the Euler-Maruyama scheme. The product of the average weights of the set of particles at the sampling time points $\boldsymbol{t}_{1:K}$ yields an unbiased estimate of the marginal likelihood $p(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta})$, denoted by $\hat{p}(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta})$. The population gamete frequency trajectories $\boldsymbol{x}_{1:K}$ are sampled once from the final set of particles with their relevant weights.

Although the PMMH algorithm has been shown to work well in He et al. (2022), in practice, its performance depends strongly on the choice of the proposal. In this work, due to the increase in the number of selection coefficients required to be estimated, choosing an appropriate proposal to ensure computational efficiency becomes challenging. To resolve this issue, we adopt a random walk proposal with covariance matrix $\boldsymbol{\Gamma}$, denoted by $q(\,\cdot \mid \boldsymbol{\vartheta}; \boldsymbol{\Gamma})$, the Gaussian probability density function with mean vector $\boldsymbol{\vartheta}$ and covariance matrix $\boldsymbol{\Gamma}$, and under ideal conditions, the optimal choice of the covariance matrix $\boldsymbol{\Gamma}$ is a rescaled version of the covariance matrix of the posterior (Roberts & Rosenthal, 2001). Given that the covariance matrix of the posterior is commonly not available in advance, we adopt the adaptation strategy (Vihola, 2012) that can dynamically align the covariance matrix of the proposal with that of the posterior based on accepted samples. More specifically, we prespecify a target acceptance rate, denoted by A^*, and a step size sequence (decaying to zero), denoted $\{\eta^i\}_{i\geq 1}$, where the superscript denotes the iteration. The covariance matrix is updated by following the iteration formula

$$\boldsymbol{\Gamma}^i = \boldsymbol{\Gamma}^{i-1} + \eta^i (A^i - A^*) \frac{(\boldsymbol{\vartheta}^i - \boldsymbol{\vartheta}^{i-1})(\boldsymbol{\vartheta}^i - \boldsymbol{\vartheta}^{i-1})^\intercal}{\|\boldsymbol{\vartheta}^i - \boldsymbol{\vartheta}^{i-1}\|^2} \tag{8}$$

with the covariance matrix $\boldsymbol{\Gamma}^1$ (*e.g.*, $\boldsymbol{\Gamma}^1 = \sigma^2 \boldsymbol{I}$) and selection coefficients $\boldsymbol{\vartheta}^1 \sim p(\boldsymbol{\vartheta})$, where

$$\boldsymbol{\vartheta}^i \sim q(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}^{i-1}; \boldsymbol{\Gamma}^{i-1})$$

and

$$A^i = \frac{p(\boldsymbol{\vartheta}^i)}{p(\boldsymbol{\vartheta}^{i-1})} \frac{\hat{p}(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta}^i)}{\hat{p}(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta}^{i-1})} \frac{q(\boldsymbol{\vartheta}^{i-1} \mid \boldsymbol{\vartheta}^i; \boldsymbol{\Gamma}^{i-1})}{q(\boldsymbol{\vartheta}^i \mid \boldsymbol{\vartheta}^{i-1}; \boldsymbol{\Gamma}^{i-1})}. \tag{9}$$

Such an adaptation strategy can also coerce the acceptance rate. In practice, the target acceptance rate is set to $A^* \in [0.234, 0.440]$, and the step size sequence is defined as $\eta^i = i^{-\gamma}$ with $\gamma \in (0.5, 1]$ (Vihola, 2012). See Luengo et al. (2020) and references therein for other adaptation strategies.

For the sake of clarity, we write down the robust adaptive version of the PMMH algorithm for our posterior computation:

Step 1: Initialise the selection coefficients $\boldsymbol{\vartheta}$ and population gamete frequency trajectories $\boldsymbol{x}_{1:K}$:

- Step 1a: Draw $\boldsymbol{\vartheta}^1 \sim p(\boldsymbol{\vartheta})$.
- Step 1b: Run a bootstrap particle filter with $\boldsymbol{\vartheta}^1$ to get $\hat{p}(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta}^1)$ and $\boldsymbol{x}^1_{1:K}$.
- Step 1c: Initialise $\boldsymbol{\Gamma}^1$.

Repeat Step 2 until enough samples of the selection coefficients $\boldsymbol{\vartheta}$ and population gamete frequency trajectories $\boldsymbol{x}_{1:K}$ have been attained:

Step 2: Update the selection coefficients $\boldsymbol{\vartheta}$ and population gamete frequency trajectories $\boldsymbol{x}_{1:K}$:

- Step 2a: Draw $\boldsymbol{\vartheta}^i \sim q(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}^{i-1}; \boldsymbol{\Gamma}^{i-1})$.
- Step 2b: Run a bootstrap particle filter with $\boldsymbol{\vartheta}^i$ to get $\hat{p}(\boldsymbol{r}_{1:K} \mid \boldsymbol{\vartheta}^i)$ and $\boldsymbol{x}^i_{1:K}$.
- Step 2c: Update $\boldsymbol{\Gamma}^i$ through Eqs. (8) and (9).
- Step 2d: Accept $\boldsymbol{\vartheta}^i$ and $\boldsymbol{x}^i_{1:K}$ with A^i and set $\boldsymbol{\vartheta}^i = \boldsymbol{\vartheta}^{i-1}$ and $\boldsymbol{x}^i_{1:K} = \boldsymbol{x}^{i-1}_{1:K}$ otherwise.

With enough samples of the selection coefficients $\boldsymbol{\vartheta}$ and population gamete frequency trajectories $\boldsymbol{x}_{1:K}$, we produce the minimum mean square error estimates for the selection coefficients $\boldsymbol{\vartheta}$ and population gamete frequency trajectories $\boldsymbol{x}_{1:K}$ by calculating their posterior means.

As in He et al. (2022), our procedure can allow the selection coefficients $s_{ij,i'j'}$ to change over time (piecewise constant), *e.g.*, let the selection coefficients $s_{ij,i'j'}(t) = s^-_{ij,i'j'}$ if $t < \tau$ otherwise $s_{ij,i'j'}(t) = s^+_{ij,i'j'}$, where τ is the time of an event that might change selection, *e.g.*, the times of plant and animal domestication. The only modification required is to simulate the population

gamete frequency trajectories $\boldsymbol{x}_{1:K}$ according to the Wright-Fisher diffusion with the selection coefficients $s^-_{ij,i'j'}$ for $t < \tau$ and $s^+_{ij,i'j'}$ for $t \geq \tau$, respectively. In this setup, we propose a scheme to test the hypothesis whether selection changes at time τ for each phenotypic trait, including estimating their selection differences, through computing the posterior $p(\Delta s_{ij,i'j'} \mid \boldsymbol{r}_{1:K})$ from the PMMH samples of the selection coefficients $s^-_{ij,i'j'}$ and $s^+_{ij,i'j'}$, where $\Delta s_{ij,i'j'} = s^+_{ij,i'j'} - s^-_{ij,i'j'}$ denotes the change in the selection coefficient at time τ. Note that our method can handle the case that the events that might change selection are different for different phenotypic traits (*i.e.*, the time τ could be taken to be different values for different phenotypic traits).

3. Results

In this section, we employ our approach to reanalyse the published ancient horse DNA data from earlier studies of Ludwig et al. (2009), Pruvost et al. (2011) and Wutke et al. (2016), where they sequenced 201 ancient horse samples in total ranging from a pre- to a post-domestication period for eight loci coding for horse coat colouration. In particular, we perform the inference of selection that acts on the base coat colour determined by *ASIP* and *MC1R* and the pinto coat pattern controlled by *KIT13* and *KIT16* and test the hypothesis that there was a shift in human preferences for horse coat colours and patterns in the Middle Ages when horses started to be differentiated by use (Wutke et al., 2016). Various simulation studies, assessing the performance of our procedure, are available in the supplement.

It has been shown by Gaunitz et al. (2018) that the ancient horse samples providing the earliest archaeological evidence of domestication (occurred 5500 years ago) are not the ancestors of modern domestic horses, which only spread out from the Pontic–Caspian steppe 4200 years ago (the domestication of modern horses) when they rapidly started to become dominant across Eurasia (Librado et al., 2021). Also, horse populations outside the Pontic–Caspian steppe and older than 4200 years ago show strong geographic structure, *e.g.*, horse populations from Iberia and Siberia are completely different (Fages et al., 2019). Hence, to eliminate these confounding effects, we exclude the ancient horse samples older than 4200 years ago in our analyses, but we assume that the respective mutations occurred at all pigmentation loci 4200 years ago.

As Wutke et al. (2016) only provided called genotypes for each gene (including missing calls), we use the same scheme as in He et al. (2022) to convert to corresponding genotype likelihoods. More specifically, we take the genotype likelihood of the called genotype to be 1 and those of the

remaining two to be 0 if the genotype is called, and otherwise, all possible (ordered) genotypes are assigned equal genotype likelihoods (normalised to sum to 1). Genotype likelihoods for each gene can be found in Table S1.

In what follows, we set the average length of a generation of the horse to be eight years and use the time-varying size of the horse population estimated by Der Sarkissian et al. (2015) (see Figure S1) with the reference population size $N_0 = 16000$ (*i.e.*, the most recent population size) like Schraiber et al. (2016) unless otherwise noted. Since the flat Dirichlet prior for the starting population gamete frequencies is more likely to produce low linkage disequilibrium, we generate the starting population gamete frequencies $\boldsymbol{x}_1$ through the following procedure:

Step 1: Draw $y_1, y_2 \sim \text{Uniform}(0, 1)$.

Step 2: Draw $D \sim \text{Uniform}(\max\{-y_1y_2, -(1-y_1)(1-y_2)\}, \min\{y_1(1-y_2), (1-y_1)y_2\})$.

Step 3: Set $\boldsymbol{x}_1 = ((1-y_1)(1-y_2) + D, (1-y_1)y_2 - D, y_1(1-y_2) - D, y_1y_2 + D)$.

Note that y_1 and y_2 denote the starting population frequencies of the mutant allele at the two loci, respectively, and D is the coefficient of linkage disequilibrium. We run our adaptive PMMH algorithm with 1000 particles and 20000 iterations, where we set the target acceptance rate to $A^* = 0.4$ and define the step size sequence as $\eta_i = i^{-2/3}$ for $i = 1, 2, \ldots, 20000$. We divide each generation into five subintervals in the Euler-Maruyama scheme. We discard a burn-in of 10000 iterations and thin the remaining iterations by keeping every fifth value.

3.1. Horse base coat colours

The horse genes *ASIP* and *MC1R* are primarily responsible for determination of base coat colours (*i.e.*, bay, black and chestnut). *ASIP*, located on chromosome 22, is epistatic to *MC1R*, located on chromosome 3 (Rieder et al., 2001). At each locus, there are two allele types, labelled A and a for *ASIP* and E and e for *MC1R*, respectively, where the capital letter represents the ancestral allele and the small letter represents the mutant allele. See Table 1 for the genotype-phenotype map at *ASIP* and *MC1R* for base coat colours.

3.1.1. Wright-Fisher diffusion for ASIP and MC1R

Let us consider a horse population represented by the alleles at *ASIP* and *MC1R* evolving under selection over time, which induces four possible haplotypes AE, Ae, aE and ae, labelled

		MC1R		
		E/E	*E/e*	*e/e*
	A/A	bay	bay	chestnut
ASIP	*A/a*	bay	bay	chestnut
	a/a	black	black	chestnut

Table 1: The genotype-phenotype map at *ASIP* and *MC1R* for horse base coat colours.

haplotypes 00, 01, 01 and 11, respectively. We take the relative viabilities of the three phenotypes, *i.e.*, the bay, black and chestnut coat, to be 1, $1 + s_b$ and $1 + s_c$, respectively, where s_b is the selection coefficient of the black coat against the bay coat and s_c is the selection coefficient of the chestnut coat against the bay coat. See Table 2 for the relative viabilities of all genotypes at *ASIP* and *MC1R*.

	AE	*Ae*	*aE*	*ae*
AE	1	1	1	1
Ae	1	$1 + s_c$	1	$1 + s_c$
aE	1	1	$1 + s_b$	$1 + s_b$
ae	1	$1 + s_c$	$1 + s_b$	$1 + s_c$

Table 2: Relative viabilities of all genotypes at *ASIP* and *MC1R*.

We measure time in units of $2N_0$ generations and scale the selection coefficients $\alpha_b = 2N_0 s_b$, $\alpha_c = 2N_0 s_c$ and recombination rate $\rho = 4N_0 r$, respectively. Let $X_{ij}(t)$ be the gamete frequency of haplotype ij at time t, which satisfies the Wright-Fisher SDE's in Eq. (5). More specifically, the drift term $\boldsymbol{\mu}(t, \boldsymbol{x})$ can be simplified with the genotype-phenotype map shown in Table 2 as

$$\begin{aligned}
\mu_{00}(t, \boldsymbol{x}) &= -\alpha_b x_{10}(x_{00}x_{11} + x_{00}x_{1*}) - \alpha_c x_{00}x_{*1}x_{*1} - \frac{\rho}{2}(x_{00}x_{11} - x_{01}x_{10}) \\
\mu_{01}(t, \boldsymbol{x}) &= -\alpha_b x_{10}(x_{01}x_{11} + x_{01}x_{1*}) + \alpha_c x_{01}x_{*0}x_{*1} + \frac{\rho}{2}(x_{00}x_{11} - x_{01}x_{10}) \\
\mu_{10}(t, \boldsymbol{x}) &= -\alpha_b x_{10}(x_{10}x_{11} + x_{10}x_{1*} - x_{1*}) - \alpha_c x_{10}x_{*1}x_{*1} + \frac{\rho}{2}(x_{00}x_{11} - x_{01}x_{10}) \\
\mu_{11}(t, \boldsymbol{x}) &= -\alpha_b x_{10}(x_{11}x_{11} + x_{11}x_{1*} - x_{11}) + \alpha_c x_{11}x_{*0}x_{*1} - \frac{\rho}{2}(x_{00}x_{11} - x_{01}x_{10}),
\end{aligned}$$

where we take the scaled recombination rate to be $\rho = 2N_0$ since the two genes are located on separate chromosomes.

3.1.2. Selection of horse base coat colours

Given that *ASIP* and *MC1R* are located on separate chromosomes, we generate the starting population gamete frequencies with the procedure described above but fix the coefficient of link-

age disequilibrium to zero. The resulting posteriors for the selection coefficients and population phenotype frequency trajectories are shown in Figure 2, and their estimates as well as the 95% highest posterior density (HPD) intervals are summarised in Table S2.

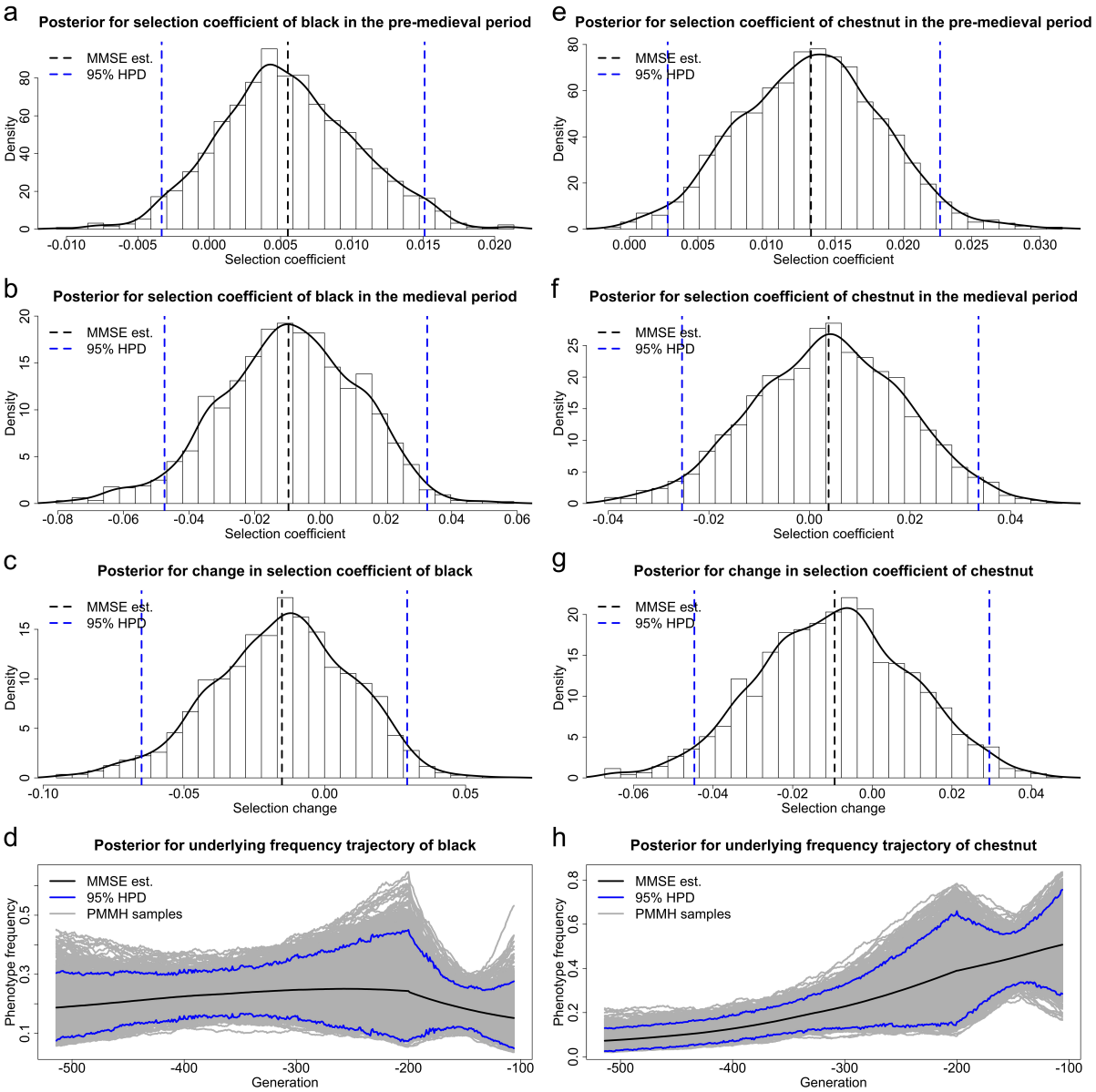

Figure 2: Posteriors for selection of base coat colours in the pre-medieval and medieval period and underlying frequency trajectories of each phenotypic trait in the population, (a)-(d) for the black coat and (e)-(h) for the chestnut coat, respectively.

Our estimate for the selection coefficient s_b^- is 0.0055 (95% HPD: $[-0.0033, 0.0151]$). Though the 95% HPD interval contains 0, we still find that the black coat was most probably favoured by selection in the pre-medieval period (with posterior probability for positive selection being 0.880). Our estimate for the selection coefficient s_b^+ is -0.0097 (95% HPD: $[-0.0475, 0.0326]$),

providing weak evidence that the black coat was selectively deleterious in the medieval period (with posterior probability for negative selection being 0.666). From Figure 2c, we observe that a negative shift took place in selection of the black coat in Middle Ages (with posterior probability for a negative change being 0.728). Our estimate for the underlying frequency trajectory of the black coat reveals a slow increase in the pre-medieval period and then a slight decrease in the medieval period.

Our estimate for the selection coefficient s_c^- is 0.0133 (95% HPD: $[-0.0028, 0.0227]$), providing enough evidence to support that the chestnut coat was positively selected in the pre-medieval period (with posterior probability for positive selection being 0.998). Our estimate for the selection coefficient s_c^+ is 0.0038 (95% HPD: $[-0.0254, 0.0336]$), suggesting that in the medieval period the chestnut coat was still favoured by selection (with posterior probability for positive selection being 0.613). Figure 2g shows that a negative shift occurred in selection of the chestnut coat in Middle Ages (with posterior probability for a negative change being 0.698). We see from Figure 2e that the underlying frequency of the chestnut keeps increasing from the domestication of modern horses.

See Figure S2 and Table S3 for the results produced with a flat Dirichlet prior for the starting population gamete frequencies, which are consistent with those shown in Figure 2.

3.2. Horse pinto coat patterns

The horse genes *KIT13* and *KIT16* are mainly responsible for determination of pinto coat patterns (*i.e.*, tobiano and sabino), both of which reside on chromosome 3, 4668 base pairs (bp) apart, with an average rate of recombination 10^{-8} crossover/bp (Dumont & Payseur, 2008). At each locus, there are two allele types, labelled *KM0* for the ancestral allele and *KM1* for the mutant allele at *KIT13* and *sb1* for the ancestral allele and *SB1* for the mutant allele at *KIT16*, respectively. See Table 3 for the genotype-phenotype map at *KIT13* and *KIT16* for pinto coat patterns, where the solid coat pattern refers to a coat that neither tobiano nor sabino is present, and the mixed coat pattern refers to a coat that is a mixture between tobiano and sabino.

3.2.1. Wright-Fisher diffusion for KIT13 and KIT16

We now consider a horse population represented by the alleles at *KIT13* and *KIT16* evolving under selection over time. Such a setup gives rise to four possible haplotypes *KM0sb1*, *KM0SB1*,

		KIT16		
		sb1/sb1	*sb1/SB1*	*SB1/SB1*
	KM0/KM0	solid	sabino	sabino
KIT13	*KM0/KM1*	tobiano	mixed	mixed
	KM1/KM1	tobiano	mixed	mixed

Table 3: The genotype-phenotype map at *KIT13* and *KIT16* for horse pinto coat patterns.

KM1sb1 and *KM1SB1*, labelled haplotypes 00, 01, 01 and 11, respectively. We take the relative viabilities of the four phenotypes, *i.e.*, the solid, tobiano, sabino and mixed coat, to be 1, $1+s_{to}$, $1+s_{sb}$ and $1+s_{mx}$, respectively, where s_{to} is the selection coefficient of the tobiano coat against the solid coat, s_{sb} is the selection coefficient of the sabino coat against the solid coat, and s_{mx} is the selection coefficient of the mixed coat against the solid coat. See Table 4 for the relative viabilities of all genotypes at *KIT13* and *KIT16*.

	KM0sb1	*KM0SB1*	*KM1sb1*	*KM1SB1*
KM0sb1	1	$1+s_{sb}$	$1+s_{to}$	$1+s_{mx}$
KM0SB1	$1+s_{sb}$	$1+s_{sb}$	$1+s_{mx}$	$1+s_{mx}$
KM1sb1	$1+s_{to}$	$1+s_{mx}$	$1+s_{to}$	$1+s_{mx}$
KM1SB1	$1+s_{mx}$	$1+s_{mx}$	$1+s_{mx}$	$1+s_{mx}$

Table 4: Relative viabilities of all genotypes at *KIT13* and *KIT16*.

We measure time in units of $2N_0$ generations and scale the selection coefficients $\alpha_{to} = 2N_0 s_{to}$, $\alpha_{sb} = 2N_0 s_{sb}$, $\alpha_{mx} = 2N_0 s_{mx}$ and recombination rate $\rho = 4N_0 r$, respectively. Let $X_{ij}(t)$ be the gamete frequency of haplotype ij at time t, which follows the Wright-Fisher SDE's in Eq. (5). In particular, the drift term $\boldsymbol{\mu}(t, \boldsymbol{x})$ can be simplified through Table 4 as

$$
\begin{aligned}
\mu_{00}(t,\boldsymbol{x}) &= -\alpha_{to}x_{00}(x_{10}(x_{00}+x_{*0})-x_{10}) - \alpha_{sb}x_{00}(x_{01}(x_{00}+x_{0*})-x_{01}) \\
&\quad - \alpha_{mx}x_{00}(2x_{01}x_{10}+x_{11}-x_{11}^2) - \frac{\rho}{2}(x_{00}x_{11}-x_{01}x_{10}) \\
\mu_{01}(t,\boldsymbol{x}) &= -\alpha_{to}x_{01}x_{10}(x_{00}+x_{*0}) - \alpha_{sb}x_{01}(x_{01}(x_{00}+x_{0*})-x_{0*}) \\
&\quad - \alpha_{mx}x_{01}((2x_{01}x_{10}+x_{11}-x_{11}^2)-x_{10}) + \frac{\rho}{2}(x_{00}x_{11}-x_{01}x_{10}) \\
\mu_{10}(t,\boldsymbol{x}) &= -\alpha_{to}x_{10}(x_{10}(x_{00}+x_{*0})-x_{*0}) - \alpha_{sb}x_{10}x_{01}(x_{00}+x_{0*}) \\
&\quad - \alpha_{mx}x_{10}((2x_{01}x_{10}+x_{11}-x_{11}^2)-x_{01}) + \frac{\rho}{2}(x_{00}x_{11}-x_{01}x_{10}) \\
\mu_{11}(t,\boldsymbol{x}) &= -\alpha_{to}x_{11}x_{10}(x_{00}+x_{*0}) - \alpha_{sb}x_{11}x_{01}(x_{00}+x_{0*}) \\
&\quad - \alpha_{mx}x_{11}((2x_{01}x_{10}+x_{11}-x_{11}^2)-(1-x_{11})) - \frac{\rho}{2}(x_{00}x_{11}-x_{01}x_{10}).
\end{aligned}
$$

3.2.2. Selection of horse pinto coat patterns

To our knowledge, the mixed coat has never been found in the horse population, so we fix the selection coefficient $s_{mx} = -1$ over time. The resulting posteriors for the selection coefficients and population phenotype frequency trajectories are illustrated in Figure 3, and their estimates as well as the 95% HPD intervals are summarised in Table S4.

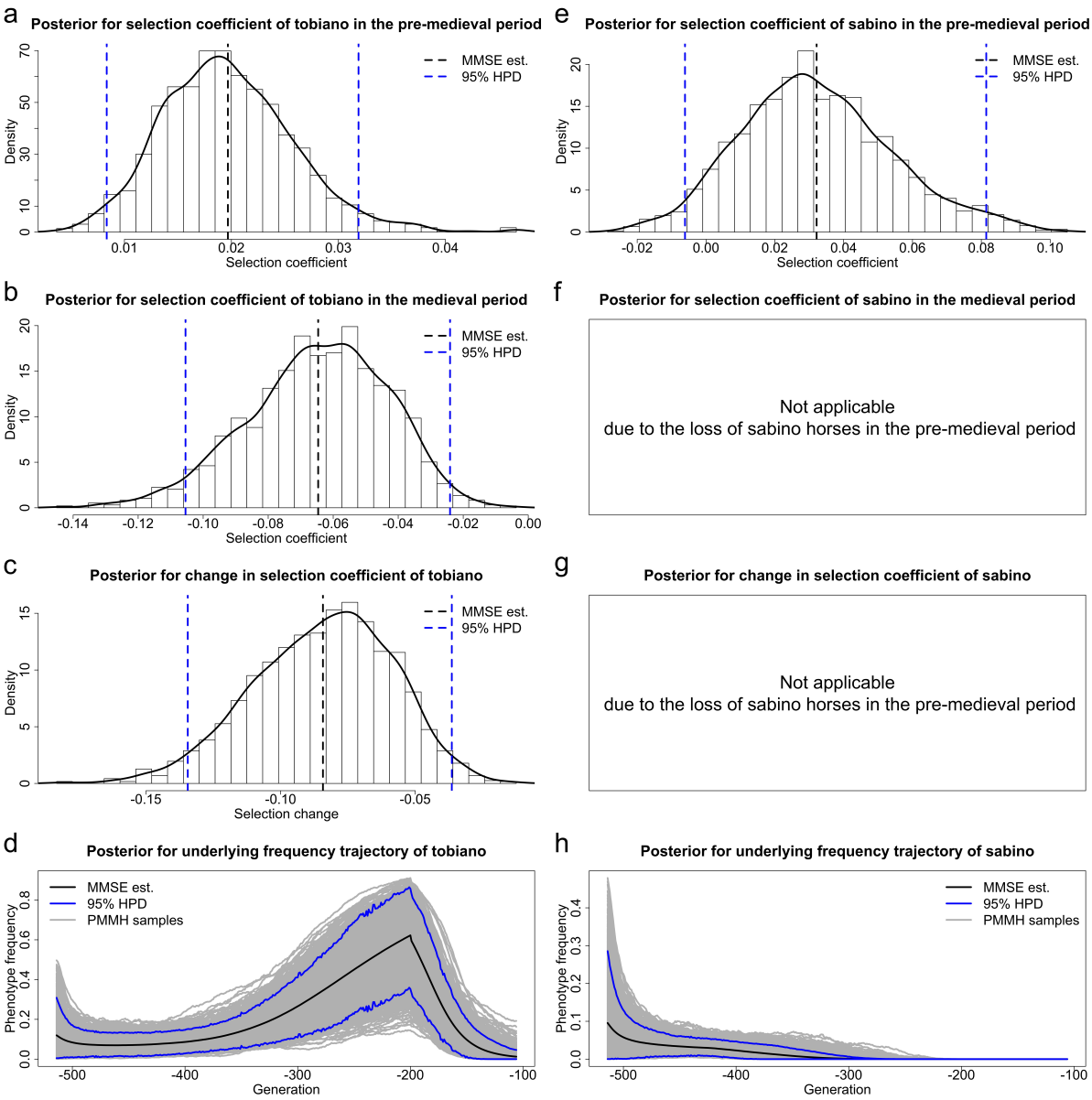

Figure 3: Posteriors for selection of pinto coat patterns in the pre-medieval and medieval period and underlying frequency trajectories of each phenotypic trait in the population, (a)-(d) for the tobiano coat and (e)-(h) for the sabino coat, respectively.

Our estimate for the selection coefficient s_{to}^{-} is 0.0197 (95% HPD: $[0.0084, 0.0319]$) and the selection coefficient s_{to}^{+} is -0.0646 (95% HPD: $[-0.1054, -0.0241]$). Our results reveal sufficient

evidence to support that the tobiano coat was favoured by selection since the domestication of modern horses (with posterior probability for positive selection being 1) but became selectively deleterious in the medieval period (with posterior probability for negative selection being 1). Figure 3c suggests that a negative shift took place in selection of the tobiano coat in the Middle Ages (with posterior probability for a negative change being 1). Our estimate for the underlying frequency trajectory of the tobiano coat indicates that the frequency of the tobiano coat grows substantially after modern horse domestication and then drops sharply in the medieval period.

Our estimate for the selection coefficient s_{sb}^{-} is 0.0321 (95% HPD interval $[-0.0061, 0.0814]$), which shows compelling evidence of positive selection acting on the sabino coat (with posterior probability for positive selection being 0.941). However, as illustrated in Figure 3c, we observe that the frequency of the sabino coat declines slowly from the domestication of modern horses until the loss of the sabino coat in approximately 24 BC (*i.e.*, the earliest time that the upper and lower bounds of the 95% HPD interval for the frequency of the sabino coat are both zero), probably caused by that the sabino coat was somewhat out-competed by the tobiano coat under the tight linkage between *KIT13* and *KIT16*.

Note, we only present the resulting posterior for the selection coefficient s_{sb}^{-}. This is because our results show that the sabino coat became extinct in the pre-medieval period (see Figure 3h). Without genetic variation data, our PMMH algorithm fails to converge in reasonable time for the selection coefficient s_{sb}^{+}, which however has little effect on estimation of the remaining three (see Figure S3, where we repeatedly run our procedure to estimate the selection coefficients s_{to}^{-}, s_{to}^{+} and s_{sb}^{-} with different prespecified values of the selection coefficient s_{sb}^{+} that are uniformly drawn from $[-1, 1]$).

We also provide the results produced with a flat Dirichlet prior for the starting population gamete frequencies (see Figure S4 and Table S5) and that we co-estimate the selection coefficient s_{mx} (see Figure S5 and Table S6). Our estimate for the selection coefficient s_{mx} is -0.5621 (95% HPD: $[-0.9645, -0.2262]$) in the pre-medieval period. Such strong negative selection resulted in a quick loss of the mixed coat right after the domestication of modern horses (see our estimate for the underlying frequency trajectory of the mixed coat in Figure S5l). All are consistent with those shown in Figure 3.

4. Discussion

To overcome the key shortcoming of He et al. (2022), which does not aim to consider genetic interactions, in this work we have introduced a novel Bayesian procedure for inferring temporally variable selection from the data on aDNA sequences with the flexibility of modelling linkage and epistasis. We build our method upon the two-layer HMM framework of He et al. (2022), thereby enabling sample uncertainty arising from the damage and fragmentation of aDNA molecules. To allow genetic interactions, we introduce a Wright-Fisher diffusion to characterise the underlying evolutionary dynamics of two linked genes under phenotypic selection, which is modelled through the differential fitness of different phenotypic traits with a genotype-phenotype map. To ensure computational efficiency, we carry out the posterior computation with a robust adaptive PMMH algorithm, where unlike the original PMMH of Andrieu et al. (2010), we adopt the adaptation strategy of Vihola (2012) to tune the covariance structure of the proposal to achieve a coerced acceptance rate in our procedure. Furthermore, our approach permits the reconstruction of the underlying population gamete frequency trajectories and provides the flexibility of modelling time-varying demographic histories.

We have analysed horse pigmentation loci, *e.g.*, *ASIP* and *MC1R* associated with base coat colours and *KIT13* and *KIT16* associated with pinto coat patterns, based on the ancient horse samples from previous studies of Ludwig et al. (2009), Pruvost et al. (2011) and Wutke et al. (2016). Our results illustrate that black, chestnut, tobiano and sabino horses were all favoured by selection from the domestication of modern horses, but selection acting on horse coat colours and patterns changed in the Middle Ages, which could be resulted from shifts in uses, cultures and religions. For example, our results provide strong evidence that tobiano horses experienced positive selection in the pre-medieval period but negative selection in the medieval period. This can be explained as resulting in part from that spotting could be the mark used to distinguish wild horses from domestic horses after modern horses were domesticated, but that differentiation might be no longer necessary as wild horses became scarcer and extinct. From historical records, we also find that initially spotted horses were associated with positive cultural image but had a negative connotation after a series of epidemics in the Middle Ages, which could have significant consequences on their popularity. In addition, the advent of weaponry in the Middle Ages could influence the preference of spotted horses since they were an easier target than solid horses. See

Wutke et al. (2016) and references therein.

We have run additional simulations to test our procedure (see File S2 for simulation settings and results). Our results reveal that our approach can accurately estimate selection coefficients and discriminate whether selection acts on the phenotypic trait and whether selection changes at the given time (see Figures S6–S9 and Tables S9–S12, where we show the results produced with simulations that mimicked the ancient horse samples but in more general settings). Although we test our method with the demographic histories shown in Figure S1 and the genotype-phenotype maps described in Tables 1 and 3, in principle, the conclusions we have drawn here hold for other demographic histories and genotype-phenotype maps. Our simulation studies also indicate that misspecifying demographic histories can greatly alter the inference of selection (see Figures S10 and S11 and Tables S13 and S14), and therefore a method that can jointly estimate demographic and selective parameters is anticipated to boost the power, which can be an important topic of future research. Although our work infers selection on phenotypes, selection on genotypes can be measured if we assume 10 distinct phenotypes. Unlike He et al. (2020b), where they defined the fitness values of single-locus genotypes with a selection coefficient and a dominance parameter and assumed the fitness values of two-locus genotypes to be determined multiplicatively from those at individual loci, our method is built upon a general model of selection, therefore enabling *e.g.*, heterozygote advantage selection.

To show the improvement attainable by modelling genetic interactions, we plot the resulting posteriors for *ASIP* and *MC1R* in Figure 4 and *KIT13* and *KIT16* in Figure 5, respectively, which are produced through the approach of He et al. (2022) with the same settings as adopted in our adaptive PMMH algorithm. We summarise the results for base coat colours and pinto coat patterns with their 95% HPD intervals in Tables S7 and S8, respectively. Moreover, additional simulation studies are left in File S2 to further show the improvement resulting from modelling linkage and epistasis (see Figures S12 and S13 and Tables S15 and S16).

The results for *ASIP* and *MC1R* produced by the method of He et al. (2022) are consistent with those illustrated in Figure 2 except for selection acting on the black coat in the medieval period. Both results suggest that in the medieval period the black coat was negatively selected, but its intensity (*i.e.*, $|s_b^+|$) estimated by the approach of He et al. (2022) is substantially lower than that produced by our method. This is because ignoring epistasis causes the *aa/ee* genotype

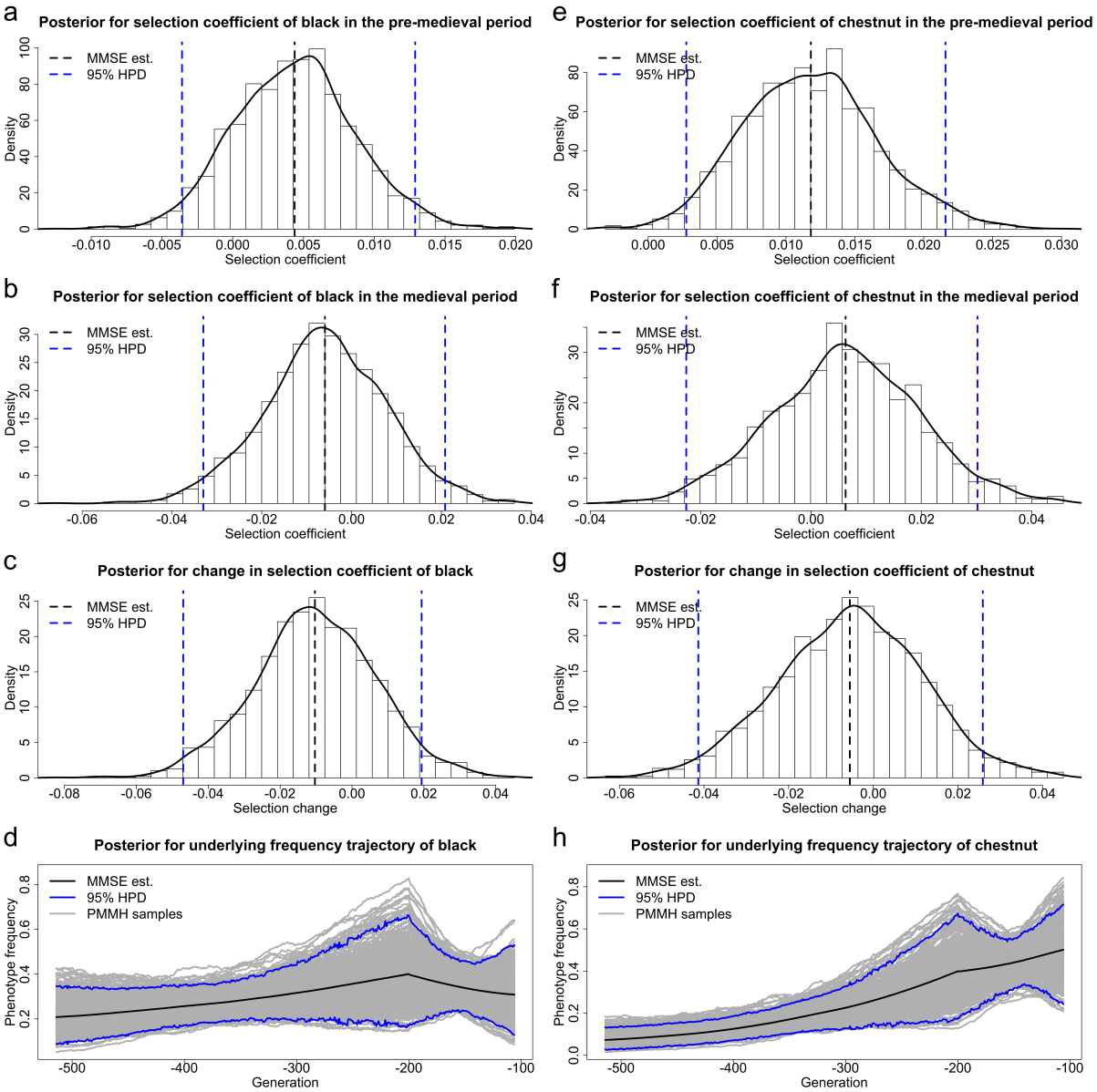

Figure 4: Posteriors for selection of base coat colours in the pre-medieval and medieval period and underlying frequency trajectories of each phenotypic trait in the population produced through the method of He et al. (2022), (a)-(d) for the black coat and (e)-(h) for the chestnut coat, respectively.

to be incorrectly attributed to the black coat, and most samples with *aa/ee* are medieval horses. The results for *KIT13* and *KIT16* produced by the approach of He et al. (2022) are compatible with those shown in Figure 3 except for selection acting on the sabino coat in the pre-medieval period. As shown in Figure 5, sabino horses experienced negative selection from modern horse domestication until extinction (in approximately AD 200), suggesting that a continuous decrease in sabino horses resulted from negative selection. However, we observe from Figure 3 that sabino horses were positively selected in the pre-medieval period but with a continuous decline in their

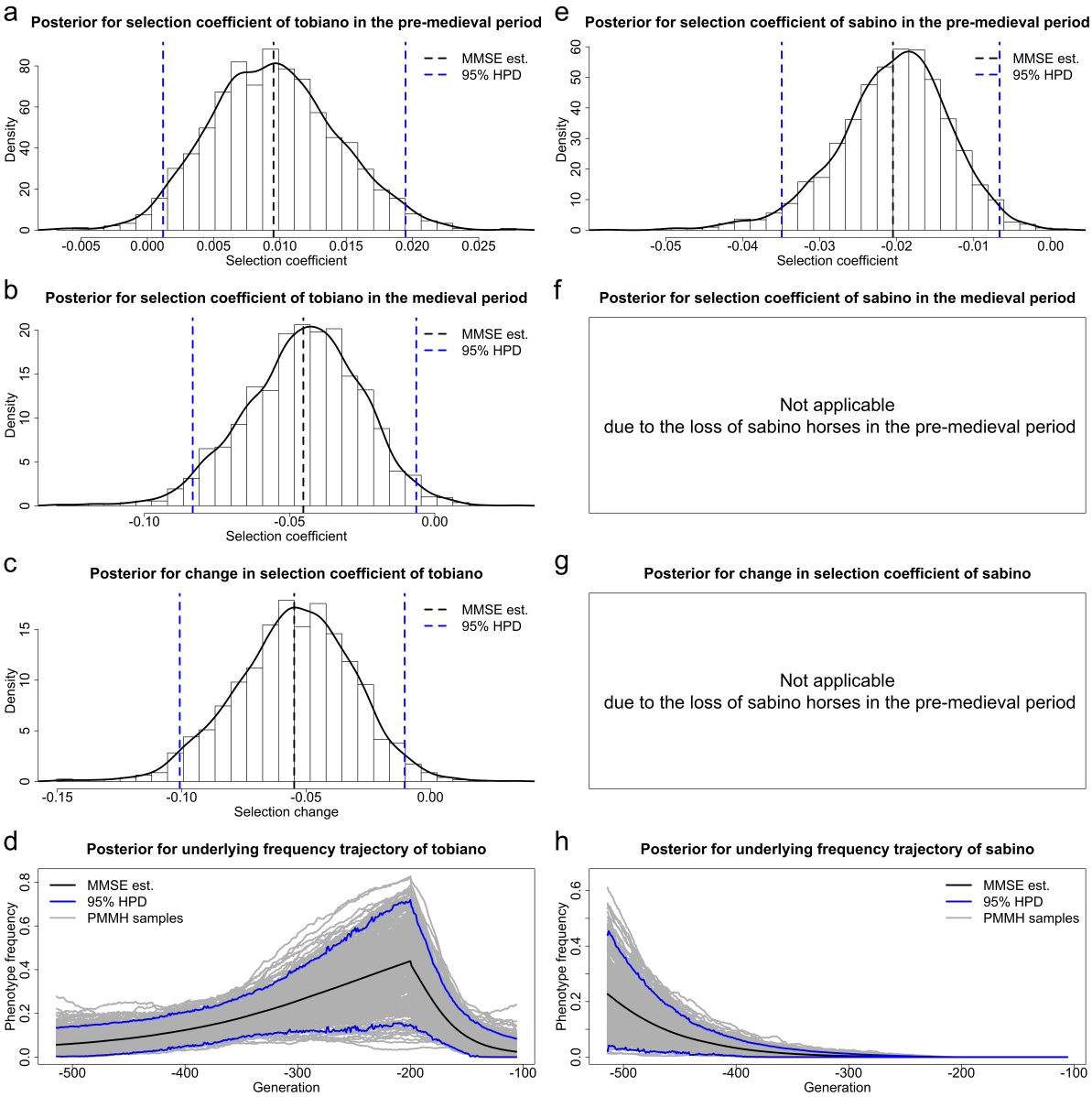

Figure 5: PPosteriors for selection of pinto coat patterns in the pre-medieval and medieval period and underlying frequency trajectories of each phenotypic trait in the population produced through the method of He et al. (2022), (a)-(d) for the tobiano coat and (e)-(h) for the sabino coat, respectively.

frequency until extinction when linkage is modelled. This drop was probably triggered by sabino horses being somewhat out-competed by tobiano horses.

For aDNA data with high degrees of uncertainty, working on genotype likelihoods rather than called genotypes has been illustrated to be beneficial in the inference of selection (He et al., 2022). Genotype likelihoods are typically calculated with the aligned reads and associated mapping and sequencing quality scores, which facilitates the incorporation of uncertainty regarding genotypes arising from base calling, alignment and assembly (Nielsen et al., 2011). Recent methodological

advances in genotype imputation, in particular those designed for imputing low-coverage data (*e.g.*, Hui et al., 2020; Rubinacci et al., 2021; Ausmees & Nettelblad, 2023), present a possibility to improve aDNA studies (da Mota et al., 2022). Our procedure can directly work on imputed genotypes (and posterior genotype probabilities produced in most imputation approaches) and can be naturally extended to allow for phased haplotypes by combining with He et al. (2020b). Note that our approach can only leverage sample uncertainty captured by genotype likelihoods. To better model sample uncertainty associated with aDNA like that introduced by post-mortem C>T and G>A deamination, an important consideration is to integrate the model for genotype likelihood calculations based on aDNA sequences, introduced by Kousathanas et al. (2017) or Renaud et al. (2019), into our HMM framework. A further assumption of this work, in common with other studies, is of temporal continuity of populations. As shown by Silva et al. (2017) and Ortega-Del Vecchyo & Slatkin (2019), this may be a strong assumption, and inferences may be confounded by a history of gene flow or population replacement. These effects can be addressed in future, using more general spatial models.

Our extension inherits the desirable features of He et al. (2022) with its key limitation that all samples are assumed to be drawn after the respective mutations occurring at both loci. Since allele age is usually unavailable, we have to restrict our inference to a certain time window, *e.g.*, from the time after which the mutant alleles at both loci have been observed in the sample or the time before which we assume that the mutant alleles at both loci have already existed in the population, which however could dramatically bias the inference of selection. Backward-in-time simulation of the Wright-Fisher diffusion (Griffiths, 2003; Coop & Griffiths, 2004) is anticipated to circumvent this problem. Moreover, how to extend our method to handle multiple interacting genes (Terhorst et al., 2015; Sohail et al., 2021, 2022) and estimate selection coefficients with their timing of changes (Shim et al., 2016; Mathieson, 2020; Mathieson & Terhorst, 2022) will also be the topics of our future investigation.

Acknowledgements

We thank the anonymous reviewers and the editor for their helpful comments on the earlier version of this work. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - `http://www.bristol.ac.uk/acrc/`.

References

Alves, J. M., Carneiro, M., Cheng, J. Y., de Matos, A. L., Rahman, M. M. et al. (2019). Parallel adaptation of rabbit populations to myxoma virus. *Science*, *363*, 1319–1326.

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*, 269–342.

Ausmees, K., & Nettelblad, C. (2023). Achieving improved accuracy for imputation of ancient DNA. *Bioinformatics*, *39*, btac738.

Bank, C., Ewing, G. B., Ferrer-Admettla, A., Foll, M., & Jensen, J. D. (2014). Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, *30*, 540–546.

Bollback, J. P., York, T. L., & Nielsen, R. (2008). Estimation of $2N_e s$ from temporal allele frequency data. *Genetics*, *179*, 497–502.

Coop, G., & Griffiths, R. C. (2004). Ancestral inference on gene trees under selection. *Theoretical Population Biology*, *66*, 219–232.

Corbin, L. J., Pope, J., Sanson, J., Antczak, D. F., Miller, D. et al. (2020). An independent locus upstream of *ASIP* controls variation in the shade of the bay coat colour in horses. *Genes*, *11*, 606.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Dehasque, M., Ávila-Arcos, M. C., Díez-del Molino, D., Fumagalli, M., Guschanski, K. et al. (2020). Inference of natural selection from ancient DNA. *Evolution Letters*, *4*, 94–108.

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P. et al. (2015). Evolutionary genomics and conservation of the endangered Przewalski's horse. *Current Biology*, *25*, 2577–2583.

Dumont, B. L., & Payseur, B. A. (2008). Evolution of the genomic rate of recombination in mammals. *Evolution*, *62*, 276–294.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A. et al. (2019). Tracking five millennia of horse management with extensive ancient genome time series. *Cell*, *177*, 1419–1435.

Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D., & Wegmann, D. (2016). An approximate Markov model for the Wright-Fisher diffusion and its application to time series data. *Genetics*, *203*, 831–846.

Fisher, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, *42*, 321–341.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N. et al. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science*, *360*, 111–114.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, *140*, 107–113.

Griffiths, R. C. (2003). The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theoretical Population Biology*, *64*, 241–251.

He, Z., Beaumont, M. A., & Yu, F. (2017). Effects of the ordering of natural selection and population regulation mechanisms on Wright-Fisher models. *G3: Genes, Genomes, Genetics*, *7*, 2095–2106.

He, Z., Beaumont, M. A., & Yu, F. (2020a). Numerical simulation of the two-locus Wright-Fisher stochastic differential equation with application to approximating transition probability densities. *bioRxiv*, (p. 213769).

He, Z., Dai, X., Beaumont, M. A., & Yu, F. (2020b). Detecting and quantifying natural selection at two linked loci from time series data of allele frequencies with forward-in-time simulations. *Genetics*, *216*, 521–541.

He, Z., Dai, X., Beaumont, M. A., & Yu, F. (2020c). Estimation of natural selection and allele age from time series allele frequency data using a novel likelihood-based approach. *Genetics*, *216*, 463–480.

He, Z., Dai, X., Lyu, W., Beaumont, M. A., & Yu, F. (2022). Estimating temporally variable selection intensity from ancient DNA data. *bioRxiv*, (p. 502345).

Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L., & Kivisild, T. (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Scientific Reports*, *10*, 1–8.

Karlin, S., & Taylor, H. E. (1981). *A Second Course in Stochastic Processes*. New York: Academic Press.

Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J. et al. (2017). Inferring heterozygosity from ancient and low coverage genomes. *Genetics*, *205*, 317–332.

Librado, P., Khan, N., Fages, A., Kusliy, M. A., Suchan, T. et al. (2021). The origins and spread of domestic horses from the Western Eurasian steppes. *Nature*, *598*, 634–640.

Loog, L., Thomas, M. G., Barnett, R., Allen, R., Sykes, N. et al. (2017). Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Molecular Biology and Evolution*, *34*, 1981–1990.

Ludwig, A., Pruvost, M., Reissmann, M., Benecke, N., Brockmann, G. A. et al. (2009). Coat color variation at the beginning of horse domestication. *Science*, *324*, 485–485.

Luengo, D., Martino, L., Bugallo, M., Elvira, V., & Särkkä, S. (2020). A survey of Monte Carlo methods for parameter estimation. *EURASIP Journal on Advances in Signal Processing*, *2020*, 1–62.

Lyu, W., Dai, X., Beaumont, M. A., Yu, F., & He, Z. (2022). Inferring the timing and strength of natural selection and gene migration in the evolution of chicken from ancient DNA data. *Molecular Ecology Resources*, *22*, 1362–1379.

Malaspinas, A.-S. (2016). Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. *Molecular Ecology*, *25*, 24–41.

Malaspinas, A.-S., Malaspinas, O., Evans, S. N., & Slatkin, M. (2012). Estimating allele age and selection coefficient from time-serial data. *Genetics*, *192*, 599–607.

Mathieson, I. (2020). Estimating time-varying selection coefficients from time series data of allele frequencies. *bioRxiv*, (p. 387761).

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N. et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*, 499–503.

Mathieson, I., & Terhorst, J. (2022). Direct detection of natural selection in Bronze Age Britain. *bioRxiv*, (p. 484330).

da Mota, B. S., Rubinacci, S., Dávalos, D. I. C., Amorim, C. E. G., Sikora, M. et al. (2022). Imputation of ancient genomes. *bioRxiv*, (p. 500636).

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*, 443–451.

Ortega-Del Vecchyo, D., & Slatkin, M. (2019). F_{ST} between archaic and present-day samples. *Heredity*, *122*, 711–718.

Pruvost, M., Bellone, R., Benecke, N., Sandoval-Castellanos, E., Cieslak, M. et al. (2011). Genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art. *Proceedings of the National Academy of Sciences*, *108*, 18626–18630.

Renaud, G., Hanghøj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics*, *212*, 587–614.

Rieder, S., Taourit, S., Mariat, D., Langlois, B., & Guérin, G. (2001). Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mammalian Genome*, *12*, 450–455.

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, *16*, 351–367.

Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J., & Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics*, *53*, 120–126.

Schraiber, J. G., Evans, S. N., & Slatkin, M. (2016). Bayesian inference of natural selection from allele frequency time series. *Genetics*, *203*, 493–511.

Shim, H., Laurent, S., Matuszewski, S., Foll, M., & Jensen, J. D. (2016). Detecting and quantifying changing selection intensities from time-sampled polymorphism data. *G3: Genes, Genomes, Genetics*, *6*, 893–904.

Silva, N. M., Rio, J., & Currat, M. (2017). Investigating population continuity with ancient DNA under a spatially explicit simulation framework. *BMC Genetics*, *18*, 1–13.

Sohail, M. S., Louie, R. H., Hong, Z., Barton, J. P., & McKay, M. R. (2022). Inferring epistasis from genetic time-series data. *Molecular Biology and Evolution*, *39*, msac199.

Sohail, M. S., Louie, R. H., McKay, M. R., & Barton, J. P. (2021). MPL resolves genetic linkage in fitness inference from complex evolutionary histories. *Nature Biotechnology*, *39*, 472–479.

Steinrücken, M., Bhaskar, A., & Song, Y. S. (2014). A novel spectral method for inferring general diploid selection from time series genetic data. *The Annals of Applied Statistics*, *8*, 2203–2222.

Tataru, P., Simonsen, M., Bataillon, T., & Hobolth, A. (2017). Statistical inference in the Wright-Fisher model using allele frequency data. *Systematic Biology*, *66*, e30–e46.

Terhorst, J., Schlötterer, C., & Song, Y. S. (2015). Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genetics*, *11*, e1005069.

Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, *22*, 997–1008.

Williamson, E. G., & Slatkin, M. (1999). Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, *152*, 755–761.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Wutke, S., Benecke, N., Sandoval-Castellanos, E., Döhle, H.-J., Friederich, S. et al. (2016). Spotted phenotypes in horses lost attractiveness in the Middle Ages. *Scientific Reports*, *6*, 38548.

Data Accessibility Statement

The authors state that all data necessary for confirming the conclusions of the present work are represented completely within the article. Source code implementing the adaptive version of the PMMH algorithm described in this work is available at `https://github.com/zhangyi-he/WFM-2L-DiffusApprox-AdaptPMMH/`, where the standard version of the PMMH algorithm is also available.

Author Contributions

Z.H. designed the project and developed the method; Z.H., X.D. and W.L. implemented the method; X.D. and W.L. analysed the data under the supervision of Z.H., M.B. and F.Y.; Z.H. wrote the manuscript; X.D., W.L., M.B. and F.Y. reviewed the manuscript.