Latent mutations in the ancestries of alleles under selection

Wai-Tong (Louis) Fan[a,b], John Wakeley[b]

[a]*Department of Mathematics, Indiana University, 831 East 3rd St, Bloomington, 47405, IN, USA*
[b]*Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Ave, Cambridge, 02138, MA, USA*

Abstract

We consider a single genetic locus with two alleles A_1 and A_2 in a large haploid population. The locus is subject to selection and two-way, or recurrent, mutation. Assuming the allele frequencies follow a Wright-Fisher diffusion and have reached stationarity, we describe the asymptotic behaviors of the conditional gene genealogy and the latent mutations of a sample with known allele counts, when the count n_1 of allele A_1 is fixed, and when either or both the sample size n and the selection strength $|\alpha|$ tend to infinity. Our study extends previous work under neutrality to the case of non-neutral rare alleles, asserting that when selection is not too strong relative to the sample size, even if it is strongly positive or strongly negative in the usual sense ($\alpha \to -\infty$ or $\alpha \to +\infty$), the number of latent mutations of the n_1 copies of allele A_1 follows the same distribution as the number of alleles in the Ewens sampling formula. On the other hand, very strong positive selection relative to the sample size leads to neutral gene genealogies with a single ancient latent mutation. We also demonstrate robustness of our asymptotic results against changing population sizes, when one of $|\alpha|$ or n is large.

Keywords: Recurrent mutation, selection, Ewens sampling formula, coalescent, Wright-Fisher diffusion

1. Introduction

The observed copies of a particular allele in a sample descend from an unknown number of distinct mutations. If k_1 is the number of these 'latent' mutations for allele A_1 when it is observed n_1 times in a sample, then $k_1 \in \{1, 2, \ldots, n_1\}$. Although latent mutations are not observed directly, they can be modeled as outcomes of the stochastic ancestral process of a sample and inferred from patterns of variation in DNA data (Harpak et al., 2016; Seplyarskiy et al., 2021; Johnson et al., 2022). Analytical results on the distribution and timing of latent mutations of rare neutral alleles are given in Wakeley et al. (2023). Here we consider non-neutral alleles which may be under strong selection and which may or may not be rare. We take two different approaches to modeling latent mutations under selection and recurrent mutation. The first approach uses the idea of coalescence in a random background of allele frequencies in the population (Barton et al., 2004). The second uses the conditional ancestral selection graph (Slade, 2000a) and demonstrates results consistent with those from the first approach.

Email addresses: `waifan@iu.edu` (Wai-Tong (Louis) Fan), `wakeley@fas.harvard.edu` (John Wakeley)

Wakeley et al. (2023) also contains an application to the frequencies of single-nucleotide sites with counts $n_1 \in \{1, 2, \ldots, 40\}$ of synonymous mutations in a subsample of 57K non-Finnish European individuals ($n = 114$K) from the *gnomAD* database (Karczewski et al., 2020). Dramatic differences in sample frequency distributions of rare alleles with different mutation rates, categorized by the 'Roulette' method of Seplyarskiy et al. (2023), were well explained by an empirical demographic model with recurrent mutation but no selection. Seplyarskiy et al. (2023, Fig. 3a) showed using simulations that a neutral, parametric demographic model fitted to these data also explained the frequencies of mutation in counts $n_1 \leq 10^4$. Polymorphic sites with small mutation counts comprise the bulk of variation in humans. They represent a rich source of information about demographic history and possibly selection. Sites with $n_1 \in \{1, 2, \ldots, 40\}$ make up about 95% of all polymorphic sites in the *gnomAD* data used in Wakeley et al. (2023).

Previous theoretical work on populations of constant size has shown that distributions of rare alleles are in fact unaffected even by moderately strong selection (Joyce and Tavaré, 1995; Joyce, 1995). Specifically, the counts of latent mutations obey the independent Poisson statistics of rare alleles in the Ewens sampling formula (Ewens, 1972; Arratia et al., 1992, 2003). This is also the case in Wakeley et al. (2023) when the population size is constant. In the present work we investigate the robustness of these results to very strong selection. Theory also predicts that rare alleles tend to be young (Kimura and Ohta, 1973; Watterson, 1976). Mathieson and McVean (2014) and Platt et al. (2019) have demonstrated empirically that rare non-synonymous or otherwise functional alleles in the human genome are even younger than non-functional rare alleles. In the present work we also investigate how strong selection and rarity affect the ages of latent mutations.

We assume there are two possible alleles, A_1 and A_2, at a single genetic locus in a large haploid population. We begin by assuming that the population size N is constant over time. In Section 3.4 we consider time-varying population size. One allele or the other is favored by directional selection. Mutation is recurrent and happens in both directions. In the diffusion approximation, time is measured in proportion to N_e generations where N_e is the effective population size (Ewens, 2004). Under the Wright-Fisher model of reproduction, $N_e = N$. Under the Moran model of reproduction (Moran, 1958, 1962), $N_e = N/2$. With these assumptions, the frequency of A_1 alleles is well approximated by a process X that solves (1) below and has parameters θ_1, θ_2 and α as $N \to \infty$. For a haploid population, $\theta_i = 2N_e u_i$ and $\alpha = 2N_e s$, in which u_i is the per-generation rate of $A_{3-i} \to A_i$ mutations and s is the selection coefficient. If there is no dominance, these results can be extended to diploids, in which case $\theta_i = 4N_e u_i$ and $\alpha = 4N_e s$.

Thus, we assume that allele-frequency dynamics in the population obey the Wright-Fisher diffusion (Fisher, 1930; Wright, 1931; Ewens, 2004) with parameters θ_1 and θ_2 for mutations $A_2 \to A_1$ and $A_1 \to A_2$, respectively, and α for the selective advantage (if $\alpha > 0$) or disadvantage (if $\alpha < 0$) of allele A_1. That is, we let $X(t)$ be the relative frequency of A_1 in the population at time t, and assume that its forward-time dynamics is described by the stochastic differential equation

$$dX(t) = \left[\frac{\theta_1}{2}(1 - X(t)) - \frac{\theta_2}{2}X(t) + \frac{\alpha}{2}X(t)(1 - X(t))\right] dt + \sqrt{X(t)(1 - X(t))}\, dW_t, \quad t > 0 \qquad (1)$$

in which W_t is the Wiener process, also called the standard Brownian motion.

Both of the approaches (random background and ancestral selection graph) we take to modeling latent mutations rely on the assumption that the population has reached equilibrium, which occurs in the limit $t \to \infty$. The stationary probability density of X is

$$\phi_\alpha(x) = C x^{\theta_1 - 1}(1 - x)^{\theta_2 - 1} e^{\alpha x}, \quad 0 < x < 1 \qquad (2)$$

(Wright, 1931; Ewens, 2004). We explicitly denote the dependence on α because this parameter plays a key role in what follows. The normalizing constant C guarantees that $\int_0^1 \phi_\alpha(x)dx = 1$. It is given by

$$C = \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2){}_1F_1(\theta_1;\theta_1+\theta_2;\alpha)} \tag{3}$$

in which $\Gamma(a)$ is the gamma function and ${}_1F_1(a;b;z)$ is the confluent hypergeometric function, or Kummer's function; see Abramowitz and Stegun (1964) and Slater (1960).

By definition, latent mutations occur in the ancestry of a sample. When a sample of total size n is taken from a population with stationary density (2), it will contain a random number $\mathcal{N}_1$ of copies of allele A_1 and $\mathcal{N}_2 = n - \mathcal{N}_1$ copies of allele A_2. The probability that $\mathcal{N}_1$ is equal to n_1 is equal to

$$\begin{aligned} q(n_1,n_2) &:= \mathbb{P}(\mathcal{N}_1 = n_1; n, \alpha, \theta_1, \theta_2) \\ &= \int_0^1 \binom{n}{n_1} x^{n_1}(1-x)^{n-n_1}\phi_\alpha(x)dx \\ &= C\binom{n_1+n_2}{n_1}\frac{\Gamma(\theta_1+n_1)\Gamma(\theta_2+n_2)}{\Gamma(\theta_1+\theta_2+n_1+n_2)}{}_1F_1(\theta_1+n_1;\theta_1+\theta_2+n_1+n_2;\alpha) \end{aligned} \tag{4}$$

for $n_1 \in \{0,1,\ldots,n\}$ and $n_2 = n - n_1$, and with C again given by (3). The notation $q(\cdot)$ is from Slade (2000a,b) and is convenient for the ancestral selection graph.

Suppose now we are given the sample count, that is, we know that among the n uniformly sampled haploid individuals, n_1 of them are of type 1 and the remaining $n_2 = n - n_1$ are of type 2. Then the posterior density of the population frequency of A_1 conditional on the sample is

$$\phi_\alpha^{(n_1,n_2)}(x) = \frac{\binom{n_1+n_2}{n_1}x^{n_1}(1-x)^{n_2}\phi_\alpha(x)}{q(n_1,n_2)} \tag{5}$$

$$= \frac{\Gamma(\theta_1+\theta_2+n_1+n_2)x^{\theta_1+n_1-1}(1-x)^{\theta_2+n_2-1}e^{\alpha x}}{\Gamma(\theta_1+n_1)\Gamma(\theta_2+n_2){}_1F_1(\theta_1+n_1;\theta_1+\theta_2+n_1+n_2;\alpha)} \tag{6}$$

from Bayes' theorem with prior density ϕ_α.

The sampling probability $q(n_1,n_2)$ in (4) and the resulting posterior density $\phi_\alpha^{(n_1,n_2)}$ play major roles in the two approaches we take to modeling latent mutations. Specifically, transition probabilities in the conditional ancestral selection graph depend on ratios of sampling probabilities (Slade, 2000b) and the allele frequency in the ancestral process of Barton et al. (2004) has initial density $\phi_\alpha^{(n_1,n_2)}(x)$ when conditioned on the sample count.

We describe the occurrence of latent mutations in the ancestry of allele A_1 conditional on the sample count n_1. We say that A_1 is *rare* when the sample size n is much larger than n_1. We enforce this rarity of A_1 by letting $n_2 \sim n$ tend to infinity with n_1 fixed, or finite. We present some results for cases in which A_1 is not rare in this sense, that is when neither n_1 nor n_2 is large. In this case we also describe the conditional ancestry of A_2, but overall our focus is on large samples and rare A_1. This is the same, sample-based concept of rarity that was used in Wakeley et al. (2023) and previously considered by Joyce and Tavaré (1995) and Joyce (1995). It may be distinguished from rarity in the population, though of course finding A_1 rare in a large sample is most likely when the population frequency x is small.

By *strong selection* we mean large $|\alpha|$. We model rarity and strong selection together under the assumption that $\alpha = \widetilde{\alpha} n_2$ for some constant $\widetilde{\alpha} \in \mathbb{R}$. We study latent mutations and the ancestral

processes which generate them under three scenarios: (i) $|\alpha|$ large with n_2 fixed, (ii) n_2 large with α fixed, and (iii) both $|\alpha|$ and n_2 large with $\widetilde{\alpha} = \alpha/n_2$ fixed. In making approximations for large n_2 and/or large $|\alpha|$, we make extensive use of asymptotic results for ratios of gamma functions and for the confluent hypergeometric function which are presented in Appendix A.

The parameters θ_1 and θ_2 are fixed constants throughout, with $\theta_1, \theta_2 > 0$. For single nucleotide sites, these population-scaled mutation rates have been estimated from pairwise sequence differences for many species and are typically about 0.01 with a range of about 0.0001 to 0.1 (Leffler et al., 2012). Values for humans are smaller but they vary almost as widely among sites in the genome, with a mean of about 0.0008 and a range of about 0.0001 to 0.02 (Seplyarskiy et al., 2021, 2023; Wakeley et al., 2023). In contrast, there is no reason to suppose that the selection parameter $|\alpha|$ is small (Eyre-Walker and Keightley, 2007; Chen et al., 2020). Note that our introduction of a constant $\widetilde{\alpha} = \alpha/n_2$ is simply a device to specify the relative importance of rarity as opposed strong selection, not a hypothesis about biology.

The case of a rare neutral allele was considered in Wakeley et al. (2023) where it was shown that the number of latent mutations in the ancestry of the n_1 copies of allele A_1 follows the same distribution as the number of alleles in the Ewens sampling formula (Ewens, 1972) with sample size n_1 and mutation parameter θ_1. Let K_1 be the random number of these latent mutations for allele A_1 in the ancestry of the sample. Then if ξ_j is a Bernoulli random variable with probability of success

$$\mathbb{P}(\xi_j = 1) = \frac{\theta_1}{\theta_1 + j - 1} \quad , \quad j = 1, 2, \ldots \tag{7}$$

then under neutrality for large sample size and conditional on $\mathcal{N}_1 = n_1$,

$$K_1 \stackrel{d}{=} \xi_{n_1} + \xi_{n_1 - 1} + \cdots + \xi_2 + \xi_1 \tag{8}$$

which gives the stated Ewens sampling result (Arratia et al., 1992). In (8) and below, $\stackrel{d}{=}$ denotes equal in distribution. Note that, because coalescence is among exchangeable lineages, the full Ewens sampling formula should apply if we were to keep track of the sizes of latent mutations; see Crane (2016) and Tavaré (2021) for recent reviews.

Here we apply the model of coalescence in a random background described by Barton et al. (2004) to prove these results (7) and (8) for rare alleles in large samples and especially to extend the analysis of latent mutations to scenarios which include selection. We investigate both the number of latent mutations and their timing in the ancestry of the sample, and we allow that selection may be strong. We also show how the same scenarios can be treated using the conditional ancestral selection graph (Slade, 2000a), giving the same limiting results for all three scenarios.

Briefly, we find that positive selection does not in general lead to (7) and (8), that very strong positive selection (relative to the sample size) leads to neutral gene genealogies with a single ancient latent mutation for the favored allele. This is described in Section 3 for scenario (i) and for the case $\widetilde{\alpha} \in (1, \infty)$ in scenario (iii). On other hand, when selection is not too strong relative to the sample size, then extreme rarity of A_1 in the sample can effectively override strong positive selection and retrieve (7) and (8). This is described in Section 3 for scenario (ii) and for the case $\widetilde{\alpha} \in (-\infty, 1)$ in scenario (iii). Figures 1, 2 and 3 illustrate our results in the three scenarios.

2. Sample frequencies and posterior population frequencies

In this section, we present asymptotic results for the sampling probability $q(n_1, n_2)$ in (4) and the posterior density $\phi_\alpha^{(n_1,n_2)}(x)$ in our three regimes of interest: (i) $|\alpha|$ large with n_2 fixed, (ii) n_2

large with α fixed, and (iii) both $|\alpha|$ and n_2 large with $\widetilde{\alpha} = \alpha/n_2$ fixed.

2.1. Asymptotics for sampling probabilities

In the case of strong selection and moderate sample size, that is $|\alpha|$ large with n_2 fixed, applying (A.4a) and (A.4b) to (4) gives

$$q(n_1, n_2) = \begin{cases} \binom{n}{n_1} \dfrac{\Gamma(\theta_1 + n_1)}{\Gamma(\theta_1)} |\alpha|^{-n_1} \left(1 + O\left(|\alpha|^{-1}\right)\right) & \text{if } \alpha < 0, \quad (9a) \\ \binom{n}{n_1} \dfrac{\Gamma(\theta_2 + n - n_1)}{\Gamma(\theta_2)} \alpha^{n_1 - n} \left(1 + O\left(\alpha^{-1}\right)\right) & \text{if } \alpha > 0. \quad (9b) \end{cases}$$

Here we focus on the leading-order terms but note that the next-order terms are straightforward to obtain using (A.4a) and (A.4b) and additional terms could be computed using (4.1.2) and (4.1.6) in Slater (1960). In (9a), each additional copy of A_1 decreases the sampling probability by a factor of $1/|\alpha|$ so the most likely sample is one which contains no copies of A_1. In (9b), each additional copy of A_1 increases the sampling probability by a factor of α so the most likely sample is monomorphic for A_1. However, these results are perfectly symmetric for the two alleles. Switching allelic labels and swapping $|\alpha|$ for α changes (9a) into (9b). That is, allele A_2 experiences the same effects of positive/negative selection in (9a)/(9b) as the focal allele A_1 does in (9b)/(9a).

In the case of large sample size and moderate selection, that is n_2 large with α fixed, applying (A.5) to (4) gives

$$q(n_1, n_2) = C \frac{\Gamma(\theta_1 + n_1)}{n_1!} n_2^{-\theta_1} \left(1 + O\left(n_2^{-1}\right)\right). \quad (10)$$

This has the same form as the neutral result, equation (22) in Wakeley et al. (2023), only with the additional factor $1/{}_1F_1(\theta_1; \theta_1 + \theta_2; \alpha)$ in the constant C. With respect to the count of the focal allele A_1, the distribution is similar to a (degenerate) negative-binomial distribution with parameters $p = 1/n_2$ and $r = \theta_1$, like the corresponding result in Theorem 2 of Watterson (1974) for neutral alleles which propagate by a linear birth-death process. The effect of selection is only to uniformly raise or lower the chances of seeing n_1 copies of A_1 in a very large sample. The additional factor $1/{}_1F_1(\theta_1; \theta_1 + \theta_2; \alpha)$ is a decreasing function of α, which is equal to 1 when $\alpha = 0$ and approaches 0 quickly from there as α increases. Greater selection against (respectively, for) A_1 increases (respectively, decreases) the chance of it being rare but does not affect the shape of the distribution of n_1, at least to leading order in $1/n_2$.

In the case of large sample size and strong selection, that is both $|\alpha|$ and n_2 large with $\widetilde{\alpha} = \alpha/n_2$ fixed, applying (A.4a), (A.4b), (A.6a), (A.6b) to (4) gives

$$q(n_1, n_2) = \begin{cases} \dfrac{\Gamma(\theta_1 + n_1)}{n_1! \Gamma(\theta_1)} \left(\dfrac{1}{1 + |\widetilde{\alpha}|}\right)^{n_1} \left(\dfrac{|\widetilde{\alpha}|}{1 + |\widetilde{\alpha}|}\right)^{\theta_1} \left(1 + O\left(n_2^{-1}\right)\right) & \text{if } \widetilde{\alpha} < 0 \quad (11a) \\ B_1 \dfrac{\Gamma(\theta_1 + n_1)}{n_1!} \left(\dfrac{1}{1 - \widetilde{\alpha}}\right)^{n_1} \left(1 + O\left(n_2^{-1}\right)\right) & \text{if } 0 < \widetilde{\alpha} < 1 \quad (11b) \\ B_2 \dfrac{1}{n_1!} \left(\dfrac{\widetilde{\alpha} - 1}{\widetilde{\alpha}} n_2\right)^{n_1} \left(1 + O\left(n_2^{-1}\right)\right) & \text{if } \widetilde{\alpha} > 1 \quad (11c) \end{cases}$$

with constants B_1 and B_2 which are unremarkable except in their dependence on n_2:

$$B_1 \propto n_2^{\theta_2-\theta_1} e^{-\widetilde{\alpha} n_2}$$

$$B_2 \propto n_2^{\theta_2-\frac{1}{2}} (\widetilde{\alpha} e)^{-n_2}$$

such that $q(n_1, n_2)$ becomes tiny as n_2 grows. In (11b) and (11c), allele A_1 is favored by selection so it will be unlikely for its sample count to be very small.

To see how (11b) and (11c) compare to (11a), consider how these three sampling probabilities change as n_1 increases:

$$\frac{q(n_1+1, n_2)}{q(n_1, n_2)} \approx \begin{cases} \dfrac{\theta_1 + n_1}{(1+|\widetilde{\alpha}|)(n_1+1)} & \text{if } \widetilde{\alpha} < 0 \quad (12a) \\[2ex] \dfrac{\theta_1 + n_1}{(1-\widetilde{\alpha})(n_1+1)} & \text{if } 0 < \widetilde{\alpha} < 1 \quad (12b) \\[2ex] \dfrac{(\widetilde{\alpha}-1) n_2}{\widetilde{\alpha}(n_1+1)} & \text{if } \widetilde{\alpha} > 1 \quad (12c) \end{cases}$$

where the approximation is for large n_2, i.e. omitting the $O\big(n_2^{-1}\big)$ parts of (11a), (11b) and (11c). The first two differ from the corresponding neutral result $(\theta_1 + n_1)/(n_1 + 1)$ by the constant factors $1/(1 + |\widetilde{\alpha}|) < 1$ in (12a) and $1/(1 - \widetilde{\alpha}) > 1$ in (12b). Note that (10) gives the neutral result, as do (12a) and (12b) as $\widetilde{\alpha} \to 0$. Relative to this, negative selection in (12a) makes additional copies of A_1 less probable whereas positive selection in (12b) makes them more probable. But (12a) and (12b) differ from the neutral result only by these constant factors. Equation (12c) is quite different. With $\widetilde{\alpha} > 1$, each additional copy of A_1 increases the sampling probability by a large factor, proportional to n_2, making this case similar to the case of strong positive selection in (9b). This is as expected. What is surprising is (12b), namely that strong selection ($\alpha \to \infty$) in favor of A_1 can be made to resemble neutrality simply by increasing the sample size relative to n_1.

2.1.1. Comparison to discrete Moran and Wright-Fisher models

We emphasize that our analyses in this work are of the Wright-Fisher diffusion model, given here as the SDE (1) with stationary density (2). It is of interest to know how well our results hold for discrete, exact models such as the Moran model and the Wright-Fisher model, especially as $|\alpha| \to \infty$ or $n \to \infty$ for finite n_1, in which cases we might expect the diffusion to be a relatively poor description of the dynamics. In this section, we focus on (11a) and show that it can be obtained in a different way from a discrete-time Moran model, without first passing to the diffusion limit, but that this cannot be done in general starting from the discrete-time Wright-Fisher model.

To leading order in $1/n_2$, (11a) is identical to the probability mass function of a negative binomial distribution with parameters $p = |\widetilde{\alpha}|/(1 + |\widetilde{\alpha}|)$ and $r = \theta_1$. Charlesworth and Hill (2019) obtained this same result using the strong-selection diffusion approximation of Nei (1968). Here selection *against* A_1 is so strong that it never reaches appreciable frequency in the population. In the limit, or ignoring the $O\big(n_2^{-1}\big)$ part in (11a), this distribution sums to one over all $n_1 \in \{0, 1, 2, \ldots\}$. The corresponding sum for the degenerate distribution in (10) diverges, because under neutrality there is a non-trivial chance that A_1 reaches appreciable frequency in the population.

Consider a discrete-time haploid Moran model with population size N, in which allele A_2 is favored by selection. Specifically, A_1 and A_2 have equal chances of being chosen to reproduce but

different chances of being chosen to die: each A_1 has an increased chance $1+s$ compared to each A_2. Upon reproduction, the offspring of an A_i, $i \in \{1,2\}$, has type A_i with probability $1-u_{3-i}$ and the other type A_{3-i} with probability u_{3-i}. If there are currently ℓ copies of A_1 and $N-\ell$ copies of A_2, then in the next time step there will $\ell+1$ copies of A_1 with probability

$$\frac{N-\ell}{N-\ell+\ell(1+s)}\frac{\ell}{N}(1-u_2)+\frac{N-\ell}{N-\ell+\ell(1+s)}\frac{N-\ell}{N}u_1 \tag{13}$$

and $\ell-1$ copies of A_1 with probability

$$\frac{\ell(1+s)}{N-\ell+\ell(1+s)}\frac{N-\ell}{N}(1-u_1)+\frac{\ell(1+s)}{N-\ell+\ell(1+s)}\frac{\ell}{N}u_2. \tag{14}$$

The fraction of A_1 converges to the Wright-Fisher diffusion process (1) as $N \to \infty$ if time is measured in units of $N(N-1)/2$ discrete steps, i.e. $dt = 2/N(N-1)$, with $u_1 = \theta_1/N$, $u_2 = \theta_2/N$ and $s = -\alpha/N$.

As another way of obtaining (11a), we assume that $s \gg u_1, u_2$. In particular, let $Nu_1 \to \theta_1$ and $Nu_2 \to \theta_2$ as $N \to \infty$ just as in the diffusion model, but let s be a constant. Then we may appeal to the analogous model and limit process (iii) of Karlin and McGregor (1964) which had no selection but instead assumed that $u_2 \gg u_1$. Similarly here we expect that allele A_1 will be held in negligible relative frequency in the population and instead be present in a finite number of copies as $N \to \infty$, only here due to strong selection rather than strong mutation.

In view of this scaling of the mutation rates by N and for comparison with the Wright-Fisher model below, we rescale time so that it is measured in unit of generations, or N time steps in this Moran model. Then with $dt = 1/N$, we can rewrite (13) and (14) as

$$\left(\ell\frac{N-\ell}{N-\ell+\ell(1+s)}(1-u_2)+\frac{N-\ell}{N-\ell+\ell(1+s)}(N-\ell)u_1\right)dt,$$

and

$$\left(\ell(1+s)\frac{N-\ell}{N-\ell+\ell(1+s)}(1-u_1)+\frac{\ell(1+s)}{N-\ell+\ell(1+s)}\ell u_2\right)dt.$$

In the limit $N \to \infty$ this converges to a continuous-time process for A_1 in which in a small time dt

$$\ell \to \begin{cases} \ell+1 & \text{with probability } \ (\ell+\theta_1)dt \\ \ell-1 & \text{with probability } \ \ell(1+s)dt \end{cases} \tag{15}$$

in other words to a birth-death process with immigration where the birth rate is $\lambda = 1$, the death rate is $\mu = 1+s$ and the immigration rate is $\kappa = \theta_1$. Then from (52) in Kendall (1949), the distribution of the number of copies of A_1 in the population at stationarity will be negative binomial with parameters $1-\lambda/\mu = s/(1+s)$ and $\kappa/\lambda = \theta_1$, or

$$p(\ell) = \binom{\ell+\theta_1-1}{\ell}\left(\frac{1}{1+s}\right)^{\ell}\left(\frac{s}{1+s}\right)^{\theta_1}, \qquad \ell \in \mathbb{Z}_+. \tag{16}$$

In getting to (11a) in the main text, which we note is for $\alpha < 0$, we first applied the diffusion limit then let the selection parameter $|\alpha|$ be large, specifically proportional to the number n_2 of copies of A_2 in a sample of large size n for a given fixed number n_1 of copies of A_1 in the

sample. In the current haploid Moran model with selection against A_1, $|\alpha| = Ns$, and the scalar $|\widetilde{\alpha}| = |\alpha|/n_2 = Ns/n_2$. Define $a := n_2/N$. Then $|\widetilde{\alpha}| = s/a$ and we may think of a as (close to) the proportion of the population sampled, because $n_2/N \sim n/N$.

If the number of copies of A_1 in the population is ℓ, then the probability there are n_1 copies in a sample of size n taken without replacement from the total population of size N is given by the hypergeometric distribution

$$p(n_1|\ell; N) = \frac{\binom{\ell}{n_1}\binom{N-\ell}{n-n_1}}{\binom{N}{n}}, \qquad n_1 = 0, 1, 2, \ldots, \ell. \tag{17}$$

Since $n - n_1 = n_2 = aN$ and taking $N \to \infty$, (17) converges to the binomial distribution

$$p(n_1|\ell) = \binom{\ell}{n_1} a^{n_1}(1-a)^{\ell-n_1}, \qquad n_1 = 0, 1, 2, \ldots, \ell. \tag{18}$$

This gives another route to (11a), namely using (16) and (18), and putting $y = \ell - n_1$,

$$\begin{aligned} p(n_1) &= \sum_{\ell=n_1}^{\infty} p(n_1|\ell)p(\ell) \\ &= \sum_{\ell=n_1}^{\infty} \binom{\ell}{n_1} a^{n_1}(1-a)^{\ell-n_1} \binom{\ell+\theta_1-1}{\ell} \left(\frac{1}{1+s}\right)^{\ell} \left(\frac{s}{1+s}\right)^{\theta_1} \\ &= \frac{1}{n_1!}\left(\frac{a}{1+s}\right)^{n_1}\left(\frac{s}{a+s}\right)^{\theta_1} \sum_{y=0}^{\infty} \frac{\Gamma(y+n_1+\theta_1)}{\Gamma(y+\theta_1)} \binom{y+\theta_1-1}{y} \left(\frac{1-a}{1+s}\right)^{y} \left(\frac{a+s}{1+s}\right)^{\theta_1} \\ &= \frac{\Gamma(n_1+\theta_1)}{n_1!\Gamma(\theta_1)} \left(\frac{a}{a+s}\right)^{n_1} \left(\frac{s}{a+s}\right)^{\theta_1} \end{aligned} \tag{19}$$

the end result (19) being equal to the leading order part of (11a) since $|\widetilde{\alpha}| = s/a$.

We can contrast this with a discrete-time haploid Wright-Fisher model with population size N, in which time is already measured in generations. Under the same assumptions that gave (15), namely s constant and $u_1, u_2 \propto 1/N$ as $N \to \infty$, we can use equation (33) in Nagylaki (1990) which specifies that, conditional on the number ℓ_g of copies of A_1 in generation g, the number ℓ_{g+1} has the Poisson distribution

$$\ell_{g+1}|\ell_g \sim \text{Poisson}\left(\frac{\theta_1}{2} + \ell_g(1-s)\right) \tag{20}$$

with $\theta_1 := 2Nu_1$. Now ℓ evolves by a Poisson branching process with Poisson immigration rather than by the birth-death process with immigration in (15). Although here too the number of copies of A_1 in the population will converge to a stationary distribution (Heathcote, 1965), it will not in general be a negative binomial distribution. Thus (11a) is consistent with the per-generation dynamics of rare alleles in the Moran model but not in the Wright-Fisher model.

2.2. Asymptotics for population frequencies conditional on the sample

Next, we obtain asymptotics for the posterior probability density $\phi_\alpha^{(n_1,n_2)}$ in (6). Let $\mathcal{P}([0,1])$ be the space of probability measures on $[0,1]$ endowed with the weak convergence topology (i.e with test functions in the space $C_b([0,1])$ of bounded continuous functions on $[0,1]$).

Lemma 1. *Let $n_1 \in \mathbb{N}$ be fixed. The following convergences in $\mathcal{P}([0,1])$ hold.*

(i) Suppose $n_2 \in \mathbb{N}$ is fixed and $\alpha \to \infty$. Then $\phi_\alpha^{(n_1,n_2)}(x)\,dx \to \delta_1$.

(ii) Suppose $\alpha \in \mathbb{R}$ is fixed and $n_2 \to \infty$. Then $\phi_\alpha^{(n_1,n_2)}(x)\,dx \to \delta_0$.

(iii) Suppose $\alpha = \widetilde{\alpha} n_2 + c$ where $\widetilde{\alpha}$, $c \in \mathbb{R}$ are fixed and $n_2 \to \infty$. Then

$$\phi_\alpha^{(n_1,n_2)}(x)\,dx \to \begin{cases} \delta_0 & \text{when} \quad \widetilde{\alpha} \in (-\infty, 1] \\ \delta_{1-1/\widetilde{\alpha}} & \text{when} \quad \widetilde{\alpha} \in (1, \infty) \end{cases} \tag{21}$$

where δ_x is the Dirac delta measure.

The proof of Lemma 1 is given in Appendix B.

3. Conditional coalescence in a random background

In this section, we extend the approach of coalescence in a random background in Barton et al. (2004) to study the number and timing of latent mutations and other asymptotic properties of the conditional gene genealogy given the sample frequencies of A_1 and A_2. We also extend our results to time-varying populations in Section 3.4. While the setting of Barton et al. (2004) covers the case of two linked loci (one selected, the other neutral) here we focus on one selected locus.

Suppose we are given a sample from the selected locus at the present time $t = 0$, and that we know the allelic types of the sample but we do not know how the sample was produced. What is the genealogy of the sample? This question was answered by Barton et al. (2004), where they modeled the ancestral process using the structured coalescent with allelic types as subpopulations. The structured coalescent can be a model of subdivision with migration between local populations (Takahata, 1988; Notohara, 1990; Herbots, 1997) or a model of selection with mutation between allelic types (Kaplan et al., 1988; Darden et al., 1989). For samples from a population at stationarity as in Section 1, Barton et al. (2004) proved that this could be done rigorously starting with a Moran model with finite N then passing to the diffusion limit. Barton and Etheridge (2004) explored some properties of gene genealogies under this model, and Etheridge et al. (2006) used the same idea to describe genetic ancestries following a selective sweep.

Even if the sample frequencies are known, the allele frequencies in the population are unknown. A key feature of this method is to model allele-frequency trajectories backward in time. As pointed out by Barton et al. (2004), the Moran model with finite N is *reversible*, meaning that under its stationary probability density, the time-reversed process is the same (in distribution) as the forward-time Moran process. This is not a property of the Wright-Fisher model with finite N but does hold for their shared diffusion limit (1) with stationary density (2); see for instance Millet et al. (1989) for why this holds. Figure 1 gives an illustration of a genealogy with mutations and allele frequencies varying backward in time.

Let $n_i^{(N)}(t)$ be the number of ancestral lineages of type i at time t backward, and let $p_t^{(N)}$ be the fraction of type 1 in the population at time t backward. By (Barton et al., 2004, Lemma 2.4), under the Moran model with stationary distribution, $(p_t^{(N)}, n_1^{(N)}(t), n_2^{(N)}(t))_{t\in\mathbb{R}_+}$ is a Markov process for each fixed N. Furthermore, (Barton et al., 2004, Lemma 3.1) describes the joint convergence of the processes $(p^{(N)}, n_1^{(N)}, n_2^{(N)})$ as $N \to \infty$.

Lemma 2 (Lemmas 2.4 and 3.1 of Barton et al. (2004))**.** *Let $p_t^{(N)}$ and $n_i^{(N)}(t)$ be the fraction of type 1 in the population and the number of ancestral lineages of type i respectively, at time t backward, under the stationary Moran model. Then $(p^{(N)}, n_1^{(N)}, n_2^{(N)})$ is a Markov process for each $N \in \mathbb{N}$. As $N \to \infty$, this process converges in distribution to a Markov process $(p_t, n_1(t), n_2(t))_{t\in\mathbb{R}_+}$ with state space $[0,1] \times \mathbb{Z}_+ \times \mathbb{Z}_+$. This limiting process is described as follows:*

(i) $(p_t)_{t\in\mathbb{R}_+}$ is a solution to equation (1) with stationary initial density ϕ_α. In particular, it does not depend on $(n_1(t), n_2(t))_{t\in\mathbb{R}_+}$.

(ii) Suppose the current state is (p, m_1, m_2). Then $(n_1(t), n_2(t))_{t\in\mathbb{R}_+}$ evolves as

$$(m_1, m_2) \to \begin{cases} (m_1-1, m_2) & \text{at rate } \frac{1}{p}\binom{m_1}{2} & \text{coalescence of type 1} \\ (m_1-1, m_2+1) & \text{at rate } \frac{1-p}{p}\, m_1\, \frac{\theta_1}{2} & \text{mutation of 1 to 2} \\ (m_1, m_2-1) & \text{at rate } \frac{1}{1-p}\binom{m_2}{2} & \text{coalescence of type 2} \\ (m_1+1, m_2-1) & \text{at rate } \frac{p}{1-p}\, m_2\, \frac{\theta_2}{2} & \text{mutation of 2 to 1} \end{cases} \tag{22}$$

We compare the notation here with that of Barton et al. (2004) and Etheridge (2011). Type 1 here is their type P, type 2 here is their type Q. So, $\frac{\theta_2}{2}$ (resp. $\frac{\theta_1}{2}$) here is μ_1 (resp. μ_2) in Barton et al. (2004), and is v_1 (resp. v_2) in (Etheridge, 2011, eqn. (2.11)). The rescaled selection coefficient $\frac{\alpha}{2}$ here is s in Barton et al. (2004). The Moran model in Barton et al. (2004) has population size $2N$ and each of the $(2N)(2N-1)/2$ unordered pairs is picked to interact (one dies and immediately the other reproduces) at rate $1/2$. Therefore, the diffusion limit in (Barton et al., 2004, Lemma 3.1) is slower by a factor of $1/2$ than the limit we consider here in this paper.

The proof of (Barton et al., 2004, Lemma 3.1) leads to more information for the limiting process. Let $n_i^{(N),obs}(t)$ be the number of type i lineages, among the $n_i^{(N)}(0)$ lineages that were present (thus observed) at $t = 0$, that remain at time t backward. Thus $n_i^{(N),obs}(t)$ is non-increasing in t, but $n_i^{(N)}(t)$ can increase as t increases due to mutations from type $3-i$ to type i. Clearly, $n_i^{(N),obs}(0) = n_i^{(N)}(0)$. Note that a mutation from type $3-i$ to type i backward in time corresponds to a mutation from type i to type $3-i$ forward in time. To keep track of the number of mutation events versus the number of coalescent events, we let $L_i^{(N)}(t)$ be the total number of latent mutations for type i during $[0,t]$ backward in time. We have the following generalization of Lemma 2.

Lemma 3 (Joint convergence)**.** *Under the stationary Moran model, the backward process*

$$(p^{(N)}; n_1^{(N)}, n_1^{(N),obs}, L_1^{(N)}; n_2^{(N)}, n_2^{(N),obs}, L_2^{(N)})$$

is a Markov process for each fixed $N \in \mathbb{N}$. As $N \to \infty$, this process converges in distribution to a continuous-time Markov process $(p_t; n_1(t), n_1^{obs}(t), L_1(t); n_2(t), n_2^{obs}(t), L_2(t))_{t\in\mathbb{R}_+}$ with state space $[0,1] \times \mathbb{Z}_+^6$ such that

(i) $(p_t)_{t\in\mathbb{R}_+}$ is a solution to equation (1) with stationary initial density ϕ_α. In particular, it does not depend on $(n_1(t), n_1^{obs}(t), L_1(t); n_2(t), n_2^{obs}(t), L_2(t))_{t\in\mathbb{R}_+}$.

(ii) At state $(p;\, m_1, a_1, \ell_1;\; m_2, a_2, \ell_2)$*, the process* $(n_1(t), n_1^{obs}(t), L_1(t);\, n_2(t), n_2^{obs}(t), L_2(t))_{t\in\mathbb{R}_+}$ *evolves as* $(m_1, a_1, \ell_1;\; m_2, a_2, \ell_2) \to$

$$\begin{cases} (m_1-1, a_1-1, \ell_1;\; m_2, a_2, \ell_2) & \textit{at rate} \quad \frac{1}{p}\binom{a_1}{2} & \textit{coalescence of two type } 1^{obs} \\ (m_1-1, a_1, \ell_1;\; m_2, a_2, \ell_2) & \textit{at rate} \quad \frac{1}{p}\left[\binom{m_1}{2}-\binom{a_1}{2}\right] & \textit{other coalescence of type 1} \\ (m_1-1, a_1-1, \ell_1+1;\; m_2+1, a_2, \ell_2) & \textit{at rate} \quad \frac{1-p}{p}\, a_1\, \frac{\theta_1}{2} & \textit{mutation of } 1^{obs} \\ (m_1-1, a_1, \ell_1;\; m_2+1, a_2, \ell_2) & \textit{at rate} \quad \frac{1-p}{p}\,(m_1-a_1)\, \frac{\theta_1}{2} & \textit{mutation of other type 1} \end{cases}$$

and, similarly,

$$\begin{cases} (m_1, a_1, \ell_1;\; m_2-1, a_2-1, \ell_2) & \textit{at rate} \quad \frac{1}{1-p}\binom{a_2}{2} & \textit{coalescence of two type } 2^{obs} \\ (m_1, a_1, \ell_1;\; m_2-1, a_2, \ell_2) & \textit{at rate} \quad \frac{1}{1-p}\left[\binom{m_2}{2}-\binom{a_2}{2}\right] & \textit{other coalescence of type 2} \\ (m_1+1, a_1, \ell_1;\; m_2-1, a_2-1, \ell_2+1) & \textit{at rate} \quad \frac{p}{1-p}\, a_2\, \frac{\theta_2}{2} & \textit{mutation of type } 2^{obs} \\ (m_1+1, a_1, \ell_1;\; m_2-1, a_2, \ell_2) & \textit{at rate} \quad \frac{p}{1-p}\,(m_2-a_2)\, \frac{\theta_2}{2} & \textit{mutation of other type 2} \end{cases}$$

Now suppose that in addition to knowing the sample counts n_1 and n_2, we also know that these are the outcome of uniform random sampling, as in (4). Let $\mathbb{P}_{\mathbf{n}}$ be the conditional probability measure of the ancestral process in Lemma 3 including both p_t and the lineage dynamics, given that a uniformly picked sample has allelic counts $\mathbf{n} = (n_1, n_2)$. Under $\mathbb{P}_{\mathbf{n}}$, the limiting process in Lemma 3 has initial frequency $p_0 \sim \phi_\alpha^{(n_1,n_2)}(x)\,dx$ given by (6), and $(n_1(0), n_1^{obs}(0), L_1(0);\, n_2(0), n_2^{obs}(0), L_2(0)) = (n_1, n_1, 0;\, n_2, n_2, 0)$. This follows from Bayes' theorem, because by part (i) of Lemmas 2 and 3, the initial frequency has prior density given by (2).

Focus on type 1 for now. We care about the sequence of events (coalescence and mutation) backward in time for type 1, and the timing of these events. At each of these events, the number of type 1 lineages decreases by 1, either by coalescence or by mutation from type 1 to type 2. Hence $n_1^{obs}(t)$ is non-increasing, but $n_1(t)$ can increase over time (backward) due to mutations from type 2 to type 1 (see Figure 1 for an illustration). Furthermore, the difference $n_1(t) - n_1^{obs}(t)$ is the number of type 1 at time t that came from lineages that are of type 2 in the sample (at $t = 0$). We do not care about these $n_1(t) - n_1^{obs}(t)$ lineages, nor the mutation events from type 2 to type 1. Analogous considerations hold for type 2.

From Lemma 3, we immediately obtain the following simplified description for the conditional ancestral process in the limit $N \to \infty$ for the two types. This description is the starting point of our analysis for constant population size; later in Proposition 2 we also obtain the analogous result for time-varying population size.

Proposition 1 (Conditional ancestral process)**.** *The process* $(p_t,\, n_1^{obs}(t),\, L_1(t),\, n_2^{obs}(t),\, L_2(t))_{t\in\mathbb{R}_+}$ *under* $\mathbb{P}_{\mathbf{n}}$ *is a Markov process with state space* $[0,1] \times \{0,1,\cdots,n_1\}^2 \times \{0,1,\cdots,n_2\}^2$ *described as follows:*

(i) $(p_t)_{t\in\mathbb{R}_+}$ *is a solution to* (1) *with initial density* $\phi_\alpha^{(n_1,n_2)}$*. In particular, it does not depend on the process* $(n_1^{obs},\, L_1,\, n_2^{obs},\, L_2)$*.*

(ii) The process $(n_1^{obs},\, L_1,\, n_2^{obs},\, L_2)$ *starts at* $(n_1, 0, n_2, 0)$*. When the current state is* $(p, a_1, \ell_1, a_2, \ell_2)$*, this process evolves as*

$$(a_1, \ell_1) \to \begin{cases} (a_1 - 1, \ell_1) & \textit{at rate} \quad \dfrac{1}{p}\dbinom{a_1}{2} \qquad \textit{coalescence of type } 1^{obs} \\[2ex] (a_1 - 1, \ell_1 + 1) & \textit{at rate} \quad \dfrac{1-p}{p}\, a_1\, \dfrac{\theta_1}{2} \qquad \textit{mutation of } 1^{obs} \textit{ to } 2 \end{cases} \tag{23}$$

and, independently,

$$(a_2, \ell_2) \to \begin{cases} (a_2 - 1, \ell_2) & \textit{at rate} \quad \dfrac{1}{1-p}\dbinom{a_2}{2} \qquad \textit{coalescence of type } 2^{obs} \\[2ex] (a_2 - 1, \ell_2 + 1) & \textit{at rate} \quad \dfrac{p}{1-p}\, a_2\, \dfrac{\theta_2}{2} \qquad \textit{mutation of } 2^{obs} \textit{ to } 1 \end{cases} \tag{24}$$

The total rate in (23), at which a_1 decreases by 1, is

$$\frac{a_1}{2p}\big((1-p)\theta_1 + a_1 - 1\big) := \lambda_{a_1}(p), \tag{25}$$

and the one-step transition probabilities are

$$(a_1, \ell_1) \to \begin{cases} (a_1 - 1, \ell_1) & \text{with probability} \quad \dfrac{a_1 - 1}{(1-p)\theta_1 + a_1 - 1} \\[2ex] (a_1 - 1, \ell_1 + 1) & \text{with probability} \quad \dfrac{(1-p)\theta_1}{(1-p)\theta_1 + a_1 - 1} =: h_{a_1}(p) \end{cases} \tag{26}$$

As t increases from 0 to ∞, the process $(n_1^{obs}(t))_{t\in\mathbb{R}_+}$ decreases from n_1 to 0, and the process $(L_1(t))_{t\in\mathbb{R}_+}$ increases from 0 to a random number $K_1 := \lim_{t\to\infty} L_1(t) \in \mathbb{N}$ which is the total number of latent mutations for type 1. Similarly, the total number of latent mutations for type 2 is defined by $K_2 := \lim_{t\to\infty} L_2(t) \in \mathbb{N}$.

We now give a more explicit description of (K_1, K_2) using the frequency process p and independent Bernoulli random variables. Note these are conditional on the sample counts, $(\mathcal{N}_1 = n_1, \mathcal{N}_2 = n_2)$ as in (8).

Let $\tau_1 < \tau_2 < \cdots < \tau_{n_1}$ be the jump times of the process n_1^{obs}. At time τ_1, the process n_1^{obs} decreases from n_1 to $n_1 - 1$, etc., until finally at τ_{n_1}, n_1^{obs} decreases from 1 to 0. It can be checked that $\tau_{n_1} < \infty$ almost surely under $\mathbb{P}_{\mathbf{n}}$, using the ergodicity of the process p and (25), so that $(n_1^{obs}(t))_{t\in\mathbb{R}_+}$ will indeed decrease to 0 eventually under $\mathbb{P}_{\mathbf{n}}$ almost surely. The allele frequencies at these random times are $p_{\tau_1}, p_{\tau_2}, \ldots, p_{\tau_{n_1}}$. By Proposition 1, under $\mathbb{P}_{\mathbf{n}}$, we have

$$K_1 \overset{d}{=} \xi_{n_1}(p_{\tau_1}) + \xi_{n_1-1}(p_{\tau_2}) + \xi_{n_1-2}(p_{\tau_3}) + \cdots + \xi_2(p_{\tau_{n_1-1}}) + 1, \tag{27}$$

where $\{\xi_k(\cdot)\}_{k=2}^{n_1}$ is a family of independent random processes such that, for a constant $p \in [0,1]$,

$$\xi_k(p) = \begin{cases} 0 & \text{with probability} \quad \dfrac{k-1}{(1-p)\theta_1 + k - 1} \\[2ex] 1 & \text{with probability} \quad \dfrac{(1-p)\theta_1}{(1-p)\theta_1 + k - 1} \end{cases} \tag{28}$$

which is a generalization of ξ_k in (7) and (8), where $\xi_k \equiv \xi_k(0)$.

Similarly, if we let $s_1 < s_2 < \cdots < s_{n_2}$ be the jump times of the process n_2^{obs}, then

$$K_2 \stackrel{d}{=} \zeta_{n_2}(p_{s_1}) + \zeta_{n_2-1}(p_{s_2}) + \zeta_{n_2-2}(p_{s_3}) + \cdots + \zeta_2(p_{s_{n_2-1}}) + 1, \tag{29}$$

where $\{\zeta_k(\cdot)\}_{k=2}^{n_2}$ is a family of independent random processes such that, for a constant $p \in [0,1]$,

$$\zeta_k(p) = \begin{cases} 0 & \text{with probability} \quad \dfrac{k-1}{p\theta_2 + k - 1} \\ 1 & \text{with probability} \quad \dfrac{p\theta_2}{p\theta_2 + k - 1} \end{cases}$$

Analogous to ξ_k in reference to K_1, in what follows we will use the notation

$$\zeta_k \equiv \zeta_k(1) \tag{30}$$

in reference to K_2.

Having described the ancestral process in Proposition 1 and the latent mutations in (27) and (29) under the conditional probability $\mathbb{P}_{\mathbf{n}}$, we study their asymptotic properties under 3 scenarios in the next 3 subsections. These 3 scenarios are a consequence of the asymptotic behaviors of the initial frequency p_0 described in Lemma 4.

Lemma 4 (Asymptotic initial frequency)**.** *Let $n_1 \in \mathbb{N}$ be fixed. The initial frequency p_0 converges in probability under $\mathbb{P}_{\mathbf{n}}$ to a deterministic constant as follows.*

(i) Suppose $n_2 \in \mathbb{N}$ is fixed and $\alpha \to \infty$. Then $p_0 \to 1$.

(ii) Suppose $\alpha \in \mathbb{R}$ is fixed and $n_2 \to \infty$. Then $p_0 \to 0$.

(iii) Suppose $\alpha = \widetilde{\alpha} n_2 + c$ where $\widetilde{\alpha}$, $c \in \mathbb{R}$ are fixed and $n_2 \to \infty$. Then

$$p_0 \to \begin{cases} 0 & \textit{when} \quad \widetilde{\alpha} \in (-\infty, 1] \\ 1 - 1/\widetilde{\alpha} & \textit{when} \quad \widetilde{\alpha} \in (1, \infty) \end{cases} \tag{31}$$

Furthermore, when $\widetilde{\alpha} \in (-\infty, 1)$, it holds that $n_2 p_0$ converges in distribution to the Gamma random variable $\mathrm{Gam}(n_1+\theta_1, 1-\widetilde{\alpha})$ with probability density function $\frac{(1-\widetilde{\alpha})^{(n_1+\theta_1)}}{\Gamma(n_1+\theta_1)}\, y^{n_1+\theta_1-1}\, e^{(\widetilde{\alpha}-1)y}$.

The proof of Lemma 4 is given in Appendix B.

3.1. Scenario (i): strong selection, arbitrary sample size

The first scenario is when $|\alpha|$ large with n_2 fixed. We consider the case $\alpha \to +\infty$ only, since the other case $\alpha \to -\infty$ follows by switching the roles of type 1 and type 2.

The conditional genealogy of the $n_1 + n_2$ sampled individuals, under $\mathbb{P}_{\mathbf{n}}$, has three parts with different timescales. First, the n_2 type-2 lineages quickly evolve (coalesce and mutate) as in the Ewens sampling formula, producing K_2 type 1 lineages at a short time s_{n_2}. Thus, $K_2 \stackrel{d}{=} \sum_{k=1}^{n_2} \zeta_k$ where $\{\zeta_k\}$ are independent Bernoulli variables taking values in $\{0,1\}$ and having means $\frac{\theta_2}{\theta_2+k-1}$. Next, the resulting $n_1 + K_2$ type 1 lineages will coalesce according to the Kingman coalescent without mutation until only one lineage remains. Finally, it takes a long time, $\tau_{n_1} \approx \frac{2\,\alpha}{\theta_1\theta_2}$, for the single lineage to mutate. In particular, $K_1 \approx 1$.

This description is justified by Theorems 1 and 2. See Figure 1 for an illustration.

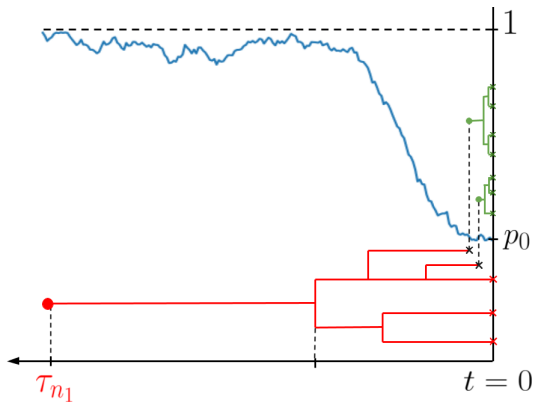

Figure 1: Conditional genealogy of a sample with $(n_1, n_2) = (3, 7)$ at the present time $t = 0$. The fluctuating blue curve shows the process $(p_t)_{t\in\mathbb{R}_+}$ of the population frequency of type 1 backward in time. In this example, p_t approaches 1 from p_0 and the 7 type-2 lineages coalesce and mutate, producing an additional $K_2 = 2$ type-1 lineages. The $5 = 3 + 2$ type-1 lineages then coalesce without mutating, reaching their common ancestral lineage at time τ_{n_1-1} and finally mutating at time τ_{n_1}. Under scenario (i), that is when α is large: p_0 will already be close to 1, coalescence and mutation among the type-2 lineages will occur quickly, with K_2 according to the Ewens sampling formula, coalesce among the type-1 lineages will follow the Kingman coalescent, and $\tau_{n_1} \approx 2\alpha/(\theta_1\theta_2)$.

Theorem 1. *Suppose $(n_1, n_2) \in \mathbb{N}^2$ is fixed. Then under $\mathbb{P}_{\mathbf{n}}$, as $\alpha \to \infty$,*

(i) $\sup_{t\in[0,T]} |1 - p_t| \to 0$ *in probability, for any $T \in (0, \infty)$; and*

(ii) the triplet $(K_1,\, K_2,\, s_{n_2})$ converges in distribution to $\big(1,\, \sum_{k=1}^{n_2} \zeta_k,\, 0\big)$, where $\{\zeta_k\}$ are independent Bernoulli variables taking values in $\{0, 1\}$ and having means $\frac{\theta_2}{\theta_2+k-1}$.

Proof. By (1), the process $q := 1 - p$ solves the stochastic differential equation

$$dq_t = \sqrt{q_t(1-q_t)}\, dW_t - \frac{\theta_1}{2} q_t + \frac{\theta_2}{2}(1-q_t) - \frac{\alpha}{2} q_t(1-q_t), \quad t \geq 0. \tag{32}$$

Fix any $\epsilon \in (0, 1)$. We shall to show that $\mathbb{P}_{\mathbf{n}}(\sup_{t\in[0,T]} q_t > 2\epsilon) \to 0$ as $\alpha \to \infty$.

By the comparison principle (Karatzas and Shreve, 1991, Proposition 2.18 in Chap. 5), we can replace the process q by another process $\widehat{q}$ that solves

$$d\widehat{q}_t = \sqrt{\widehat{q}_t(1-\widehat{q}_t)}\, dW_t + \left[\frac{\theta_2}{2} - \frac{\alpha}{2}\widehat{q}_t(1-\widehat{q}_t)\right] dt \tag{33}$$

with an initial condition $\widehat{q}_0$ that is equal in distribution to q_0. By Girsanov's theorem, we can further take away the constant drift $\frac{\theta_2}{2} dt$. That is, it suffices to show that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which $\mathbb{P}(\sup_{t\in[0,T]} \widehat{Q}_t > 2\epsilon) \to 0$ as $\alpha \to \infty$, where the process $\widehat{Q}$ solves

$$d\widehat{Q}_t = \sqrt{\widehat{Q}_t(1-\widehat{Q}_t)}\, dW_t - \frac{\alpha}{2}\widehat{Q}_t(1-\widehat{Q}_t)\, dt \tag{34}$$

with an initial condition $\widehat{Q}_0$ that is equal in distribution to q_0. The initial frequency $q_0 \to 0$ in probability under $\mathbb{P}_{\mathbf{n}}$, by Lemma 1(i). Hence it suffices to show that

$$\mathbb{P}\left(\sup_{t\in[0,T]} \left\{\int_0^t \sqrt{\widehat{Q}_s(1-\widehat{Q}_s)}\, dW_s - \frac{\alpha}{2}\int_0^t \widehat{Q}_s(1-\widehat{Q}_s)\, ds\right\} > \epsilon\right) \to 0 \quad \text{as } \alpha \to \infty. \tag{35}$$

This is true by the time-change representation of the martingale $M_t := \int_0^t \sqrt{\widehat{Q}_s(1-\widehat{Q}_s)}\, dW_s$ (Karatzas and Shreve, 1991, Theorem 4.6 in Chap. 3) and the fact that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,T]}\left\{B_{\langle M\rangle_t} - \frac{\alpha}{2}\langle M\rangle_t\right\} > \epsilon\right)\\
&= \mathbb{P}\left(\sup_{r\in[0,\,\langle M\rangle_T]}\left\{B_r - \frac{\alpha}{2}r\right\} > \epsilon\right)\\
&\le \mathbb{P}\left(\sup_{r\in[0,\,T/4]}\left\{B_r - \frac{\alpha}{2}r\right\} > \epsilon\right)\\
&= \int_0^{T/4} \frac{1}{\sqrt{2\pi t^3}} \exp\left\{\frac{-(\epsilon + \frac{\alpha}{2}t)^2}{2t}\right\} dt \to 0 \quad \text{as} \quad \alpha \to \infty,
\end{aligned}$$

where in the inequality we used the fact that the quadratic variation $\langle M\rangle_T \le T/4$ almost surely (since $q_t(1-q_t) \le 1/4$ for all $t \in \mathbb{R}_+$). Convergence (i) is proved.

Note that the coalescence rate and the mutation rate in (23) converge to $a_1(a_1-1)/2$ and 0 respectively as $p \to 1$, and that in (24), both rates converge to infinity but their ratio converges in such a way that the limiting one-step transition probabilities are

$$(a_2, \ell_2) \to \begin{cases} (a_2-1, \ell_2) & \text{with probability} \quad \frac{a_2-1}{\theta_2+a_2-1} \\ (a_2-1, \ell_2+1) & \text{with probability} \quad \frac{\theta_2}{\theta_2+a_2-1} \end{cases} \tag{36}$$

Convergence (ii) then follows from (i) and the representations (27) and (29). □

Next, we quantify the difference $q_t := 1 - p_t$ by rescaling. Let $(Q_t)_{t\in\mathbb{R}_+}$ be the $\mathbb{R}_+$-valued process that solves the stochastic differential equation

$$dQ_t = \sqrt{Q_t}\, dW_t + \frac{1}{2}(\theta_2 - Q_t)\, dt, \qquad t \in \mathbb{R}_+, \tag{37}$$

where W is the Wiener process.

Equation (37) (also (42) and (46) below) is equivalent to the Cox-Ingersoll-Ross (CIR) model for interest rates in financial mathematics. It has several other names including the Feller process and the square-root process (Dufresne, 2001). It has a unique strong solution. This equation is not explicitly solvable, but its transition density is explicitly known (Vanyolos et al., 2014) and its moments and distributions have been intensively studied.

In each of Lemma 5, Theorem 3 and Lemma 6 below, we consider weak convergence of a family of random variables in the canonical path space $C([0,\infty),\mathbb{R})$ equiped with the topology of local uniform convergence.

Lemma 5. *Let $q_t = 1 - p_t$ for $t \in \mathbb{R}_+$. The process $(\alpha q_{\frac{t}{\alpha}})_{t\in\mathbb{R}_+}$ converges in distribution under $\mathbb{P}_{\mathbf{n}}$ to the process Q solving* (37) *and starting at* 0.

Proof. Let $Y_t := \alpha q_{\frac{t}{\alpha}}$. Similar to (44),

$$\begin{aligned}
Y_t - Y_0 &= \alpha\,(q_{t/\alpha} - q_0) \\
&= \alpha \int_0^{t/\alpha} \sqrt{q_s(1-q_s)}\,dW_s + \frac{\alpha}{2}\int_0^{t/\alpha} \theta_2(1-q_s) - \theta_1 q_s - \alpha q_s(1-q_s)\,ds \\
&\overset{d}{=} \frac{\alpha}{\sqrt{\alpha}} \int_0^{t} \sqrt{q_{r/\alpha}(1-q_{r/\alpha})}\,dW_r + \frac{\alpha}{2}\int_0^{t} \theta_2(1-q_{r/\alpha}) - \theta_1 q_{r/\alpha} - \alpha q_{r/\alpha}(1-q_{r/\alpha})\,\frac{dr}{\alpha} \\
&= \sqrt{\alpha} \int_0^{t} \sqrt{\frac{Y_r}{\alpha}\left(1-\frac{Y_r}{\alpha}\right)}\,dW_r \;+\; \frac{1}{2}\int_0^{t} \theta_2\left(1-\frac{Y_r}{\alpha}\right) - \theta_1\frac{Y_r}{\alpha} - \alpha\frac{Y_r}{\alpha}\left(1-\frac{Y_r}{\alpha}\right)\,dr \\
&= \int_0^{t} \sqrt{Y_r\left(1-\frac{Y_r}{\alpha}\right)}\,dW_r \;+\; \frac{1}{2}\int_0^{t} \theta_2\left(1-\frac{Y_r}{\alpha}\right) - \theta_1\frac{Y_r}{\alpha} - Y_r\left(1-\frac{Y_r}{\alpha}\right)\,dr. \qquad (38)
\end{aligned}$$

By Lemma 1(i), $Y_0 \to 0$ in probability.

The desired weak convergence of the family $\{\alpha q_{\frac{\cdot}{\alpha}}\}$ of random variables in the canonical path space $C([0,\infty),\mathbb{R})$, can be proved by a standard compactness argument (Billingsley, 1999, Chap. 2) or (Ethier and Kurtz, 2005, Chap. 3). That is, we first show that the family is relatively compact: any subsequence has a further subsequence that converges in distribution as $n \to \infty$. This can be done using the Prohorov's theorem. Next, we identify that any subsequential limit is equal in distribution to X, by showing that they solve the same martingale problem. □

In Theorem 2 below, we obtain that the mean of the age τ_{n_1} is about $\frac{2\,\alpha}{\theta_1\theta_2}$.

Theorem 2 (Age of the oldest latent mutation of a favorable allele)**.** *Suppose $(n_1, n_2) \in \mathbb{N}^2$ is fixed. Then under $\mathbb{P}_{\mathbf{n}}$, as $\alpha \to \infty$, $\frac{\tau_{n_1}}{\alpha}$ converges in distribution to an exponential random variable with mean $\frac{2}{\theta_1\theta_2}$. That is,*

$$\frac{\tau_{n_1}}{\alpha} \xrightarrow{d} \mathrm{Exp}\left(\frac{\theta_1\,\theta_2}{2}\right) \qquad \textit{as} \quad \alpha \to \infty.$$

Proof. Recall the rate of mutation in (23) with $a_1 = 1$. For any $t \in \mathbb{R}_+$, as $\alpha \to \infty$,

$$\begin{aligned}
\mathbb{P}_{\mathbf{n}}\left(\frac{\tau_{n_1}}{\alpha} > t\right) &\approx \mathbb{E}\left[e^{\frac{-\theta_1}{2}\int_0^{\alpha t}\frac{1-p_s}{p_s}\,ds}\right] \qquad (39) \\
&= \mathbb{E}\left[e^{\frac{-\theta_1}{2}\int_0^{\alpha^2 t}\frac{q_{r/\alpha}}{1-q_{r/\alpha}}\,\frac{dr}{\alpha}}\right] \\
&= \mathbb{E}\left[e^{\frac{-\theta_1\,t}{2\alpha^2 t}\int_0^{\alpha^2 t}\frac{\alpha q_{r/\alpha}}{1-q_{r/\alpha}}\,dr}\right].
\end{aligned}$$

Note that in the exponent, by Lemma 5,

$$\lim_{\alpha\to\infty} \frac{-\theta_1\,t}{2\alpha^2 t}\int_0^{\alpha^2 t} \frac{\alpha q_{r/\alpha}}{1-q_{r/\alpha}}\,dr \overset{d}{=} \frac{-\theta_1\,t}{2}\lim_{T\to\infty}\frac{1}{T}\int_0^T Q_r\,dr \;=\; \frac{-\theta_1\,t}{2}\,\mathbb{E}_\infty[Q],$$

where the last equality holds almost surely because the process Q is ergodic. Here $\mathbb{E}_\infty$ is the expectation with respect to the stationary distribution of the process Q to which it converges to as $t \to \infty$.

The limiting distribution of Q_t, as $t \to \infty$, is known to be the Gamma distribution $\mathrm{Gam}(\theta_2, 1)$ with probability density function $\frac{1}{\Gamma(\theta_2)}\, y^{\theta_2 - 1}\, e^{-y}$. Hence $\mathbb{E}_\infty[Q] = \theta_2$. $\square$

In Theorem 2, (39), as in all proofs in this paper, $A \approx B$ means that $A/B \to 1$ in the limit specified, which is either $\alpha \to \infty$ or $n \to \infty$. This is equivalent to $A = B + o(1)$ where B converges and $o(1)$ represents terms which tends to 0 in the limit.

3.2. Scenario (ii): arbitrary selection, large sample size

The second scenario is when n_2 large with α fixed. This scenario has the same asymptotic picture as that of the next scenario when $\widetilde{\alpha} = 0$, as described in Remark 1.

The conditional genealogy of the n_1 type 1 individuals in the sample can be described as follows. Events among the type 1 lineages occur quickly under $\mathbb{P}_{\mathbf{n}}$ in the sense that τ_{n_1} is of order $O(1/n)$. However, if we measure time in proportion to $1/n$ coalescent time units and measure frequency on the scale of numbers of copies of alleles, then the n_1 type-1 lineages evolve (coalesce and mutate) as in the Ewens sampling formula. In particular, $K_1 \approx 1 + \sum_{k=2}^{n_1} \xi_k$, where $\{\xi_k\}$ are independent Bernoulli variables taking values in $\{0, 1\}$ and having means $\frac{\theta_1}{\theta_1 + k - 1}$.

The rescaled frequency process for type 1 can be described precisely under the rescaling above by the Feller diffusion with drift:

$$dZ_t = \sqrt{Z_t}\, dW_t + \frac{\theta_1}{2}\, dt, \qquad t \in \mathbb{R}_+, \tag{40}$$

with initial distribution being the Gamma random variable $\mathrm{Gam}(n_1 + \theta_1, 1)$. See Figure 2 for an illustration.

3.3. Scenario (iii): strong selection, large sample size

The third scenario is when both $|\alpha|$ and n_2 large with $\widetilde{\alpha} = \alpha/n_2$ fixed. We shall further consider two cases: $\widetilde{\alpha} \in (-\infty, 1)$ and $\widetilde{\alpha} \in (1, \infty)$ in two subsections below. Lemma 4 implies that

$$\mathbb{E}_{\mathbf{n}}[p_0] \approx \begin{cases} \frac{\theta_1 + n_1}{1 - \widetilde{\alpha}}\, \frac{1}{n} & \text{when} \quad \widetilde{\alpha} \in (-\infty, 1) \\ 1 - 1/\widetilde{\alpha} & \text{when} \quad \widetilde{\alpha} \in (1, \infty) \end{cases} \quad \text{as } n \to \infty. \tag{41}$$

Therefore, it makes sense to consider the two cases in scenario (iii).

3.3.1. Case $\widetilde{\alpha} \in (-\infty, 1)$

In this case, under $\mathbb{P}_{\mathbf{n}}$ and as $n \to \infty$, we have that $p_0 = O(1/n)$ by Lemma 4. The genealogy of the n_1 type 1 lineages are the same as that in Senario (ii); see Figure 2. This description is precised and justified by Theorems 3-4 and Corollary 1 below.

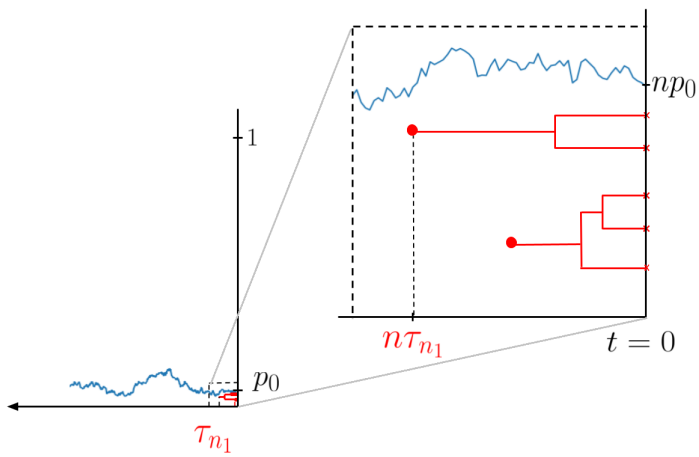

Figure 2: Conditional genealogy of a sample with observed frequencies (n_1, n_2) at the present time $t = 0$, where $n_1 = 5$ and n_2 is large, and $\alpha = \widetilde{\alpha} n_2$ for a constant $\widetilde{\alpha} \in (-\infty, 1)$. The n_2 samples are not shown. In this figure, $K_1 = 2$ and the two red bullets are mutation events from type 1 to type 2. In this scenario (iii), K_1 is distributed like the number of alleles in the Ewens sampling formula, and the timing of the type-1 events are small (of order $O(1/n)$ on the coalescent time scale). The rescaled process $\left(np_{\frac{t}{n}}\right)_{t\in\mathbb{R}_+}$ is well approximated by the diffusion process (42) with initial distribution $\mathrm{Gam}(n_1 + \theta_1, 1 - \widetilde{\alpha})$.

Let $(Z_t)_{t\in\mathbb{R}_+}$ be the $\mathbb{R}_+$-valued process that has initial state $Z_0 \sim \mathrm{Gamma}(n_1 + \theta_1, 1 - \widetilde{\alpha})$ and solves the stochastic differential equation

$$dZ_t = \sqrt{Z_t}\, dW_t + \frac{1}{2}(\theta_1 + \widetilde{\alpha}\, Z_t)\, dt, \qquad t \in \mathbb{R}_+, \tag{42}$$

where W is the Wiener process.

Theorem 3 (Convergence of rescaled genealogy)**.** *Suppose $\widetilde{\alpha} = \alpha/n_2 \in (-\infty, 1)$ is fixed. As $n \to \infty$, the process $\left(np_{\frac{t}{n}},\, n_1^{obs}(\frac{t}{n}),\, L_1(\frac{t}{n})\right)_{t\in\mathbb{R}_+}$ converges in distribution under $\mathbb{P}_{\mathbf{n}}$ to a Markov process $\left(Z_t,\, \widetilde{n}_1(t),\, \widetilde{L}_1(t)\right)_{t\in\mathbb{R}_+}$ with state space $\mathbb{R}_+ \times \{0, 1, \cdots, n_1\} \times \{0, 1, \cdots, n_1\}$ described as follows:*

(i) $(Z_t)_{t\in\mathbb{R}_+}$ is a solution to (42) with initial state $Z_0 \sim \mathrm{Gamma}(n_1 + \theta_1, 1 - \widetilde{\alpha})$. In particular, its transition kernel does not depend on $(\widetilde{n}_1,\, \widetilde{L}_1)$.

(ii) Suppose the current state is (z, a_1, ℓ_1). Then $(\widetilde{n}_1,\, \widetilde{L}_1)$ evolves as

$$(a_1, \ell_1) \to \begin{cases} (a_1 - 1, \ell_1) & \textit{at rate} \quad \frac{1}{z}\binom{a_1}{2} \qquad \textit{coalescence of type 1} \\ (a_1 - 1, \ell_1 + 1) & \textit{at rate} \quad \frac{1}{z}\, a_1\, \frac{\theta_1}{2} \qquad \textit{mutation of 1 to 2} \end{cases} \tag{43}$$

Proof. Let $Y_t := np_{t/n}$. By Lemma 4, under $\mathbb{P}_{\mathbf{n}}$ we have $Y_0 = np_0 \to \mathrm{Gam}(n_1+\theta_1, 1-\widetilde{\alpha})$. By (1),

$$\begin{aligned}
Y_t - Y_0 &= n\,(p_{t/n} - p_0) \\
&= n\int_0^{t/n} \sqrt{p_s(1-p_s)}\,dW_s + \frac{n}{2}\int_0^{t/n} \theta_1(1-p_s) - \theta_2 p_s + \alpha p_s(1-p_s)\,ds \\
&\overset{d}{=} \frac{n}{\sqrt{n}}\int_0^{t} \sqrt{p_{r/n}(1-p_{r/n})}\,dW_r + \frac{n}{2}\int_0^{t} \theta_1(1-p_{r/n}) - \theta_2 p_{r/n} + \alpha p_{r/n}(1-p_{r/n})\,\frac{dr}{n} \\
&= \sqrt{n}\int_0^{t} \sqrt{\frac{Y_r}{n}\left(1-\frac{Y_r}{n}\right)}\,dW_r \;+\; \frac{1}{2}\int_0^{t} \theta_1\left(1-\frac{Y_r}{n}\right) - \theta_2\frac{Y_r}{n} + \alpha\frac{Y_r}{n}\left(1-\frac{Y_r}{n}\right)\,dr \\
&= \int_0^{t} \sqrt{Y_r\left(1-\frac{Y_r}{n}\right)}\,dW_r \;+\; \frac{1}{2}\int_0^{t} \theta_1\left(1-\frac{Y_r}{n}\right) - \theta_2\frac{Y_r}{n} + \frac{\alpha}{n}Y_r\left(1-\frac{Y_r}{n}\right)\,dr, \qquad (44)
\end{aligned}$$

where in the third line above we used the fact that the processes $\left(\int_0^{t/n} f(s)\,dW_s\right)_{t\in\mathbb{R}_+}$ and $\left(\frac{1}{\sqrt{n}}\int_0^t f(r/n)dW_r\right)_{t\in\mathbb{R}_+}$ are equal in distribution, where $f(s) = \sqrt{p_s(1-p_s)}$.

Using (44), the fact $\sup_n \mathbb{E}[Y_0^2] < \infty$ and the assumption $\widetilde{\alpha} = \alpha/n_2 \in \mathbb{R}$ is fixed, we can check by Gronwall's inequality that $\limsup_{n\to\infty} \mathbb{E}_{\mathbf{n}}[\sup_{t\in[0,T]} Y_t] < \infty$ for all $T > 0$.

Now, note that equation (42) is the same as (44) after we get rid of the terms $\frac{Y_r}{n}$ and replace $\frac{\alpha}{n}$ by $\widetilde{\alpha}$. As $n \to \infty$, the process $(Y_t)_{t\in[0,T]}$ converges in distribution under $\mathbb{P}_{\mathbf{n}}$ to a process $(Z_t)_{t\in[0,T]}$ with initial state $Z_0 \sim \mathrm{Gam}(n_1+\theta_1, 1-\widetilde{\alpha})$ and solving (42). This weak convergence can be checked using a standard compactness argument (Billingsley, 1999, Chap. 2) or (Ethier and Kurtz, 2005, Chap. 3), outlined in the last paragraph in the proof of Lemma 5. Using (23), the desired joint convergence follows similarly. □

Let $\widetilde{\tau}_1 < \widetilde{\tau}_2 < \cdots < \widetilde{\tau}_{n_1}$ be the jump times of the process $\widetilde{n}_1$ in Theorem 3, at each of which the process decreases by 1.

Theorem 4 (Convergence of timing of events)**.** *Suppose $\widetilde{\alpha} = \alpha/n_2 \in (-\infty, 1)$ fixed. As $n \to \infty$, the random vector $(n_2\,\tau_i,\ n_2\,p_{\tau_i})_{i=1}^{n_1}$ under $\mathbb{P}_{\mathbf{n}}$ converges in distribution to $(\widetilde{\tau}_i,\ Z_{\widetilde{\tau}_i})_{i=1}^{n_1}$.*

Proof. We first give a more explicit description of the jump times $\widetilde{\tau}_1 < \widetilde{\tau}_2 < \cdots < \widetilde{\tau}_{n_1}$, in terms of the function

$$\widetilde{\lambda}_{a_1}(z) := \frac{a_1}{2z}(\theta_1 + a_1 - 1)$$

that comes from (43) in Theorem 3. At the first jump time $\widetilde{\tau}_1$, the process $\widetilde{n}_1$ decreases from n_1 to $n_1 - 1$. Thus $\widetilde{\tau}_1$ is the first jump time of a Poisson process with time inhomogeneous rate $\left(\widetilde{\lambda}_{n_1}(Z_t)\right)_{t\in\mathbb{R}_+}$, given the trajectory $(Z_t)_{t\in\mathbb{R}_+}$. Hence,

$$\mathbb{P}(\widetilde{\tau}_1 > t) = \mathbb{E}\left[e^{-\int_0^t \widetilde{\lambda}_{n_1}(Z_s)\,ds}\right], \qquad t \in \mathbb{R}_+. \qquad (45)$$

Given $(\widetilde{\tau}_1,\ Z_{\widetilde{\tau}_1})$, the difference $\widetilde{\tau}_2 - \widetilde{\tau}_1$ is the first jump time of an independent Poisson process with time inhomogeneous rate $\left(\widetilde{\lambda}_{n_1-1}(Z_{t+\widetilde{\tau}_1})\right)_{t\in\mathbb{R}_+}$. Given $(\widetilde{\tau}_2,\ Z_{\widetilde{\tau}_2})$, the difference $\widetilde{\tau}_3 - \widetilde{\tau}_2$ is the first jump time of an independent Poisson process with time inhomogeneous rate $\left(\widetilde{\lambda}_{n_1-2}(Z_{t+\widetilde{\tau}_2})\right)_{t\in\mathbb{R}_+}$;

and so on. Finally, given $(\widetilde{\tau}_{n_1-1},\, Z_{\widetilde{\tau}_{n_1-1}})$, the difference $\widetilde{\tau}_{n_1} - \widetilde{\tau}_{n_1-1}$ is the first jump time of an independent Poisson process with time inhomogeneous rate $\left(\widetilde{\lambda}_1(Z_{t+\widetilde{\tau}_{n_1-1}})\right)_{t\in\mathbb{R}_+}$.

Using the total rate of type-1 events, $\lambda_{a_1}(p)$ defined in (25), and Theorem 3, as $n \to \infty$ we have

$$\int_0^{t/n} \lambda_{a_1}(p_s)\, ds = \int_0^t \frac{\lambda_{a_1}(p_{s/n})}{n}\, ds = \int_0^t \frac{a_1}{n\, 2p_{s/n}} \big(a_1 - 1 + (1-p_{s/n})\theta_1\big)\, ds$$
$$\to \int_0^t \frac{a_1}{2\, Z_s}(\theta_1 + a_1 - 1)\, ds = \int_0^t \widetilde{\lambda}_{a_1}(Z_s)\, ds.$$

Hence $\mathbb{P}_{\mathbf{n}}(n_2\tau_1 > t) \to \mathbb{P}(\widetilde{\tau}_1 > t)$ for all $t \geq 0$, by (45). Combining with Theorem 3, we have that $n_2\,(\tau_1,\ p_{\tau_1})$ under $\mathbb{P}_{\mathbf{n}}$ converges in distribution to $(\widetilde{\tau}_1,\ Z_{\widetilde{\tau}_1})$ as $n_2 \to \infty$.

Applying the strong Markov property of the process Z at $\widetilde{\tau}_1$ and that of the process p at τ_1, we can similarly show that $n_2\,(\tau_1, \tau_2 - \tau_1,\ p_{\tau_1}, p_{\tau_2} - p_{\tau_1})$ under $\mathbb{P}_{\mathbf{n}}$ converges to $(\widetilde{\tau}_1, \widetilde{\tau}_2 - \widetilde{\tau}_1,\ Z_{\widetilde{\tau}_1}, Z_{\widetilde{\tau}_2} - Z_{\widetilde{\tau}_1})$ in distribution. Continuing in the same way, we obtain that $n_2\left(\tau_i - \tau_{i-1},\ p_{\tau_i} - p_{\tau_{i-1}}\right)_{i=1}^{n_1}$ under $\mathbb{P}_{\mathbf{n}}$ converges in distribution to $\left(\widetilde{\tau}_i - \widetilde{\tau}_{i-1},\ Z_{\widetilde{\tau}_i} - Z_{\widetilde{\tau}_{i-1}}\right)_{i=1}^{n_1}$, where $\tau_0 = \widetilde{\tau}_0 = 0$. The desired convergence then follows. □

In Wakeley et al. (2023, Appendix), we showed that for the case $\alpha = 0$ (no selection) and $n_1 > 1$, the jump times $\tau_1 < \tau_2 < \cdots < \tau_{n_1}$ of n_1^{obs} are of order $1/n$ and the re-scaled vector $(n\tau_i)_{i=1}^{n_1}$ converges in distribution. Theorem 4 therefore generalizes the latter convergence in the presence of selection, in scenario (ii) and in the case $\widetilde{\alpha} \in (-\infty, 1)$ within scenario (iii). This can further be generalized to time-varying populations, as we shall show below. By equation (18) in Wakeley et al. (2023), $\mathbb{E}_{\mathbf{n}}[\tau_1] \approx \frac{2\log n_2}{n_2}$ and $\mathbb{E}_{\mathbf{n}}[\widetilde{\tau}_1] = \infty$ when $n_1 = 1$ and $\alpha = 0$.

Theorem 4 implies that when $\widetilde{\alpha} \in (-\infty, 1)$, the age τ_{n_1} of the oldest latent mutation is well approximated by $\frac{1}{n_2}\widetilde{\tau}_{n_1}$, and the total number of latent mutations for type 1 is predicted by the Ewens sampling formula, as $n \to \infty$.

Corollary 1. *Suppose $\widetilde{\alpha} = \alpha/n_2 \in (-\infty, 1)$ fixed. Then as $n \to \infty$,*

(i) K_1 converges in distribution under $\mathbb{P}_{\mathbf{n}}$ to $1 + \sum_{k=2}^{n_1} \xi_k$, where $\{\xi_k\}$ are independent Bernoulli variables taking values in $\{0, 1\}$ and having means $\frac{\theta_1}{\theta_1+k-1}$.

(ii) $n_2\,\tau_{n_1}$ converges in distribution under $\mathbb{P}_{\mathbf{n}}$ to $\widetilde{\tau}_{n_1}$.

Proof. The vector $(p_{\tau_i})_{i=1}^{n_1-1}$ converges in probability to the zero vector in $\mathbb{R}^{n_1-1}$ as $n \to \infty$, by Theorem 3. This implies that

$$\left(h_k(p_{\tau_{n_1-k+1}})\right)_{k=2}^{n_1} \to \left(\frac{\theta_1}{\theta_1 + k - 1}\right)_{k=2}^{n_1} \in \mathbb{R}^{n_1-1},$$

where we recall $h_k(p) := \frac{(1-p)\theta_1}{(1-p)\theta_1+k-1}$ defined in (26). Hence the following weak convergence in $\mathbb{R}^{n_1-1}$ holds:

$$\left(\xi_k(p_{\tau_{n_1-k+1}})\right)_{k=2}^{n_1} \xrightarrow{d} (\xi_k)_{k=2}^{n_1}.$$

The proof is complete by (27). □

Remark 1. Theorems 3-4 and Corollary 1 remain valid if "$\widetilde{\alpha} = \alpha/n_2 \in (-\infty, 1)$ is fixed" is replaced by "$\alpha = \widetilde{\alpha}n_2 + c$ where $\widetilde{\alpha}$, $c \in \mathbb{R}$ are fixed". In particular, these results hold for Scenario (ii).

3.3.2. Case $\widetilde{\alpha} \in (1, \infty)$

In this case, under $\mathbb{P}_{\mathbf{n}}$ and as $n \to \infty$, we have that $p_0 \to 1 - 1/\widetilde{\alpha} > 0$ by (31). The process $(p_t)_{t \in \mathbb{R}_+}$ increases very quickly and stays close to 1. As a result, the conditional ancestral process for the n_1 type 1 samples has two parts with different timescales. First it coalesces only as the Kingman coalescent (without mutation) until there is only one lineage. Then it takes a very long time (about $n \frac{2\widetilde{\alpha}}{\theta_1 \theta_2} \approx \frac{2\alpha}{\theta_1 \theta_2}$) for the single latent mutation to occur. In particular, $K_1 \approx 1$.

This description is justified by Theorems 5-6. See Figure 3 for an illustration.

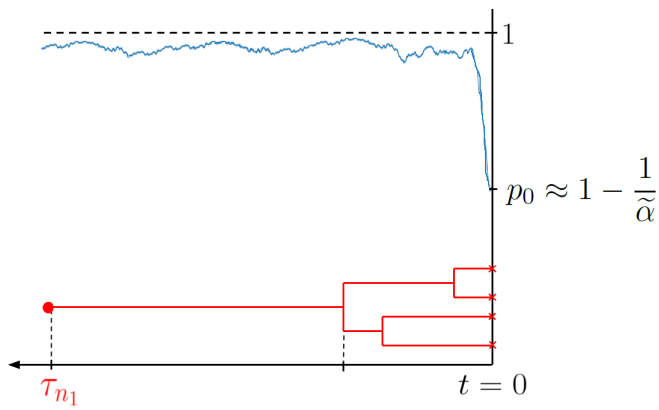

Figure 3: Conditional genealogy of a sample with observed frequencies (n_1, n_2) at the present time $t = 0$, where $n_1 = 4$ and n_2 is large, and $\alpha = \widetilde{\alpha} n_2$ for a constant $\widetilde{\alpha} \in (1, \infty)$. The n_2 samples are not shown. This scenario is reminiscent of scenario (i) if we focus on the genealogy of only the A_1 lineages. These lineages first coalesce as the Kingman coalescent (without mutation) until there is only one lineage, which take $O(1)$ amount of time. Then it takes $O(n_2)$ amount of time for the single latent mutation to occur at time τ_{n_1}.

Theorem 5. *Suppose $\widetilde{\alpha} = \alpha / n_2 \in (1, \infty)$ is fixed. As $n \to \infty$, under $\mathbb{P}_{\mathbf{n}}$,*

(i) $\sup_{t \in [S,T]} |1 - p_t| \to 0$ *in probability, for any $0 < S < T < \infty$; and*

(ii) For any $T > 0$, the process $(n_1^{obs}(t), L_1(t))_{t \in [0,T]}$ converges in distribution to a process $(\widetilde{n}^{obs}(t), 0)_{t \in [0,T]}$ where $\widetilde{n}^{obs}$ is a pure death process with jump rate $\binom{k}{2} = \frac{k(k-1)}{2}$ from k to $k - 1$.

(iii) $K_1 \to 1$ in probability.

Proof. We first observe that the process p gets close to 1 quickly, when $\widetilde{\alpha} \in (1, \infty)$, in the sense that for any $\epsilon \in (0, 1)$,

$$\lim_{n \to \infty} \mathbb{P}_{\mathbf{n}} (\tau_{1-\epsilon} > S) = \lim_{n \to \infty} \mathbb{P}_{\mathbf{n}} \left(\sup_{t \in [0,S]} p_t < 1 - \epsilon \right) = 0,$$

where $\tau_{1-\epsilon} := \inf\{t \in \mathbb{R}_+ : p_t > 1 - \epsilon\}$ is the first time p hits above $1 - \epsilon$. This is true because $p_0 \to 1 - 1/\widetilde{\alpha} \in (0, 1)$ in probability, so that the growth term $\widetilde{\alpha} n_2 \, p_t (1 - p_t)$ is large at least when $t > 0$ is small. Next, suppose the process p starts at $1 - \epsilon$ (i.e. the process $q = 1 - p$ starts at ϵ), we show that the exit time of the process q out of the interval $[0, 2\epsilon)$ is longer than T with probability tending to 1, as $n \to \infty$. More precisely,

$$\mathbb{P}_{1-\epsilon} \left(\inf_{t \in [0,T]} p_t < 1 - 2\epsilon \right) = \mathbb{P}_{\epsilon} \left(\sup_{t \in [0,T]} q_t > 2\epsilon \right)$$

which tends to 0 as $n_2 \to \infty$, as in the proof of Theorem 1(i).

From these two estimates and the strong Markov property of the process p, we have that for any $\epsilon \in (0,1)$,

$$\begin{aligned}
&\mathbb{P}_{\mathbf{n}}\left(\sup_{t\in[S,T]} |1-p_t| > 2\epsilon\right) \\
&\le \mathbb{P}_{\mathbf{n}}\left(\sup_{t\in[S,T]} |1-p_t| > 2\epsilon,\ \tau_{1-\epsilon} \le S\right) + \mathbb{P}_{\mathbf{n}}\left(\tau_{1-\epsilon} > S\right) \\
&= \mathbb{E}_{\mathbf{n}}\left[1_{\{\tau_{1-\epsilon}\le S\}}\, \mathbb{P}_{1-\epsilon}\left(\inf_{t\in[S-\tau_{1-\epsilon},\, T-\tau_{1-\epsilon}]} p_t < 1-2\epsilon\right)\right] + \mathbb{P}_{\mathbf{n}}\left(\tau_{1-\epsilon} > S\right) \\
&\le \mathbb{P}_{1-\epsilon}\left(\inf_{t\in[0,\,T]} p_t < 1-2\epsilon\right) + \mathbb{P}_{\mathbf{n}}\left(\tau_{1-\epsilon} > S\right) \to 0 \qquad \text{as } n\to\infty.
\end{aligned}$$

The proof of part (i) is complete.

By Part (i) and (25), the times for the type 1 events are of order $O(1)$ and $h_k(p_t) \to 0$ where $h_k(p)$ is defined in (26). Hence Parts (ii) and (iii) follow. □

Now we consider the second part of the genealogy, when there is a single lineage left (i.e. during τ_{n_1-1} and τ_{n_1}).

To estimate the age τ_{n_1} of the single latent mutation, we can ignore the $n_1 - 1$ jump times of the process $\widetilde{n}^{obs}$ (since they are of order 1 by Theorem 5). The frequency of type 1 is tightly regulated in the sense that it is close to 1 in the sense of Theorem 5(i). However, we need to know "how close it is to 1" in order to get an estimate of τ_{n_1}, because simply setting $p = 1$ in $\frac{1-p}{2p}\theta_1$ will give us zero.

We quantify the difference $1 - p_t$ by rescaling the process $q := 1 - p$ as in Theorem 3. The emergent process is very similar to (42). Let $(\widetilde{q}_t)_{t\in\mathbb{R}_+}$ be the $\mathbb{R}_+$-valued process that solves the stochastic differential equation

$$d\widetilde{q}_t = \sqrt{\widetilde{q}_t}\, dW_t + \frac{1}{2}(\theta_2 - \widetilde{\alpha}\widetilde{q}_t)\, dt, \qquad t \in \mathbb{R}_+, \tag{46}$$

where W is the Wiener process. This process $\widetilde{q}$ is a mean reverting Cox-Ingersoll-Ross (CIR) model. It will never touch zero if $\theta_2 \ge 1$; otherwise it can occasionally touch the boundary zero point. The CIR model is an ergodic process with stationary density given by the Gamma distribution $\mathrm{Gam}(\theta_2, \widetilde{\alpha})$, which has probability density function $\frac{\widetilde{\alpha}^{\theta_2}}{\Gamma(\theta_2)}\, y^{\theta_2 - 1}\, e^{-\widetilde{\alpha} y}$.

Lemma 6. *Suppose $\widetilde{\alpha} = \alpha/n_2 \in \mathbb{R}$ is fixed. Let $q_t = 1 - p_t$ for $t \in \mathbb{R}_+$. Suppose $nq_0 \to c_0$ as $n \to \infty$. Then the process $(nq_{\frac{t}{n}})_{t\in\mathbb{R}_+}$ converges in distibution to the solution process $\widetilde{q}$ to* (46) *starting at c_0.*

Proof. Following the proof of Lemma 5, we use (32), let $Y_t := nq_{\frac{t}{n}}$ and suppose $Y_0 \to c_0$. Then, similar to (38),

$$\begin{aligned}
Y_t - Y_0 &= n\left(q_{t/n} - q_0\right) \\
&\stackrel{d}{=} \int_0^t \sqrt{Y_r\left(1 - \frac{Y_r}{n}\right)}\, dW_r + \frac{1}{2}\int_0^t \theta_2\left(1 - \frac{Y_r}{n}\right) - \theta_1 \frac{Y_r}{n} - \frac{\alpha}{n} Y_r\left(1 - \frac{Y_r}{n}\right) dr.
\end{aligned} \tag{47}$$

Using the assumption that $\widetilde{\alpha} = \alpha/n_2 \in \mathbb{R}$ is fixed, the desired weak convergence can be proved by a standard compactness argument (Billingsley, 1999, Chap. 2) or (Ethier and Kurtz, 2005, Chap. 3), outlined in the last paragraph in the proof of Lemma 5. □

Theorem 6 below is analogous to Theorem 2. We obtain that the mean of the age τ_{n_1} is about $n\frac{2\,\widetilde{\alpha}}{\theta_1\theta_2} \approx \frac{2\,\alpha}{\theta_1\theta_2}$ when it is larger than $\frac{2n}{\theta_1\theta_2}$ and n is large.

Theorem 6 (Age of the unique latent mutation)**.** *Suppose $\widetilde{\alpha} = \alpha/n_2 \in (1,\infty)$ is fixed. As $n \to \infty$, $\frac{\tau_{n_1}}{n}$ converges in distribution under $\mathbb{P}_{\mathbf{n}}$ to an exponential random variable with mean $\frac{2\,\widetilde{\alpha}}{\theta_1\theta_2}$. That is,*

$$\frac{\tau_{n_1}}{n} \xrightarrow{d} \mathrm{Exp}\left(\frac{\theta_1\,\theta_2}{2\,\widetilde{\alpha}}\right) \qquad \textit{as} \quad n \to \infty.$$

Proof. We argue as in the proof of Theorem 2. For any $t \in \mathbb{R}_+$, as $n \to \infty$, by Theorem 5,

$$\mathbb{P}_{\mathbf{n}}\left(\frac{\tau_{n_1}}{n} > t\right) \approx \mathbb{E}\left[\exp\left\{\frac{-\theta_1\,t}{2n^2 t}\int_0^{n^2 t}\frac{nq_{r/n}}{1-q_{r/n}}\,dr\right\}\right].$$

Note that in the exponent, by Lemma 6,

$$\lim_{n\to\infty}\frac{-\theta_1\,t}{2n^2 t}\int_0^{n^2 t}\frac{nq_{r/n}}{1-q_{r/n}}\,dr \overset{d}{=} \frac{-\theta_1\,t}{2}\lim_{T\to\infty}\frac{1}{T}\int_0^T \widetilde{q}_r\,dr \;=\; \frac{-\theta_1\,t}{2}\,\mathbb{E}_\infty[\widetilde{q}],$$

where the last equality holds almost surely because the process $\widetilde{q}$ solving (46) is ergodic. Here $\mathbb{E}_\infty$ is the expectation with respect to the stationary distribution of the process $\widetilde{q}$ to which it converges as $t \to \infty$.

The limiting distribution of $\widetilde{q}_t$, as $t \to \infty$, is known to be the Gamma distribution $\mathrm{Gam}(\theta_2, \widetilde{\alpha})$. Hence $\mathbb{E}_\infty[\widetilde{q}] = \theta_2/\widetilde{\alpha}$. □

3.4. Time-varying population size

For a population with time-varying size $\rho(t)N$ at forward time t where ρ is a non-constant function, neither the Moran process nor its diffusion approximation possess a stationary distribution. However, the random background approach of Barton et al. (2004) can be generalized to this setting by considering the time-reversed frequency process.

Our main message in this section is that the limiting pictures in scenarios (i) and (ii) are robust against continuously-changing population sizes and the initial distribution μ_0 of the initial (ancient) frequency X_0. Roughly speaking, this robustness comes from the fact that by taking either α or $n_2 \sim n$ large, the events among finite-count alleles in the sample are so sped up that the population size will have hardly changed by the time all their coalescent and latent mutation events have occurred. The limiting picture for scenario (iii) is more subtle. It depends on the large deviation behavior under the of the present day frequency as $n_2 \to \infty$.

Let $T > 0$ be the present. For comparison with our results for constant population size, we keep the same definitions of θ_1, θ_2 and α and we set $\rho(T) = 1$. Thus, N is the population size at the present time T, and θ_1, θ_2 and α are the present-day values of these variables. The demographic function ρ could for example represent exponential population growth, in which case $\rho(t) = \rho(0)e^{\beta t}$ for some positive constant β. This model was used in Wakeley et al. (2023) to illustrate the effects of rapid growth on neutral rare variation in humans. Here we allow that $\rho(t)$ is piecewise continuous. As will become clear, the key feature of ρ for our results is that it is continuous at T.

Since the random background approach of Barton et al. (2004) was formulated based on the lineage dynamics of the Moran model, we begin by describing the diffusion process arising from a Moran model with time-varying population size.

Lemma 7 (Diffusion limit for time-varying Moran model)**.** *Let $\rho : \mathbb{R}_+ \to (0, \infty)$ be a piecewise continous function with finitely many jumps, and N be a positive integer. Consider the discrete-time Moran process in which, at step $k = [N(N-1)t/2]$, the total population size is $[\rho(t)N]$ and N is replaced by $\rho(t)N$ in the one-step transition* (13)-(14)*. Suppose $u_1 = \theta_1/N$, $u_2 = \theta_2/N$ and $s = -\alpha/N$. Then as $N \to \infty$, the relative frequency of A_1 at step $[N(N-1)t/2]$ converges in distribution to X_t solving*

$$dX_t = \left[\frac{\theta_1}{2\rho(t)}(1 - X_t) - \frac{\theta_2}{2\rho(t)}X_t + \frac{\alpha}{2\rho(t)}X_t(1 - X_t)\right] dt + \sqrt{\frac{X_t(1 - X_t)}{\rho^2(t)}}\, dW_t, \tag{48}$$

where W_t is the Wiener proces, provided that the initial relative frequency converges to $X(0)$.

Putting $\rho(t) \equiv 1$ for all $t \in \mathbb{R}_+$, or $\beta = 0$ in the exponential growth model, makes (48) identical to (1). The term $\rho^2(t)$ in the denominator inside the square root comes from the diffusion timescale of the Moran model: for a population of constant size, one unit of time in the diffusion is $N(N-1)/2 \propto N^2$ times steps in the discrete model. To explain the term $\frac{\alpha}{2\rho(t)} = \frac{-s\rho(t)N}{2\rho^2(t)}$, note that the rate of change of X_t due to selection is proportional to the product of the total size $\rho(t)N$ and the parameter s, which is then multiplied by $1/\rho^2(t)$ because the timescale in (48) is defined in terms of the present-day population size N. The proof of Lemma 7 is given in Appendix B.

Remark 2 (Wright-Fisher model with varying size)**.** The analogous diffusion process X^{WF} for the discrete Wright-Fisher model with total size $[\rho(t)N]$ in generation $[Nt]$ is *different from* the process X in (48), except in the case $\rho(t) \equiv 1$ for all $t \in \mathbb{R}_+$. This diffusion solves the SDE

$$dX_t^{\mathrm{WF}} = \left[\frac{\theta_1}{2}(1 - X_t^{\mathrm{WF}}) - \frac{\theta_2}{2}X_t^{\mathrm{WF}} + \frac{\alpha}{2}X_t^{\mathrm{WF}}(1 - X_t^{\mathrm{WF}})\right] dt + \sqrt{\frac{X_t^{\mathrm{WF}}(1 - X_t^{\mathrm{WF}})}{\rho(t)}}\, dW_t,$$

which is the adaptation of equation (1) in Schraiber et al. (2016) to our haploid model of selection and recurrent mutation; see also equation (21) in Evans et al. (2007). The generators of X^{WF} and X are related by $\mathcal{A}_t^{\mathrm{WF}} = \rho(t)\mathcal{A}_t$ for all $t \in \mathbb{R}_+$. In other words, the diffusion X^{WF} from the discrete Wright-Fisher model is sped up by the factor $\rho(t)$ at time t.

To compare X and X^{WF}, we can perform deterministic time-changes to normalize their diffusion coefficients to be the same. In general, suppose X satisfies the SDE $dX_t = b(t, X_t)dt + \sigma(t, X_t)\, dW_t$ and Y is a time-change of X defined by $Y_r := X_{\psi(r)}$, where ψ is any fixed continuous and strictly increasing function, then

$$dY_r = b(\psi(r), Y_r)\, \psi'(r)\, dr + \sigma(\psi(r), Y_r)\, \sqrt{\psi'(r)}\, dW_r.$$

Hence, when X is a (weak) solution to (48) and $\psi = g^{-1}$ where g is the unique continuous function such that $g(t) = \int_0^t \frac{1}{\rho^2(s)} ds$, we obtain that $Y_r := X_{g^{-1}(r)}$ solves

$$dY_r = \rho(g^{-1}(r))\, b(Y_r)\, dr + \sqrt{Y_r(1 - Y_r)}\, dW_r, \tag{49}$$

where $b(y) = \frac{\theta_1}{2}(1 - y) - \frac{\theta_2}{2}y + \frac{\alpha}{2}y(1 - y)$. Analogously, following Schraiber et al. (2016)—see their equation (6) and the SDE below it—and taking f such that $f(t) = \int_0^t \frac{1}{\rho(s)} ds$, we find that $Y_r^{\mathrm{WF}} := X_{f^{-1}(r)}^{\mathrm{WF}}$ solves

$$dY_r^{\mathrm{WF}} = \rho(f^{-1}(r))\, b(Y_r^{\mathrm{WF}})\, dr + \sqrt{Y_r^{\mathrm{WF}}(1 - Y_r^{\mathrm{WF}})}\, dW_r. \tag{50}$$

Since $f \neq g$ unless $\rho(t) \equiv 1$ for all $t \in \mathbb{R}_+$, we have that $Y \neq Y^{\mathrm{WF}}$, i.e. $X_{g^{-1}(r)} \neq X^{\mathrm{WF}}_{f^{-1}(r)}$ in general. Nonetheless, (49) and (50) have the same form, the only difference being the way time r in these diffusions is related back to time t in the discrete models.

Note that the law of the present-time frequency of A_1 in the model of Lemma 7 depends on the distribution μ_0 of the initial frequency $X(0)$. This law is denoted by $\mathbb{P}_{\mu_0}(X_T \in dy)$.

Suppose a sample of n individuals are picked uniformly at random at the present time $T > 0$, i.e. when the frequency of A_1 is X_T, and we know that n_1 of them are of type 1 (and $n - n_1$ are of type 2). Let $\mathbb{P}_{\mathbf{n}} = \mathbb{P}_{\mathbf{n},\mu_{\mathbf{0}}}$ be the conditional probability measure given the sample count $\mathbf{n} = (n_1, n_2)$. We also denote the conditional law of the present frequency $p_0 = X_T$, under $\mathbb{P}_{\mathbf{n}}$, by $\mathcal{L}^{\mathbf{n}} := \mathbb{P}_{\mathbf{n}}(X_T \in dy) = \mathbb{P}_{\mu_0}(X_T \in dy \,|\, \mathbf{n})$. Then

$$\mathcal{L}^{\mathbf{n}} = C_T \, y^{n_1}(1-y)^{n_2} \, \mathbb{P}_{\mu_0}(X_T \in dy), \tag{51}$$

where $C_T = \left(\int_0^1 y^{n_1}(1-y)^{n_2}\mathbb{P}_{\mu_0}(X_T \in dy)\right)^{-1}$ is a normalizing constant. This follows from Bayes' theorem, just like (5) did, but with prior distribution $\mathbb{P}_{\mu_0}(X_T \in dy)$.

Similar to Proposition 1, the conditional ancestral process in the diffusion limit can be described as follows. This description involves the backward frequency process

$$p_t := X_{T-t} \qquad \text{for} \quad t \in [0,T] \tag{52}$$

which is by definition the time-reversal of the process X.

Proposition 2 (Conditional ancestral process)**.** *Let $T > 0$ and $\mu_0 \in \mathcal{P}([0,1])$ be fixed, and the demographic function ρ be as in Lemma 7. The process $(p_t,\, n_1^{obs}(t),\, L_1(t),\, n_2^{obs}(t),\, L_2(t))_{t\in[0,T]}$ under $\mathbb{P}_{\mathbf{n}}$ is a time-inhomogeneous Markov process with state space $[0,1] \times \{0,1,\cdots,n_1\}^2 \times \{0,1,\cdots,n_2\}^2$ described as follows:*

(i) $(p_t)_{t\in[0,T]}$, defined by (52), has the law of $(X_{T-t})_{t\in[0,T]}$ under $\mathbb{P}_{\mathbf{n}}$. In particular, it has initial distribution $\mathcal{L}^{\mathbf{n}}$ in (51) and it does not depend on the process $(n_1^{obs},\, L_1,\, n_2^{obs},\, L_2)$.

(ii) The process $(n_1^{obs},\, L_1,\, n_2^{obs},\, L_2)$ starts at $(n_1, 0, n_2, 0)$. When this process is at time t and the current state is $(p, a_1, \ell_1, a_2, \ell_2)$, this process evolves as

$$(a_1,\ell_1) \to \begin{cases} (a_1-1,\ell_1) & \text{at rate} \quad \dfrac{1}{p}\dbinom{a_1}{2}\dfrac{1}{\rho^2(T-t)} \qquad \text{coalescence of type } 1^{obs} \\[2ex] (a_1-1,\ell_1+1) & \text{at rate} \quad \dfrac{1-p}{p}\, a_1 \, \dfrac{\theta_1}{2\rho(T-t)} \qquad \text{mutation of } 1^{obs} \text{ to } 2 \end{cases} \tag{53}$$

and, independently,

$$(a_2,\ell_2) \to \begin{cases} (a_2-1,\ell_2) & \text{at rate} \quad \dfrac{1}{1-p}\dbinom{a_2}{2}\dfrac{1}{\rho^2(T-t)} \qquad \text{coalescence of type } 2^{obs} \\[2ex] (a_2-1,\ell_2+1) & \text{at rate} \quad \dfrac{p}{1-p}\, a_2 \, \dfrac{\theta_2}{2\rho(T-t)} \qquad \text{mutation of } 2^{obs} \text{ to } 1 \end{cases} \tag{54}$$

Note the term $\rho(T-t)$ in (53)-(54) indicates the dependence of the conditional ancestral process on the demographic function. Nonetheless, Proposition 2 still gives a description for K_1 and K_2 in terms of Bernoulli random variables, like (27) and (29) respectively. For example, under $\mathbb{P}_{\mathbf{n}}$, (27) still holds but (28) needs to be modified. Indeed, given $\{(\tau_i, p_{\tau_i})\}_{i=1}^{n_1-1}$, the random variables $\{\xi_k(\cdot)\}_{k=2}^{n_1}$ are independent and

$$\xi_{n_1-i+1}(p_{\tau_i}) = \begin{cases} 0 & \text{with probability} \quad \dfrac{n_1 - i}{(1-p_{\tau_i})\theta_1\rho(T-\tau_i) + n_1 - i} \\[2ex] 1 & \text{with probability} \quad \dfrac{(1-p_{\tau_i})\theta_1\rho(T-\tau_i)}{(1-p_{\tau_i})\theta_1\rho(T-\tau_i) + n_1 - i} \end{cases} \tag{55}$$

which further generalizes $\xi_k(p)$ in (28) to include $\rho(t)$. An analogous description holds for K_2.

Remark 3. A more explicit description for the process $(X_t)_{t\in[0,T]}$ under $\mathbb{P}_{\mathbf{n}}$, hence also that for its time-reversal $(p_t)_{t\in[0,T]}$, can be obtained by Doob's h-transform (Doob, 1957, 2001). More precisely, we define the function

$$h(t,x) := \mathbb{P}(\mathbf{n} \,|\, X_t = x) = \mathbb{E}\left[\binom{n}{n_1} X_T^{n_1}(1-X_T)^{n_2} \,\middle|\, X_t = x\right].$$

Then $(X_t)_{t\in[0,T]}$ under the conditional probability $\mathbb{P}_{\mathbf{n}}$ solves the SDE

$$dX_t = \left[b(t,X_t) + \sigma^2(t,X_t)\,\partial_x \log h(t,X_t)\right] dt + \sigma(t,X_t)\, d\widetilde{W}_t,$$

where $b(t,x) := \frac{\theta_1}{2\rho(t)}(1-x) - \frac{\theta_2}{2\rho(t)}x + \frac{\alpha}{2\rho(t)}x(1-x)$ and $\sigma(t,x) = \sqrt{\frac{x(1-x)}{\rho^2(t)}}$ are the coefficients in (48), and $\widetilde{W}$ is a Brownian motion. Sufficient conditions on the function ρ for which the process $(p_t)_{t\in[0,T]}$ satisfies a stochastic differential equation may be deduced from an integration by parts argument as in Millet et al. (1989).

Next, we look at asymptotics. The following analogue of Lemma 1 holds for any initial distribution μ_0 of $X(0)$ and any demographic function ρ that is bounded and positive. Note that in Lemma 1, $\mu_0 = \phi_\alpha$ depends on α, but here μ_0 is fixed.

Proposition 3. *Let $T > 0$ and $\mu_0 \in \mathcal{P}([0,1])$ be fixed, and the demographic function ρ be as in Lemma 7. The following convergences in $\mathcal{P}([0,1])$ hold.*

(i) Suppose n_2 is fixed and $\alpha \to \infty$. Then $\mathcal{L}^{\mathbf{n}} \to \delta_1$.

(ii) Suppose α is fixed and $n_2 \to \infty$. Then $\mathcal{L}^{\mathbf{n}} \to \delta_0$.

(iii) Suppose $\widetilde{\alpha} \in \mathbb{R}$ is fixed and $\alpha = \widetilde{\alpha} n_2 \to \infty$. Suppose $\mathbb{P}_{\mu_0}(X_T \in dy)$ has a density $p(T,\mu_0,y)\,dy$ and there exists a large deviation rate function $\mathcal{I} : [0,1] \to [0,\infty]$ such that for each $y \in [0,1]$,

$$\frac{\log p(T,\mu_0,y)}{n_2} \to \mathcal{I}(y) \qquad \text{as } n \to \infty.$$

Suppose also that $(1-y)\,e^{\mathcal{I}(y)}$ has a unique maximum at $y_* \in [0,1]$. Then $\mathcal{L}^{\mathbf{n}} \to \delta_{y_*}$, and $\mathcal{I}(y_*) = \frac{1}{1-y_*}$.

Remark 4. The assumptions in (iii) holds when $\rho(t) \equiv 1$ and $\mu_0 = \phi_\alpha$ in (2), i.e. constant population size with stationary initial condition. In this case, $\mathcal{I}(y) = \widetilde{\alpha}\, y$ is the linear function, and $\mathcal{L}^{\mathbf{n}} = \phi_\alpha^{(n_1,n_2)}$ in (6). The rate function has a phase transition at $\widetilde{\alpha} = 1$, as shown in Lemma 1. Namely, $y_* = 0$ when $\widetilde{\alpha} \in (-\infty, 1)$ and $y_* = 1 - 1/\widetilde{\alpha}$ when $\widetilde{\alpha} \in (1, \infty)$.

Remark 5. The large deviation principle for $\mathbb{P}_{\mu_0}(X_T \in dy)$ as $n_2 \to \infty$ can be checked using the Gärtner-Ellis theorem (Dembo and Zeitouni, 2009, Theorem 2.3.6). When it holds, the rate function $\mathcal{I}$ is equal to the Legendre transform of the function

$$\Lambda(\gamma) := \lim_{n_2 \to \infty} \frac{1}{n_2} \log \mathbb{E}_{\mu_0}[e^{\gamma\, n_2\, X_T}].$$

Proof. A proof follows from that of Lemma 1. Let $f \in C_b([0,1])$, a bounded continuous function on $[0,1]$. Then

$$\mathbb{E}_{\mathbf{n}}[f(p_0)] = \int_0^1 f(y)\, \mathcal{L}^{\mathbf{n}}(dy) = \frac{\mathbb{E}_{\mu_0}[f(X_T)\, X_T^{n_1}(1-X_T)^{n_2}]}{\mathbb{E}_{\mu_0}[X_T^{n_1}(1-X_T)^{n_2}]}. \tag{56}$$

For part (i), note that if (n_1, n_2) is fixed and $\alpha \to \infty$, then $\mathbb{P}_{\mu_0}(X_T < 1 - \epsilon) \to 0$ for any $\epsilon > 0$ as in the proof of Theorem 5(i). Hence $\mathbb{E}_{\mu_0}[|f(X_T) - f(1)|] \to 0$ for any $f \in C_b([0,1])$. In particular, $\mathbb{P}_{\mu_0}(X_T \in dy) \to \delta_1$. Hence $\mathcal{L}^{\mathbf{n}}$ tends to δ_1 in $\mathcal{P}([0,1])$, as $\alpha \to \infty$.

For part (ii), note that $(1-y)^{n_2}$ has maximum at $y = 0$, and $y^{n_1}\, \mathbb{P}_{\mu_0}(X_T \in dy)$ does not depend on n_2. Hence $\mathcal{L}^{\mathbf{n}}$ tends to δ_0 in $\mathcal{P}([0,1])$, as $n_2 \to \infty$.

For part (iii), the numerator of (56) is

$$\begin{aligned}\mathbb{E}_{\mu_0}[f(X_T)\, X_T^{n_1}(1-X_T)^{n_2}] &= \int_0^1 f(y)\, y^{n_1}(1-y)^{n_2}\, \mathbb{P}_{\mu_0}(X_T \in dy) \\ &\approx \int_0^1 f(y)\, y^{n_1}\, [(1-y)\, e^{\mathcal{I}(y)}]^{n_2}\, \varphi(y)\, dy,\end{aligned}$$

for some function φ such that $\frac{\ln \varphi(y)}{n_2} \to 0$, by assumptions of part (iii). Since $(1-y)\, e^{\mathcal{I}(y)}$ has a unique maximum at y_*, $\lim_{n\to\infty} \mathbb{E}_{\mathbf{n}}[f(p_0)] = f(y_*)$ by (56) and a standard argument as in the proof of Lemma 1. □

By Proposition 2 and Proposition 3, similar limiting pictures for the conditional coalescence process for scenarios (i) and (ii) hold for any positive function $\rho : [0,T] \to (0,\infty)$ that is *continuous near the current time* T. Note that ρ is bounded away from zero on any compact time interval $[0,T]$ and therefore analogous approximations for the frequency process $(p_t)_{t\in[0,T]}$ under $\mathbb{P}_{\mathbf{n}}$ (such as Lemma 5) still hold, where the new approximating functions now involve the h function in Remark 3.

More precisely, in scenario (i), Theorem 1 still holds, and Theorem 2 still holds but with a possibly different limiting random variable. Hence $s_{n_2} \sim O(1/\alpha)$ is very small and $p_{\tau_i} \sim 1$, so $K_1 \to 1$ by (55), and the single latent mutation for type 1 is very old. The population size may have changed by the time this ancient mutation occurs, so the approximation in Lemma 5 will be replaced by an analogous statement that involves the h function in Remark 3. In scenario (ii), $\tau_{n_1} \sim O(1/n)$ is very small, and recalling that $\rho(T) = 1$, we have $K_1 \approx 1 + \sum_{k=2}^{n_1} \xi_k$ by (55), where $\{\xi_k\}$ are independent Bernoulli variables taking values in $\{0,1\}$ and having means $\frac{\theta_1}{\theta_1+k-1}$. Theorem

3 with $\alpha \in \mathbb{R}$ fixed still holds, but the statement needs to be modified because the approximating process Z in (42) will be replaced by another one that involves the h function in Remark 3, as was needed in modifying Lemma 5.

Scenario (iii) is harder to analyze and we leave it for future work. We conjecture that if $y_* = 0$, then the conditional genealogy behaves like scenario (ii); and if $y_* \in (0, 1]$, then the conditional genealogy behaves like scenario (i).

4. Conditional ancestral selection graph

Our aim in this section is to see how the results of the previous section can be obtained from a different model: the ancestral selection graph. We concentrate on the ancestry of focal allele A_1 and on constant population size, and proceed more heuristically than in the previous section.

The ancestral selection graph is an augmented coalescent model for the joint distribution of the gene genealogy and the allelic states of the sample (Krone and Neuhauser, 1997; Neuhauser and Krone, 1997). It includes the usual coalescent rate 1 per pair of lineages and mutation rate $\theta/2$ per lineage. Additionally, under the stationary model of Section 1, it includes a *branching* rate of $|\alpha|/2$ per lineage. When a branching event occurs, the lineage splits into an *incoming* lineage and a *continuing* lineage. One of these is *real*, meaning it is included in the gene genealogy. The other is *virtual*, meaning it is there only to model the gene genealogy correctly with selection. Which is which could be resolved if their allelic states were known: the incoming lineage is real if its allelic type is the one favored by selection, otherwise the continuing lineage is real. But the allelic states are not known in the construction of the ancestral selection graph.

The conditional ancestral selection graph models gene genealogies given a sample with allelic states specified (Slade, 2000a,b; Fearnhead, 2001, 2002; Stephens and Donnelly, 2003; Baake and Bialowons, 2008). In this case it is known which lineages are real and which are virtual. This allows a simplification in which there is a reduced rate of branching and only virtual lineages of the disfavored type are produced (Slade, 2000a). A second simplification is possible if mutation is parent-independent: then any lineage which mutates may be discarded (Fearnhead, 2002).

We assume parent-independent mutation, specifically $\theta_1 = \theta\pi_1$ and $\theta_2 = \theta\pi_2$, with $\pi_1 + \pi_2 = 1$. Any two-allele mutation model can be restated in this way, leaving the stationary probability density (2) and the sampling probability (4) unchanged. But doing so introduces 'spurious mutations to one's own type' (Donnelly, 1986) or 'empty mutations' (Baake and Bialowons, 2008) which occur only in the model and do not correspond to a biological process. These are not latent mutations. Including them allows us to discard real A_2 lineages and any virtual lineage once these mutate, but we must distinguish between empty and actual mutations in the ancestry of A_1.

The resulting conditional process tracks the numbers real and virtual ancestral lineages from the present time $t = 0$ back into the past. Let $r_1(t)$, $r_2(t)$ and $v_i(t)$, where $i = 1$ if $\alpha < 0$ or $i = 2$ if $\alpha > 0$, be the numbers of real type-1, real type-2 and virtual type-i lineages at past time t. The process begins in state $r_1(0) = n_1$, $r_2(0) = n_2$, $v_i(0) = 0$ and stops the first time $r_1(t) + r_2(t) = 1$. We suppress t in what follows, and focus on the instantaneous rates of transition of the process.

The conditional ancestral process is obtained by considering rates of events in the unconditional process, which has total rate $(r_1 + r_2 + v_i)(\theta + |\alpha| + r_1 + r_2 + v_i - 1)/2$, then weighting rates of events depending on how likely they are to produce the sample. Rates of some events are down-weighted to zero. For instance, the sample could not have been obtained if there were a coalescent event between lineages with different allelic types, whereas in the unconditional process these happen with rate $r_1 r_2$ plus either $r_2 v_1$ or $r_1 v_2$, depending on whether $\alpha < 0$ or $\alpha > 0$.

Rates of events for which the sample has a non-zero chance of being observed are up-weighted or down-weighted by ratios of sampling probabilities like (4). This method of conditioning a Markov process on its eventual outcome is stated simply in Kemeny and Snell (1960, p. 64), a familiar example being the Wright-Fisher diffusion conditioned on eventual fixation (Ewens, 2004, p. 89), and is characterized more generally by Doob's h-transform (Doob, 1957, 2001). In the conditional ancestral selection graph, the Markov process is the (unconditional) ancestral process of Krone and Neuhauser (1997) and the eventual outcome is the sample with allelic states specified.

In our formulation, the samples and their ancestral lineages all are distinguishable, which we denote with a subscript "o" for ordered as in Wakeley et al. (2023). The probability of any particular allelic configuration in the ancestry of the sample, in which there are r_1 lineages of type 1, r_2 lineages of type 2 and v_i lineages of type $i \in \{1, 2\}$, is

$$q_o(r_1, r_2, v_1) = \int_0^1 x^{r_1+v_1}(1-x)^{r_2}\phi_\alpha(x)dx \quad \text{if } \alpha < 0 \tag{57}$$

$$q_o(r_1, r_2, v_2) = \int_0^1 x^{r_1}(1-x)^{r_2+v_2}\phi_\alpha(x)dx \quad \text{if } \alpha > 0 \tag{58}$$

with $\phi_\alpha(x)$ as in (2). Note, the additional binomial coefficient in the sampling probability (4) is the number of possible orderings of a sample containing n_1 and n_2 copies of A_1 and A_2.

The rate of any particular event with non-zero probability in the conditional process is the product of its rate in the unconditional process and a ratio of sampling probabilities from either (57) or (58). For event $(r_1, r_2, v_i) \to (r_1', r_2', v_i')$, the required ratio is $q_o(r_1', r_2', v_i')/q_o(r_1, r_2, v_i)$. The denominator $q_o(r_1, r_2, v_i)$ is the probability of the sample given all events so far in the conditional ancestral process, which have led to the current state (r_1, r_2, v_i), and the numerator $q_o(r_1', r_2', v_i')$ is the probability of the sample given these events and the event $(r_1, r_2, v_i) \to (r_1', r_2', v_i')$. Appendix C provides the details of how the minimal ancestral process we use here to model latent mutations in the ancestry of the sampled copies of allele A_1 is obtained from the full conditional ancestral process, using the simplifications of Slade (2000a) and Fearnhead (2002) that we have just described.

The resulting conditional ancestral process differs depending on whether $\alpha < 0$ or $\alpha > 0$ but in either case it includes five possible transitions from state (r_1, r_2, v_i). If $\alpha < 0$,

$$(r_1, r_2, v_1) \to \begin{cases} (r_1-1, r_2+1, v_1) & \text{at rate} \quad r_1\frac{\theta\pi_1}{2}\frac{q_o(r_1-1,r_2+1,v_1)}{q_o(r_1,r_2,v_1)} \\ (r_1-1, r_2, v_1) & \text{at rate} \quad \binom{r_1}{2}\frac{q_o(r_1-1,r_2,v_1)}{q_o(r_1,r_2,v_1)} \\ (r_1, r_2, v_1+1) & \text{at rate} \quad (r_1+r_2+v_1)\frac{|\alpha|}{2}\frac{q_o(r_1,r_2,v_1+1)}{q_o(r_1,r_2,v_1)} \\ (r_1, r_2, v_1-1) & \text{at rate} \quad \left(v_1\frac{\theta\pi_1}{2} + r_1v_1 + \binom{v_1}{2}\right)\frac{q_o(r_1,r_2,v_1-1)}{q_o(r_1,r_2,v_1)} \\ (r_1, r_2-1, v_1) & \text{at rate} \quad \left(r_2\frac{\theta\pi_2}{2} + \binom{r_2}{2}\right)\frac{q_o(r_1,r_2-1,v_1)}{q_o(r_1,r_2,v_1)} \end{cases} \tag{59}$$

whereas if $\alpha > 0$,

$$(r_1, r_2, v_2) \to \begin{cases} (r_1-1, r_2+1, v_2) & \text{at rate} \quad \frac{r_1\theta\pi_1}{2}\frac{q_o(r_1-1,r_2+1,v_2)}{q_o(r_1,r_2,v_2)} \\ (r_1-1, r_2, v_2) & \text{at rate} \quad \binom{r_1}{2}\frac{q_o(r_1-1,r_2,v_2)}{q_o(r_1,r_2,v_2)} \\ (r_1, r_2, v_2+1) & \text{at rate} \quad (r_1+r_2+v_2)\frac{\alpha}{2}\frac{q_o(r_1,r_2,v_2+1)}{q_o(r_1,r_2,v_2)} \\ (r_1, r_2, v_2-1) & \text{at rate} \quad \left(v_2\frac{\theta\pi_2}{2} + r_2v_2 + \binom{v_2}{2}\right)\frac{q_o(r_1,r_2,v_2-1)}{q_o(r_1,r_2,v_2)} \\ (r_1, r_2-1, v_2) & \text{at rate} \quad \left(r_2\frac{\theta\pi_2}{2} + \binom{r_2}{2}\right)\frac{q_o(r_1,r_2-1,v_2)}{q_o(r_1,r_2,v_2)} \end{cases} \tag{60}$$

which differ owing to the different resolutions of branching events when $\alpha < 0$ versus $\alpha > 0$. We may note that the total rates of events in (59) and (60) are less than in the unconditional ancestral process because the conditional process has a reduced rate of branching (Slade, 2000a) and because empty mutations do not change the number or types of ancestral lineages. If $\alpha < 0$, the total rate is $r_1\theta\pi_1/2 + r_2|\alpha|/2$ less, whereas if $\alpha > 0$, it is $r_1\theta\pi_1/2 + r_1\alpha/2$ less.

Asymptotic approximations for the ratios $q_o(r_1', r_2', v_i')/q_o(r_1, r_2, v_i)$ in these rates of events can be obtained using the results in Appendix A. In the following three subsections we present approximations to the conditional ancestral process for our three scenarios of interest: (i) $|\alpha|$ large with n_2 fixed, (ii) n_2 large with α fixed, and (iii) both $|\alpha|$ and n_2 large with $\widetilde{\alpha} = \alpha/n_2$ fixed. Because initially $r_2 = n_2$, we consider r_2 large in the scenarios with n_2 large. For each scenario, we compute the rates events up to leading order in $|\alpha|$ or r_2, then consider how these conform to the corresponding results of Section 3.

4.1. Scenario (i): strong selection, arbitrary sample size

Here $|\alpha|$ is large with n_2 fixed, along with n_1 and θ. In Section 3.1, Theorem 1, we treated the ancestries of A_1 and A_2 simultaneously as $\alpha \to +\infty$, so that A_1 was favored and A_2 was disfavored. Here we cover these same two possibilities by modeling the ancestry of A_1, using (59) when A_1 is disfavored ($\alpha < 0$) and (60) when A_1 is favored ($\alpha > 0$). We disregard the ancestry of the non-focal allele A_2 except insofar as it is needed to model events in the ancestry of A_1.

When $\alpha < 0$, using (A.4a) in (59) gives

$$(r_1, r_2, v_1) \to \begin{cases} (r_1-1, r_2+1, v_1) & \text{at rate} \quad r_1\frac{|\alpha|}{2}\frac{\theta\pi_1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1-1, r_2, v_1) & \text{at rate} \quad r_1\frac{|\alpha|}{2}\frac{r_1-1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1, r_2, v_1+1) & \text{at rate} \quad \frac{(r_1+r_2+v_1)(\theta\pi_1+r_1+v_1)}{2} + O\big(|\alpha|^{-1}\big) \\ (r_1, r_2, v_1-1) & \text{at rate} \quad v_1\frac{|\alpha|}{2}\frac{\theta\pi_1+2r_1+v_1-1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1, r_2-1, v_1) & \text{at rate} \quad r_2\frac{\theta\pi_2}{2} + \binom{r_2}{2} + O\big(|\alpha|^{-1}\big) \end{cases} \tag{61}$$

for the ancestry of a disfavored allele under strong selection (as $\alpha \to -\infty$). Latent mutations and coalescent events occur with rates proportional to $|\alpha|$. Virtual lineages are removed similarly quickly but are produced at a much lower rate. So v_1 will stay zero during the $O(1/|\alpha|)$ time it takes for the requisite n_1 latent mutations or coalescent events to occur. Then the analogous result to (36), namely (7) and (8), follows from the first two lines of (61). Coalescence and mutation among the

copies of A_2 occur at the slower rate, so none of these should occur before all the type-1 lineages disappear. These results were first suggested in Wakeley (2008).

When $\alpha > 0$, using (A.4b) in (60) gives

$$(r_1, r_2, v_2) \to \begin{cases} (r_1-1, r_2+1, v_2) & \text{at rate} \quad r_1 \frac{1}{\alpha} \frac{\theta\pi_1(\theta\pi_2+r_2+v_2)}{2} + O\big(\alpha^{-2}\big) \\ (r_1-1, r_2, v_2) & \text{at rate} \quad \binom{r_1}{2} + O\big(\alpha^{-1}\big) \\ (r_1, r_2, v_2+1) & \text{at rate} \quad \frac{(r_1+r_2+v_2)(\theta\pi_2+r_2+v_2)}{2} + O\big(\alpha^{-1}\big) \\ (r_1, r_2, v_2-1) & \text{at rate} \quad v_2 \frac{\alpha}{2} \frac{\theta\pi_2+2r_2+v_2-1}{\theta\pi_2+r_2+v_2-1} + O(1) \\ (r_1, r_2-1, v_2) & \text{at rate} \quad r_2 \frac{\alpha}{2} \frac{\theta\pi_2+r_2-1}{\theta\pi_2+r_2+v_2-1} + O(1) \end{cases} \tag{62}$$

for the ancestry of a favored allele under strong selection (as $\alpha \to +\infty$). Now A_2 is undergoing the fast process just described for A_1 in (61), so these lineages will disappear quickly. Again the rate of removal of virtual lineages greatly exceeds their rate of production. In $O(1/\alpha)$ time, the ancestral state will become $(r_1, r_2, v_2) = (n_1, 0, 0)$. But now with A_1 favored, the rates of coalescence and latent mutation differ by a factor of α, so the first $n_1 - 1$ events will be coalescent events, followed by a long wait for a single latent mutation with rate $\theta^2\pi_1\pi_2/(2\alpha)$ as in Theorem 2.

4.2. Scenario (ii): arbitrary selection, large sample size

Here n_2 is large with α fixed, along with n_1 and θ. Because $r_2 = n_2$ at the start of the ancestral process, we present rates of events to leading order in $1/r_2$. In Section 3.2 we deferred this scenario to Section 3.3, because in the limit it is equivalent to $\widetilde{\alpha} = 0$. Of course, there are two ways for $\widetilde{\alpha}$ to approach zero, and the sign of $\widetilde{\alpha}$ matters in (42) for any $\widetilde{\alpha}$ not strictly equal to zero. Here we consider the two cases, $\alpha < 0$ and $\alpha > 0$, separately.

When $\alpha < 0$, using (A.5) in (59) gives

$$(r_1, r_2, v_1) \to \begin{cases} (r_1-1, r_2+1, v_1) & \text{at rate} \quad r_1 \frac{r_2}{2} \frac{\theta\pi_1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1-1, r_2, v_1) & \text{at rate} \quad r_1 \frac{r_2}{2} \frac{r_1-1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1, r_2, v_1+1) & \text{at rate} \quad \frac{|\alpha|(\theta\pi_1+r_1+v_1)}{2} + O\big(r_2^{-1}\big) \\ (r_1, r_2, v_1-1) & \text{at rate} \quad v_1 \frac{r_2}{2} \frac{\theta\pi_1+2r_1+v_1-1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1, r_2-1, v_1) & \text{at rate} \quad \frac{r_2^2}{2} + O(r_2) \end{cases} \tag{63}$$

This differs from the neutral case (Wakeley et al., 2023) only by the possibility of virtual lineages. As in (61), these will be removed quickly if they are produced. The process of latent mutation and coalescence happens in $O(1/r_2)$ time, with relative rates in the first two lines of (63) again giving (7) and (8). Because $r_2 \to \infty$, this approximation will hold long enough for the required fixed number of events among the A_1 lineages to occur, despite the rapid decrease of r_2 in the last line of (63). A proof of this is given in Wakeley et al. (2023, Appendix). Theorem 4 addresses the corresponding issues for the model of Section 3.

When $\alpha > 0$, using (A.5) in (60) gives

$$(r_1, r_2, v_2) \to \begin{cases} (r_1-1, r_2+1, v_2) & \text{at rate} \quad r_1 \frac{r_2}{2} \frac{\theta\pi_1}{\theta\pi_1+r_1-1} + O(1) \\ (r_1-1, r_2, v_2) & \text{at rate} \quad r_1 \frac{r_2}{2} \frac{r_1-1}{\theta\pi_1+r_1-1} + O(1) \\ (r_1, r_2, v_2+1) & \text{at rate} \quad \frac{r_2|\alpha|}{2} + O(1) \\ (r_1, r_2, v_2-1) & \text{at rate} \quad v_2 r_2 + O(1) \\ (r_1, r_2-1, v_2) & \text{at rate} \quad \frac{r_2^2}{2} + O(r_2) \end{cases} \tag{64}$$

which differs from (63) in two ways. Now the rate of production of virtual lines is non-negligible. But here their presence does not affect the rates of latent mutation and coalescence. Again we have (7) and (8), and the process of latent mutation and coalescence happens in $O(1/r_2)$ time.

4.3. Scenario (iii): strong selection, large sample size

Here both $|\alpha|$ and n_2 are large with $\widetilde{\alpha} = \alpha/n_2$ fixed, along with n_1 and θ. Again since the process begins with $r_2 = n_2$, we present rates of events to leading order in $1/r_2$. Because the conditional ancestral process differs for $\alpha < 0$ versus $\alpha > 0$, i.e. with (59) and (60), and the asymptotic approximation we use for the hypergeometric function differs for $\widetilde{\alpha} < 1$ versus $\widetilde{\alpha} > 1$, i.e. with (A.6a) and (A.6b), here we have three cases. Note these are the same three cases in (11a), (11b) and (11c).

When $\widetilde{\alpha} < 0$, using (A.6a) in (59) gives

$$(r_1, r_2, v_1) \to \begin{cases} (r_1-1, r_2+1, v_1) & \text{at rate} \quad r_1 \frac{r_2(1+|\widetilde{\alpha}|)}{2} \frac{\theta\pi_1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1-1, r_2, v_1) & \text{at rate} \quad r_1 \frac{r_2(1+|\widetilde{\alpha}|)}{2} \frac{r_1-1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1, r_2, v_1+1) & \text{at rate} \quad \frac{|\widetilde{\alpha}|(\theta\pi_1+r_1+v_1)}{2(1+|\widetilde{\alpha}|)} + O\big(r_2^{-1}\big) \\ (r_1, r_2, v_1-1) & \text{at rate} \quad v_1 \frac{r_2(1+|\widetilde{\alpha}|)}{2} \frac{\theta\pi_1+2r_1+v_1-1}{\theta\pi_1+r_1+v_1-1} + O(1) \\ (r_1, r_2-1, v_1) & \text{at rate} \quad \frac{r_2^2}{2} + O(r_2) \end{cases} \tag{65}$$

which is comparable to (61) and (63). Again we may effectively ignore virtual lineages. The rates of latent mutation and coalescence in (61) and (63) differ only by the interchange of r_2 for $|\alpha|$. In (65), the factor $r_2(1 + |\widetilde{\alpha}|)$ encompasses the effects of both. The larger $|\widetilde{\alpha}|$ is, the more quickly these events will occur, and again (7) and (8) describe the number of latent mutations.

When $0 < \widetilde{\alpha} < 1$, using (A.6a) in (60) gives

$$(r_1, r_2, v_2) \to \begin{cases} (r_1-1, r_2+1, v_2) & \text{at rate} \quad r_1 \frac{r_2(1-\widetilde{\alpha})}{2} \frac{\theta\pi_1}{\theta\pi_1+r_1-1} + O(1) \\ (r_1-1, r_2, v_2) & \text{at rate} \quad r_1 \frac{r_2(1-\widetilde{\alpha})}{2} \frac{r_1-1}{\theta\pi_1+r_1-1} + O(1) \\ (r_1, r_2, v_2+1) & \text{at rate} \quad \frac{r_2\widetilde{\alpha}}{2} + O(1) \\ (r_1, r_2, v_2-1) & \text{at rate} \quad v_2 r_2 + O(1) \\ (r_1, r_2-1, v_2) & \text{at rate} \quad \frac{r_2^2}{2} + O(r_2) \end{cases} \tag{66}$$

which is comparable to (64). In contrast to (65), now with A_1 favored, the larger $\widetilde{\alpha}$ is (i.e. the closer it is to 1) the smaller the rates of latent mutation and coalescence become. Otherwise, for any given

$\widetilde{\alpha}$, the same conclusions regarding latent mutations and their timing follow from (66) as from (65), and these conform to what is stated in Corollary 1.

When $\widetilde{\alpha} > 1$, using (A.6b) in (60) gives

$$(r_1, r_2, v_2) \to \begin{cases} (r_1 - 1, r_2 + 1, v_2) & \text{at rate} \quad r_1 \frac{\theta \pi_1}{2} \frac{1}{\widetilde{\alpha}-1} + O\big(r_2^{-1}\big) \\ (r_1 - 1, r_2, v_2) & \text{at rate} \quad \binom{r_1}{2} \frac{\widetilde{\alpha}}{\widetilde{\alpha}-1} + O\big(r_2^{-1}\big) \\ (r_1, r_2, v_2 + 1) & \text{at rate} \quad \frac{r_2}{2} + O(1) \\ (r_1, r_2, v_2 - 1) & \text{at rate} \quad v_2 r_2 \widetilde{\alpha} + O(1) \\ (r_1, r_2 - 1, v_2) & \text{at rate} \quad \frac{r_2^2}{2} \widetilde{\alpha} + O(r_2) \end{cases} \tag{67}$$

which paints a very different picture. Whereas (61), (63), (64), (65) and (66) all give the Ewens sampling result described by (7) and (8) and have these events occurring quickly on the coalescent time scale, (67) is rather like (62) in that the rates of latent mutation and coalescence are too slow to register on the time scale of events involving the non-focal allele A_2. The overwhelmingly most frequent events in (67) will be coalescent events between A_2 lineages at rate $\propto r_2^2$, so an effectively instantaneous transition will occur from r_2 large to r_2 comparable to r_1. Then this case (67) will collapse quickly to the corresponding case (62) where coalescence without mutation will happen among the A_1 followed by a long wait for a single latent mutation. For the model in Section 3.3.2, this is described by Theorem 5 and Theorem 6. Finally we may note that initially the rates of latent mutation and coalescence in (67) are precisely those predicted for the model in Section 3.3.2 from (23) starting at $p_0 \to 1 - 1/\widetilde{\alpha}$ as specified for $\widetilde{\alpha} \in (1, \infty)$ in (31).

5. Discussion

In this paper, we have considered a two allele model at a single genetic locus subject to recurrent mutation and selection in a large haploid population with possibly time-varying size. We assumed that a sample of size n was drawn uniformly from the infinite population under the diffusion approximation. By extending the framework of Barton et al. (2004), we described the asymptotic behaviors of the conditional genealogy and the number of latent mutations of the sample, given the sample frequencies of the two alleles. This moves beyond what is in Wakeley et al. (2023) by the inclusion of selection and by the use of an entirely different model, i.e. coalescence in a random background (Barton et al., 2004). This yields novel results. For example, in the strong selection case in which the selection strength α is proportional to the sample size n and both go to infinity (our scenario (iii)), the genealogy of the rare allele can be described in terms of a Cox-Ingersoll-Ross (CIR) diffusion with an initial Gamma distribution.

The concept of rare allele in this paper and in Wakeley et al. (2023) is the same as the one considered by Joyce and Tavaré (1995) and Joyce (1995). It focuses on the counts of the alleles in a large sample rather than their relative frequencies in the population. In scenarios (ii) and (iii) we consider a fixed number n_1 of the rare type 1 when the sample size n tends to infinity. Joyce and Tavaré (1995) considered rare alleles in a large sample drawn from the stationary distribution of a d-dimensional Wright-Fisher diffusion with selection and mutation. They showed that the counts of rare alleles, from different latent mutations in our terminology, have approximately independent Poisson distributions with parameters that do *not* depend on the selection parameters, and that the

Ewens sampling formula describes their distribution. Their model with $d = 2$ and genic selection corresponds to our scenario (ii). Our results for very strong selection ($\alpha \to \infty$) in scenario (iii) differ from those of Joyce and Tavaré (1995) in that the rare-allele sampling probabilities (11a), (11b) and (11c) do depend on selection. Interestingly, the number of latent mutations given n_1 still follows the Ewens sampling formula when $\lim_{n\to\infty} \alpha/n \in (-\infty, 1)$. But this is not true when $\lim_{n\to\infty} \alpha/n \in (1, \infty)$, in which case the number of latent mutations is always $k_1 \equiv 1$.

Some of our results for rare alleles have empirical relevance, specifically those for scenario (ii) including their robustness to time-varying population size demonstrated in Section 3.4, and those for scenario (iii) with $\widetilde{\alpha} < 0$. In scenario (ii), as n increases for fixed but arbitrary α, the distributions of latent mutations and the ages of those latent mutations become identical to those for neutral alleles described in Wakeley et al. (2023). Our results also show that selection does have an effect in this case, but it is only to raise or lower the rare-allele sampling probability (10) by the constant factor C for every value of n_1. This relative insensitivity to selection suggests confidence in using rare alleles for demographic inference and genome-wide association studies (O'Connor et al., 2015; Nait Saada et al., 2020; Zaidi and Mathieson, 2020). Slatkin and Rannala (1997b), who obtained the Ewens sampling formula result for rare deleterious alleles by assuming they evolve independently according to a linear birth-death process, cf. Slatkin and Rannala (1997a), suggested that deviations from this neutral prediction at two human-disease-associated loci were due to population growth. Reich and Lander (2001) made a similar argument for a number of other disease-associated loci starting from the mutation-selection balance model of Hartl and Campbell (1982) and Sawyer (1983) which also gives the Ewens sampling formula result for rare disease alleles.

Our exploration of time-varying populations in Section 3.4, namely the robustness of the Ewens sampling formula result for the number of latent mutations, suggests that rare alleles may not always be well suited for demographic inference. With only a mild constraint on the trajectory of population sizes through time, increasing the sample size will eventually make the distribution of latent mutations of rare alleles look as if the population size has been constant at its current size. There is no doubt that demographic inferences improve as sample sizes increase. What Section 3.4 implies is that these improvements will not come from focusing exclusively on the lower end of sample allele frequencies (i.e. any fixed n_1 as $n \to \infty$). How relevant this is for a given sample will depend the actual ages of its latent mutations and the degree of population-size change between those times and the present. To illustrate, consider the $O(1/n)$ ages of latent mutations under the exponential growth model with rate β. If $\beta/n \ll 1$, the ancestral process of tracing back to these mutations will be complete before the population has changed much in size and the results of Section 3.4 will hold. But this is clearly not the case for the *gnomAD* data in Wakeley et al. (2023) and Seplyarskiy et al. (2023). The distribution of n_1 in the non-Finnish European sample with $n = 114K$ is well fit by $\beta/n = 3$. See for example Fig. 3 in Seplyarskiy et al. (2023). Sample sizes would need to be orders of magnitude greater for the results in Section 3.4 to hold in this case.

Scenario (iii) with $\widetilde{\alpha} < 0$ is applicable to strongly deleterious alleles. An appreciable fraction of new mutations are strongly deleterious (Eyre-Walker and Keightley, 2007; Kim et al., 2017; Weghorn et al., 2019; Dukler et al., 2022). Previous theoretical work includes Nei (1968), who found a gamma density analogous to ours in Lemma 4 but for the population allele frequency of partially recessive lethal mutations, and Charlesworth and Hill (2019), who used Nei's approximation to derive the negative binomial distribution for n_1, our (11a). In this case, (12a) shows that the sampling probabilities of rare alleles fall off quickly as n_1 grows: each additional copy of A_1 in the sample lowers its probability by a factor of $1/(1 + |\widetilde{\alpha}|)$ compared to the neutral case. Even so, the

distribution of k_1 given n_1 follows the Ewens sampling formula. Hartl and Campbell (1982) and Sawyer (1983) obtained similar results by assuming that both selection and mutation are strong. Our analysis of scenario (iii) with $\widetilde{\alpha} < 0$ also shows that latent mutations of rare strongly deleterious alleles are especially young: selection speeds up the ancestral process of latent mutation by a factor of $1 + |\widetilde{\alpha}|$ on top of the factor of n already present under neutrality. This is most easily seen by comparing the first two lines of (65) to the first two lines of (63).

Our results for scenario (i) with $\alpha < 0$, which hold as $\alpha \to -\infty$ for arbitrary sample size and alleles at any sample frequencies, are also applicable to strongly deleterious alleles. They are similar to the results just discussed for scenario (iii) with $\widetilde{\alpha} < 0$. We expect that our results for very strong positive selection, i.e. scenario (i) with $\alpha > 0$ and scenario (iii) with $\widetilde{\alpha} > 0$, will be of limited applicability. Mutations to strongly positively selected alleles are uncommon and observing such an allele a small number of times in a very large sample would be exceedingly unlikely.

Many open questions remain. Joyce (1995) obtained a result similar to that of Joyce and Tavaré (1995), for a Wright-Fisher diffusion with selection and infinite-alleles mutation. This diffusion process is a particular case of the Fleming-Viot process (see Ethier and Kurtz (1993) for a review) and it has a unique stationary distribution denoted ν_{selec}. Joyce (1995) considered a large sample of size n drawn from ν_{selec}. Let $C_b(n) \in \mathbb{Z}_+^b$ be the first b allele counts in a sample of size n drawn from the stationary distribution, and K_n be the total number of alleles in the sample. Joyce (1995) showed that for any fixed b, the distribution of $(C_b(n), K_n)$ under ν_{selec} is arbitrarily close to that under the neutral model. It would be interesting to know if analogous results for our scenario (iii) also hold for the infinite allele model. In particular, is there a threshold for the selection strength relative to n that controls whether selection is washed out or not in the limit as $n \to \infty$?

For time-varying populations, little is known in scenario (iii). For example, will the assumptions in Proposition 3 hold for a general demographic function? Will there be a phase transition for the value of y_* in terms of $\widetilde{\alpha}$ and if so, what will determine the phase transition? Also, both our results and those of Joyce and Tavaré (1995) and Joyce (1995) are for the infinite-population diffusion limit. Further consideration of the issues raised in Section 2.1.1 is needed to assess the relevance of these results to various kinds of finite populations.

The critical case $\widetilde{\alpha} = 1$ in scenario (iii) is omitted in this paper. Results for this case are expected to lie between those of $\widetilde{\alpha} > 1$ and $\widetilde{\alpha} < 1$, and require more in-depth asymptotic analysis. For example, one can first obtain an asymptotics for the hypergeometric function in (A.6a)-(A.6b) for the case $\widetilde{\alpha} = 1$, and then follow the argument in Lemma 4 to obtain the asymptotic of the expectation $\mathbb{E}_{\mathbf{n}}[p_0]$ as $n_2 \to \infty$ in this critical case. Lemma 4 asserts that $\mathbb{E}_{\mathbf{n}}[p_0] = O(1)$ when $\widetilde{\alpha} > 1$ and $\mathbb{E}_{\mathbf{n}}[p_0] = O(1/n_2)$ when $\widetilde{\alpha} < 1$. We conjecture that $\mathbb{E}_{\mathbf{n}}[p_0] = O(n_2^{-\sigma})$ for some $\sigma \in (0, 1)$ in the critical case.

Finally, we have ignored the possibility of spatial structure. Spatially heterogeneous populations in which reproduction rates, death rates, mutation rates and selection strength can depend both on spatial position and local population density present challenges. This is because the population dynamics now take place in high or infinite dimension (Hallatschek and Nelson, 2008; Barton et al., 2010; Durrett and Fan, 2016; Louvet and Véber, 2023; Etheridge et al., 2023). For example, the spatial version of (1), the stochastic Fisher-Kolmogorov-Petrovsky-Piscunov (FKPP) equation introduced by Shiga (1988), is a stochastic partial differential equation that arises as the scaling limit of various discrete models under weak selection (Müller and Tribe, 1995; Durrett and Fan, 2016; Fan, 2021). Under the stochastic FKPP, Hallatschek and Nelson (2008) and Durrett and Fan (2016) studied the backward-time lineage dynamics of a single sample individual, conditioned on

knowing its type. It would be interesting to see if our results in this paper can be extended to spatial stochastic models with selection.

Acknowledgements

We thank Alison Etheridge for raising the question about the applicability of our limiting results to Wright-Fisher reproduction (cf. Section 2.1.1). We also thank Shamil Sunyaev, Evan Koch and Joshua Schraiber for helpful discussions, and Daniel Rickert and Kejia Geng for assistance in producing the figures. Research partially supported by National Science Foundation grants DMS-1855417 and DMS-2152103, and Office of Naval Research grant N00014-20-1-2411 to Wai-Tong (Louis) Fan.

Appendix A. Asymptotic approximations used in the text

From the series expansion for a ratio of gamma functions with a common large parameter, 6.1.47 in Abramowitz and Stegun (1964) or equation (1) in Tricomi and Erdélyi (1951), we have

$$\frac{\Gamma(a+n_2)}{\Gamma(b+n_2)} = n_2^{a-b}\left(1+\frac{(a-b)(a+b-1)}{2n_2}+O\left(n_2^{-2}\right)\right) \tag{A.1}$$

for constants a and b which will depend on the application. For example, we can apply (A.1) twice in the sampling probability (4) when n_2 is large: once with $a = n_1 + 1$ and $b = 1$ (in the binomial coefficient) and once with $a = \theta_2$ and $b = \theta_1 + \theta_2 + n_1$.

The confluent hypergeometric function is commonly defined in terms of the integral

$${}_1F_1\left(a;b;z\right) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 e^{zu}u^{a-1}(1-u)^{b-a-1}du \tag{A.2}$$

or in terms of the series

$${}_1F_1\left(a;b;z\right) = \sum_{k=0}^{\infty}\frac{a^{(k)}z^k}{b^{(k)}k!} \tag{A.3}$$

which converges for all $z \in \mathbb{R}$ and $b > a > 0$, where $a^{(k)}$ is the rising factorial $a(a+1)\cdots(a+k-1)$ with $a^{(0)} = 1$. Again a and b depend on the context, e.g. as in (3) and (4). The parameter z corresponds to the selection parameter α.

For large $|\alpha|$ and with constant a and b,

$${}_1F_1\left(a;b;\alpha\right) = \begin{cases} \dfrac{\Gamma(b)}{\Gamma(b-a)}|\alpha|^{-a}\left(1-\dfrac{a(b-a-1)}{|\alpha|}+O\left(|\alpha|^{-2}\right)\right) & \text{if } \alpha < 0 \quad \text{(A.4a)} \\ \dfrac{\Gamma(b)}{\Gamma(a)}e^{\alpha}\alpha^{a-b}\left(1-\dfrac{(b-a)(a-1)}{\alpha}+O\left(\alpha^{-2}\right)\right) & \text{if } \alpha > 0 \quad \text{(A.4b)} \end{cases}$$

where the middle, neutral case is given only for completeness. Equation (A.4a) is from (4.1.2) in Slater (1960), and (A.4b) is from (4.1.6) in Slater (1960) or may be obtained from (A.4a) using Kummer's first theorem which appears as (1.4.1) in Slater (1960).

For large n_2, with constants a, b and z,

$${}_1F_1\left(a;b+n_2;z\right) = 1+\frac{az}{n_2}+O\left(n_2^{-2}\right) \tag{A.5}$$

directly from (A.3).

For large n_2 and $\alpha = \widetilde{\alpha} n_2$, with constants a and b,

$$
{}_1F_1\left(a; b+n_2; \widetilde{\alpha} n_2\right) \approx \begin{cases} (1-\widetilde{\alpha})^{-a} & \text{if } \widetilde{\alpha} < 1 \qquad \text{(A.6a)} \\ \dfrac{\sqrt{2\pi}}{\Gamma(a)} \left(1 - \dfrac{1}{\widetilde{\alpha}}\right)^{a-1} \left(\dfrac{1}{\widetilde{\alpha}}\right)^{b-a+n_2} n_2^{a-\frac{1}{2}} e^{n_2(\widetilde{\alpha}-1)} & \text{if } \widetilde{\alpha} > 1 \qquad \text{(A.6b)} \end{cases}
$$

which we present only to leading order for simplicity. Equation (A.6a) follows from (A.3) and (A.6b) was obtained by applying Laplace's method to the integral in (A.2) for this case.

Appendix B. Proofs of Lemma 1, Lemma 4 and Lemma 7

Proof of Lemma 1. Part (ii) then follows from part (iii) with $\widetilde{\alpha} = 0$, and the proof of part (i) follows from a similar argument.

To prove part (iii), we let $a = \theta_1 + n_1$ for simplicity. The function $\phi_\alpha^{(n_1,n_2)}$ is a constant multiple of the function

$$
\begin{aligned}
x^{a-1}(1-x)^{n_2+\theta_2-1} e^{\alpha x} &= \left[(1-x)e^{\widetilde{\alpha} x}\right]^{n_2} x^{a-1}(1-x)^{\theta_2-1} e^{cx} \\
&= e^{n_2 S(x)} x^{a-1}(1-x)^{\theta_2-1} e^{cx},
\end{aligned}
$$

where the function $S : [0,1) \to \mathbb{R}$ defined by $S(x) := \widetilde{\alpha} x + \ln(1-x)$

$$
\begin{cases} \text{is strictly decreasing} & \text{when} \quad \widetilde{\alpha} \in (-\infty, 1] \\ \text{has a global maximum at } x = 1 - 1/\widetilde{\alpha} \in (0,1) & \text{when} \quad \widetilde{\alpha} \in (1, \infty) \end{cases} \tag{B.1}
$$

Part (iii) then follows from asymptotic expansion of integrals such as the Laplace method.

Let $x^* \in [0,1]$ be the global maximum of the function S. Then $x^* = 0$ when $\widetilde{\alpha} \in (-\infty, 1]$ and $x^* = 1 - 1/\widetilde{\alpha}$ when $\widetilde{\alpha} \in (1, \infty)$. Fix an arbitrary $\epsilon \in (0,1)$. There exists $\delta \in (0,1)$ small enough such that $\sup_{y \in [0,1]:\, |y - x^*| < \delta} |f(y) - f(x^*)| < \epsilon$. For each of the two cases, by (B.1), the ratio

$$
\int_{x \in [0,1]:\, |x - x^*| > \delta} e^{n_2 S(x)} x^{a-1}(1-x)^{\theta_2-1} e^{cx}\, dx \Big/ \int_0^1 e^{n_2 S(x)} x^{a-1}(1-x)^{\theta_2-1} e^{cx}\, dx \to 0 \tag{B.2}
$$

as $n_2 \to \infty$. For any $f \in C_b([0,1])$,

$$
\begin{aligned}
&\left| \int_0^1 f(x) \phi_\alpha^{(n_1,n_2)}(x)\, dx - f(x^*) \right| \\
\le &\left| \int_{x \in [0,1]:\, |x - x^*| > \delta} f(x) \phi_\alpha^{(n_1,n_2)}(x)\, dx \right| + \left| \int_{x \in [0,1]:\, |x - x^*| \le \delta} f(x) \phi_\alpha^{(n_1,n_2)}(x)\, dx - f(x^*) \right| \\
\le &\|f\| \int_{x \in [0,1]:\, |x - x^*| > \delta} \phi_\alpha^{(n_1,n_2)}(x)\, dx + \epsilon + |f(x^*)| \int_{x \in [0,1]:\, |x - x^*| > \delta} \phi_\alpha^{(n_1,n_2)}(x)\, dx.
\end{aligned}
$$

Hence by (B.2), $\limsup_{n_2 \to \infty} \left| \int_0^1 f(x) \phi_\alpha^{(n_1,n_2)}(x)\, dx - f(x^*) \right| \le \epsilon$. Since $\epsilon > 0$ is arbitrary, we have shown that $\left| \int_0^1 f(x) \phi_\alpha^{(n_1,n_2)}(x)\, dx - f(x^*) \right| \to 0$ as $n_2 \to \infty$. $\square$

Proof of Lemma 4. Convergence in distribution to a constant is equivalent to convergence in probability. Hence Lemma 4, except the last statement about the convergence in distribution of $n_2 p_0$, follows from Lemma 1. As in the main text, $A \approx B$ below means $A/B \to 1$ in the specified limit.

When $\widetilde{\alpha} \in (-\infty, 1)$, we let $a = n_1 + \theta_1$ for simplicity. The probability density function of np_0 under $\mathbb{P}_{\mathbf{n}}$ is

$$\begin{aligned}
\frac{1}{n}\phi_{\alpha}^{(n_1,n_2)}\left(\frac{y}{n}\right) &= \frac{1}{n}\,\frac{1}{\text{alpha}(a, n_2+\theta_2)\,{}_1F_1\left(a; n+\theta_1+\theta_2; \alpha\right)}\,\left(\frac{y}{n}\right)^{a-1}\left(1-\left(\frac{y}{n}\right)\right)^{n_2+\theta_2-1} e^{\alpha\frac{y}{n}} \\
&\approx \frac{1}{n}\,\frac{1}{\text{alpha}(a, n_2+\theta_2)\,{}_1F_1\left(a; n+\theta_1+\theta_2; \widetilde{\alpha} n_2\right)}\,\left(\frac{y}{n}\right)^{a-1} e^{-y}\, e^{\widetilde{\alpha} y} \\
&\approx \frac{1}{n^a}\,\frac{1}{\text{alpha}(a, n_2+\theta_2)\,(1-\widetilde{\alpha})^{-a}}\, y^{a-1}\, e^{-y}\, e^{\widetilde{\alpha} y} \\
&\approx \frac{1}{\Gamma(a)\,(1-\widetilde{\alpha})^{-a}}\, y^{a-1}\, e^{-y}\, e^{\widetilde{\alpha} y}
\end{aligned}$$

as $n_2 \to \infty$, where we used (A.6a) and then (A.1) in the last two approximations above. Hence the probability density function of $n_2 p_0$ (under $\mathbb{P}_{\mathbf{n}}$) converges pointwise to that of the $\text{Gam}(n_1+\theta_1, 1-\widetilde{\alpha})$ random variable. This implies the desired convergence in distribution. □

Proof of Lemma 7. Fix $t \in \mathbb{R}_+$ and let $k = [N(N-1)t/2]$. Suppose $A(k)$ is the number of type 1 at step k of the discrete-time Moran process. Direct calculations from (13) and (14) show that, as $N \to \infty$,

$$\mathbb{E}\left[\frac{A(k+1)-A(k)}{\rho(t)N} \,\Big|\, \frac{A(k)}{\rho(t)N} = x\right] \approx b(t,x)\frac{2}{N^2}$$

and

$$\mathbb{E}\left[\left(\frac{A(k+1)-A(k)}{\rho(t)N}\right)^2 \,\Big|\, \frac{A(k)}{\rho(t)N} = x\right] \approx \sigma^2(t,x)\frac{2}{N^2},$$

where $b(t,x) := \frac{\theta_1}{2\rho(t)}(1-x) - \frac{\theta_2}{2\rho(t)}x + \frac{\alpha}{2\rho(t)}x(1-x)$ and $\sigma(t,x) = \sqrt{\frac{x(1-x)}{\rho^2(t)}}$ are the coefficients in (48). The condition on ρ guarantees that the SDE (48) has a unique weak solution and that the desired weak convergence follows from standard (martingale problem) method; for reference see Stroock and Varadhan (1979, Chapter 11). □

Appendix C. Events in the conditional ancestral selection graph

Here we show how the minimal conditional ancestral process in Section 4 is obtained from the full conditional ancestral process. To begin, we assume that at some time in the conditional ancestral process there were r_1, r_2, v_1 and v_2 real and virtual lineages of type 1 and type 2. The associated sampling probability is $q_o(r_1, r_2, v_1, v_2)$, the straightforward extension of (57) or (58) to include both type-1 and type-2 virtual lineages. How branching events are resolved depends on which allele is favored by selection. We begin here by assuming that A_2 is favored, or $\alpha < 0$. Grouping events

by the types of lineages involved (real or virtual of type 1 or type 2) then by whether it is mutation, branching or coalescence gives fourteen possibilities which occur at the following rates.

$$r_1 \frac{\theta\pi_1}{2}\left(\frac{q_o(r_1,r_2,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)} + \frac{q_o(r_1-1,r_2+1,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)}\right) \tag{C.1}$$

$$r_2 \frac{\theta\pi_2}{2}\left(\frac{q_o(r_1,r_2,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)} + \frac{q_o(r_1+1,r_2-1,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)}\right) \tag{C.2}$$

$$v_1 \frac{\theta\pi_1}{2}\left(\frac{q_o(r_1,r_2,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)} + \frac{q_o(r_1,r_2,v_1-1,v_2+1)}{q_o(r_1,r_2,v_1,v_2)}\right) \tag{C.3}$$

$$v_2 \frac{\theta\pi_2}{2}\left(\frac{q_o(r_1,r_2,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)} + \frac{q_o(r_1,r_2,v_1+1,v_2-1)}{q_o(r_1,r_2,v_1,v_2)}\right) \tag{C.4}$$

$$r_1 \frac{|\alpha|}{2}\frac{q_o(r_1,r_2,v_1+1,v_2)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.5}$$

$$r_2 \frac{|\alpha|}{2}\left(2\frac{q_o(r_1,r_2,v_1+1,v_2)}{q_o(r_1,r_2,v_1,v_2)} + \frac{q_o(r_1,r_2,v_1,v_2+1)}{q_o(r_1,r_2,v_1,v_2)}\right) \tag{C.6}$$

$$v_1 \frac{|\alpha|}{2}\frac{q_o(r_1,r_2,v_1+1,v_2)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.7}$$

$$v_2 \frac{|\alpha|}{2}\left(2\frac{q_o(r_1,r_2,v_1+1,v_2)}{q_o(r_1,r_2,v_1,v_2)} + \frac{q_o(r_1,r_2,v_1,v_2+1)}{q_o(r_1,r_2,v_1,v_2)}\right) \tag{C.8}$$

$$\binom{r_1}{2}\frac{q_o(r_1-1,r_2,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.9}$$

$$\binom{r_2}{2}\frac{q_o(r_1,r_2-1,v_1,v_2)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.10}$$

$$r_1 v_1 \frac{q_o(r_1,r_2,v_1-1,v_2)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.11}$$

$$r_2 v_2 \frac{q_o(r_1,r_2,v_1,v_2-1)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.12}$$

$$\binom{v_1}{2}\frac{q_o(r_1,r_2,v_1-1,v_2)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.13}$$

$$\binom{v_2}{2}\frac{q_o(r_1,r_2,v_1,v_2-1)}{q_o(r_1,r_2,v_1,v_2)} \tag{C.14}$$

The sum of (C.1) through (C.14) is equal to the total rate of events in the unconditional ancestral process, $(r_1+r_2+v_1+v_2)(\theta+|\alpha|+r_1+r_2+v_1+v_2-1)/2$. Twenty-two distinct events $(r_1,r_2,v_1,v_2) \to (r_1',r_2',v_1',v_2')$ are represented, one for each of the ratios of sampling probabilities, $q_0(r_1',r_2',v_1',v_2')/q_0(r_1,r_2,v_1,v_2)$. Note that the assumption of parent-independent mutation leads to the four kinds of spurious or empty mutation events in (C.1) through (C.4) which do not change the ancestral state of the sample ($r_1'=r_1, r_2'=r_2, v_1'=v_1, v_2'=v_2$). Also, only those events which have

have non-zero probabilities of giving the data appear in (C.1) through (C.14); coalescent events between lineages with different types and type-i mutation events on type $3-i$ lineages would make the data impossible.

Recall that the resolution of branching events depends on which allele is favored by selection. The events and their probabilities in (C.5) through (C.8) are just for the case $\alpha < 0$, where A_2 is the favored allele. Each branching event creates an incoming lineage and a continuing lineage, both of which may be of type 1 or type 2. Let (I, C) be the types of these lineages. In (C.5) and (C.7), only one of the four (I, C) pairs has non-zero probability of producing the data: $(I = 1, C = 1)$ corresponding to the event $(r_1, r_2, v_1, v_2) \to (r_1, r_2, v_1 + 1, v_2)$. In (C.6) and (C.8), the possibility $(I = 1, C = 1)$ is discarded as it would then be impossible for the descendant lineage to be of type 2. The other three possibilities have non-zero chances of producing the data, and associated events

$$(r_1, r_2, v_1, v_2) \to \begin{cases} (r_1, r_2, v_1 + 1, v_2) & \text{when } (I = 1, C = 2) & \text{(C.15a)} \\ (r_1, r_2, v_1 + 1, v_2) & \text{when } (I = 2, C = 1) & \text{(C.15b)} \\ (r_1, r_2, v_1, v_2 + 1) & \text{when } (I = 2, C = 2)\ . & \text{(C.15c)} \end{cases}$$

In contrast, if $\alpha > 0$ then branching events on type-2 lineages are the ones for which only one of the four (I, C) pairs has non-zero probability of producing the data: $(I = 2, C = 2)$ corresponding to the event $(r_1, r_2, v_1, v_2) \to (r_1, r_2, v_1, v_2 + 1)$. When $\alpha > 0$, if the branching event occurs on a type-1 lineage, then in place of (C.15a), (C.15b) and (C.15c) we have

$$(r_1, r_2, v_1, v_2) \to \begin{cases} (r_1, r_2, v_1 + 1, v_2) & \text{when } (I = 1, C = 1) & \text{(C.16a)} \\ (r_1, r_2, v_1, v_2 + 1) & \text{when } (I = 1, C = 2) & \text{(C.16b)} \\ (r_1, r_2, v_1, v_2 + 1) & \text{when } (I = 2, C = 1)\ . & \text{(C.16c)} \end{cases}$$

Therefore, when $\alpha > 0$, (C.5) through (C.8) must be replaced with

$$r_1 \frac{\alpha}{2} \left(\frac{q_o(r_1, r_2, v_1 + 1, v_2)}{q_o(r_1, r_2, v_1, v_2)} + 2 \frac{q_o(r_1, r_2, v_1, v_2 + 1)}{q_o(r_1, r_2, v_1, v_2)} \right) \tag{C.17}$$

$$r_2 \frac{\alpha}{2} \frac{q_o(r_1, r_2, v_1, v_2 + 1)}{q_o(r_1, r_2, v_1, v_2)} \tag{C.18}$$

$$v_1 \frac{\alpha}{2} \left(\frac{q_o(r_1, r_2, v_1 + 1, v_2)}{q_o(r_1, r_2, v_1, v_2)} + 2 \frac{q_o(r_1, r_2, v_1, v_2 + 1)}{q_o(r_1, r_2, v_1, v_2)} \right) \tag{C.19}$$

$$v_2 \frac{\alpha}{2} \frac{q_o(r_1, r_2, v_1, v_2 + 1)}{q_o(r_1, r_2, v_1, v_2)} \tag{C.20}$$

Equations (C.1) through (C.7) and (C.9) through (C.14) are the same for $\alpha > 0$ and $\alpha < 0$.

The simplifications discovered by Slade (2000a) and Fearnhead (2002) follow from the simple fact that each sampled lineage is either of type 1 or type 2. Slade (2000a) noticed that when both the descendant lineage and the incoming lineage have the favored type, the type of the continuing lineage does not matter so there is no need to introduce a new virtual lineage. Instead, these two possibilities can be collapsed into a single null event which does not change the numbers and types of ancestral lineages. That is, we can use

$$q_o(r_1, r_2, v_1 + 1, v_2) + q_o(r_1, r_2, v_1, v_2 + 1) = q_o(r_1, r_2, v_1, v_2) \tag{C.21}$$

in (C.6), (C.8), (C.17) and (C.19). As a result, no type-2 virtual lineages will be created.

Along the same lines, Fearnhead (2002) noticed that when mutation is parent-independent there is no need to follow ancestral lineages once they have mutated, because the ancestral lineage could be of either type. Any such lineage can be removed from the ancestral process. Here we use

$$q_o(r_1, r_2, v_1, v_2) + q_o(r_1 + 1, r_2 - 1, v_1, v_2) = q_o(r_1, r_2 - 1, v_1, v_2) \tag{C.22}$$

in (C.2), and other appropriate identities in (C.3) and (C.4). But we do not make use of this simplification in (C.1) because our specific goal is to model latent mutations in the ancestry of A_1. These are actual mutations, where the ancestral type was A_2. The remaining $A_1 \to A_1$ empty mutations are null events, which do not change the numbers and types of ancestral lineages.

The conditional ancestral processes for $\alpha < 0$ and $\alpha > 0$ given by (59) and (60) in the main text each include just five kinds of (non-null) events. We obtain these by applying the simplifications of Slade (2000a) and Fearnhead (2002) then grouping events by their outcomes. For example, the coalescent events in (C.10) have effect $r_2 \to r_2 - 1$, as do the combined mutations in (C.2) once the simplification of Fearnhead (2002) is applied. So these appear together as one kind of event, the fifth case in both (59) and (60).

We do not include null events in (59) and (60) since these by definition have no effect on the ancestral lineages. In the case $\alpha < 0$, the null events are empty mutations on type-1 real lineages and branching events on type-2 real lineages where the incoming line is also of type 2. These occur with total rate $r_1\theta\pi_1/2 + r_2|\alpha|/2$. In the case $\alpha > 0$, the null events are empty mutations on type-1 real lineages and branching events on type-1 real lineages where the incoming line is also of type 1. These occur with total rate $r_1\theta\pi_1/2 + r_1\alpha/2$.

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of Mathematical Functions. Dover, New York.

Arratia, R., Barbour, A.D., Tavaré, S., 1992. Poisson process approximations for the Ewens sampling formula. The Annals of Applied Probability 2, 519 – 535. doi:10.1214/aoap/1177005647.

Arratia, R., Barbour, A.D., Tavaré, S., 2003. Logarithmic Combinatorial Structures: A Probabilistic Approach. EMS monographs in mathematics, European Mathematical Society.

Baake, E., Bialowons, R., 2008. Ancestral processes with selection: branching and Moran models. Banach Center Publications 80, 33–52. doi:10.4064/bc80-0-2.

Barton, N., Etheridge, A., Véber, A., 2010. A new model for evolution in a spatial continuum. Electronic Journal of Probability 15, 162–216. doi:10.1214/EJP.v15-741.

Barton, N.H., Etheridge, A.M., 2004. The effect of selection on genealogies. Genetics 166, 1115–1131. doi:10.1093/genetics/166.2.1115.

Barton, N.H., Etheridge, A.M., Sturm, A.K., 2004. Coalescence in a random background. The Annals of Applied Probability 14, 754 – 785. doi:10.1214/105051604000000099.

Billingsley, P., 1999. Convergence of Probability Measures. John Wiley & Sons, New York. doi:10.1002/9780470316962.

Charlesworth, B., Hill, W.G., 2019. Selective effects of heterozygous protein-truncating variants. Nature Genetics 51, 2–2. doi:10.1038/s41588-018-0291-9.

Chen, J., Glémin, S., Lascoux, M., 2020. From drift to draft: How much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? Genetics 214, 1005–1018. doi:10.1534/genetics.119.302869.

Crane, H., 2016. The ubiquitous Ewens sampling formula. Statistical Science 31, 1–19. doi:10.1214/15-STS529.

Darden, T., Kaplan, N.L., Hudson, R.R., 1989. A numerical method for calculating moments of coalescent times in finite populations with selection. Journal of Mathematical Biology 27, 355–368. doi:10.1007/BF00275818.

Dembo, A., Zeitouni, O., 2009. Large Deviations Techniques and Applications. Springer Berlin, Heidelberg. doi:10.1007/978-3-642-03311-7.

Donnelly, P., 1986. Dual processes in population genetics, in: Tautu, P. (Ed.), Stochastic Spatial Processes, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 94–105. doi:10.1007/BFb0076240.

Doob, J.L., 1957. Conditional brownian motion and the boundary limits of harmonic functions. Bulletin de la Société mathématique de France 86, 431–458. URL: http://www.numdam.org/article/BSMF_1957__85__431_0.pdf.

Doob, J.L., 2001. Classical Potential Theory and Its Probabilistic Counterpart. Springer-Verlag Berlin, Heidelberg. doi:10.1007/978-3-642-56573-1.

Dufresne, D., 2001. The integrated square-root process. Centre for Actuarial Studies, Department of Economics, University of Melbourne, Victoria 3010, Australia. URL: https://fbe.unimelb.edu.au/__data/assets/pdf_file/0006/2592078/90.pdf.

Dukler, N., Mughal, M.R., Ramani, R., Huang, Y.F., Siepel, A., 2022. Extreme purifying selection against point mutations in the human genome. Nature Communications 13, 4312. doi:10.1038/s41467-022-31872-6.

Durrett, R., Fan, W.T.L., 2016. Genealogies in expanding populations. Ann. Appl. Probab. 26, 3456–3490. doi:10.1214/16-AAP1181.

Etheridge, A., 2011. Some Mathematical Models from Population Genetics: École d'Été de Probabilités de Saint-Flour XXXIX-2009. Springer Berlin, Heidelberg. doi:10.1007/978-3-642-16632-7.

Etheridge, A., Pfaffelhuber, P., Wakolbinger, A., 2006. An approximate sampling formula under genetic hitchhiking. The Annals of Applied Probability 16, 685–729. doi:10.1214/105051606000000114.

Etheridge, A.M., Kurtz, T.G., Letter, I., Ralph, P.L., Lung, T.T.H., 2023. Looking forwards and backwards: dynamics and genealogies of locally regulated populations. arXiv:2305.14488.

Ethier, S.N., Kurtz, T.G., 1993. Fleming-Viot processes in population genetics. SIAM Journal on Control and Optimization 31, 345–386. doi:10.1137/0331019.

Ethier, S.N., Kurtz, T.G., 2005. Markov Processes: Characterization and Convergence. John Wiley & Sons, New York. doi:10.1002/9780470316658.

Evans, S.N., Shvets, Y., Slatkin, M., 2007. Non-equilibrium theory of the allele frequency spectrum. Theoretical Population Biology 71, 109–119. doi:10.1016/j.tpb.2006.06.005.

Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theoretical Population Biology 3, 87–112. doi:10.1016/0040-5809(72)90035-4.

Ewens, W.J., 2004. Mathematical Population Genetics, Volume I: Theoretical Foundations. Springer-Verlag, Berlin.

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8, 610–618. doi:10.1038/nrg2146.

Fan, W.T.L., 2021. Stochastic PDEs on graphs as scaling limits of discrete interacting systems. Bernoulli 27, 1899–1941. doi:10.3150/20-bej1296.

Fearnhead, P., 2001. Perfect simulation from population genetic models with selection. Theoretical Population Biology 59, 263–279. doi:https://doi.org/10.1006/tpbi.2001.1514.

Fearnhead, P., 2002. The common ancestor at a nonneutral locus. Journal of Applied Probability 39, 38–54. doi:10.1239/jap/1019737986.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Clarendon, Oxford.

Hallatschek, O., Nelson, D.R., 2008. Gene surfing in expanding populations. Theoretical Population Biology 73, 158–170. doi:10.1016/j.tpb.2007.08.008.

Harpak, A., Bhaskar, A., Pritchard, J.K., 2016. Mutation rate variation is a primary determinant of the distribution of allele frequencies in humans. PLoS Genetics 12, e1006489. doi:10.1371/journal.pgen.1006489.

Hartl, D.L., Campbell, R.B., 1982. Allelic multiplicity in simple Mendelian disorders. American Journal of Human Genetics 34, 866–873. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1685710.

Heathcote, C.R., 1965. A branching process allowing immigration. Journal of the Royal Statistical Society. Series B (Methodological) 27, 138–143. URL: http://www.jstor.org/stable/2984491.

Herbots, H.M., 1997. The structured coalescent, in: Donnelly, P., Tavaré, S. (Eds.), Progress in Population Genetics and Human Evolution (IMA Volumes in Mathematics and its Applications, vol. 87. Springer-Verlag, New York, pp. 231–255.

Johnson, K.E., Adams, C.J., Voight, B.F., 2022. Identifying rare variants inconsistent with identity-by-descent in population-scale whole-genome sequencing data. Methods in Ecology and Evolution 13, 2429–2442. doi:10.1111/2041-210X.13991.

Joyce, P., 1995. Robustness of the Ewens sampling formula. Journal of Applied Probability 32, 609–622. doi:10.2307/3215116.

Joyce, P., Tavaré, S., 1995. The distribution of rare alleles. Journal of Mathematical Biology 33, 602–618. doi:10.1007/BF00298645.

Kaplan, N.L., Darden, T., Hudson, R.R., 1988. The coalescent process in models with selection. Genetics 120, 819–829. doi:10.1093/genetics/120.3.819.

Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer New York, NY. doi:10.1007/978-1-4612-0949-2.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Aguilar Salinas, C.A., Ahmad, T., Albert, C.M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E.J., Boehnke, M., Bonnycastle, L.L., Bottinger, E.P., Bowden, D.W., Bown, M.J., Chambers, J.C., Chan, J.C., Chasman, D., Cho, J., Chung, M.K., Cohen, B., Correa, A., Dabelea, D., Daly, M.J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P.T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S.J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M.M., Hultman, C.M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.J., Kim, Y.J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H.M., Kugathasan, S., Kwak, S.H., Laakso, M., Lehtimäki, T., Loos, R.J.F., Lubitz, S.A., Ma, R.C.W., MacArthur, D.G., Marrugat, J., Mattila, K.M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Meigs, J.B., Melander, O., Metspalu, A., Neale, B.M., Nilsson, P.M., O'Donovan, M.C., Ongur, D., Orozco, L., Owen, M.J., Palmer, C.N.A., Palotie, A., Park, K.S., Pato, C., Pulver, A.E., Rahman, N., Remes, A.M., Rioux, J.D., Ripatti, S., Roden, D.M., Saleheen, D., Salomaa, V., Samani, N.J., Scharf, J., Schunkert, H., Shoemaker, M.B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P.F., Suvisaari, J., Tai, E.S., Teo, Y.Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Vawter, M.P., Ware, J.S., Watkins, H., Weersma, R.K., Wessman, M., Wilson, J.G., Xavier, R.J., Neale, B.M., Daly, M.J., MacArthur, D.G., Genome Aggregation Database Consortium, 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. doi:10.1038/s41586-020-2308-7.

Karlin, S., McGregor, J., 1964. On some stochastic models in genetics, in: Gurland, J. (Ed.), Stochastic Models in Medicine and Biology. The University of Wisconsin Press, Madison, pp. 245–271.

Kemeny, J.G., Snell, J.L., 1960. Finite Markov Chains. D Van Nostrand Company, New York.

Kendall, D.G., 1949. Stochastic processes and population growth. Journal of the Royal Statistical Society. Series B (Methodological) 11, 230–282. URL: http://www.jstor.org/stable/2984078.

Kim, B.Y., Huber, C.D., Lohmueller, K.E., 2017. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics 206, 345–361. doi:10.1534/genetics.116.197145.

Kimura, M., Ohta, T., 1973. The age of a neutral mutant persisting in a finite population. Genetics 75, 199–212. doi:10.1093/genetics/75.1.199.

Krone, S.M., Neuhauser, C., 1997. Ancestral processes with selection. Theoretical Population Biology 51, 210–237. doi:10.1006/tpbi.1997.1299.

Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Séurel, L., Venkat, A., Andolfatto, P., Przeworski, M., 2012. Revisiting an old riddle: What determines genetic diversity levels within species? PLOS Biology 10, 1–9. doi:10.1371/journal.pbio.1001388.

Louvet, A., Véber, A., 2023. Measure-valued growth processes in continuous space and growth properties starting from an infinite interface. arXiv:2205.03937.

Mathieson, I., McVean, G., 2014. Demography and the age of rare variants. PLOS Genetics 10, 1–11. doi:10.1371/journal.pgen.1004528.

Millet, A., Nualart, D., Sanz, M., 1989. Integration by parts and time reversal for diffusion processes. The Annals of Probability 17, 208 – 238. doi:10.1214/aop/1176991505.

Moran, P.A.P., 1958. Random processes in genetics. Proc. Camb. Phil. Soc. 54, 60–71. doi:10.1017/S0305004100033193.

Moran, P.A.P., 1962. Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

Müller, C., Tribe, R., 1995. Stochastic p.d.e.'s arising from the long range contact and long range voter processes. Probability Theory and Related Fields 102, 519–545. doi:10.1007/BF01198848.

Nagylaki, T., 1990. Models and approximations for random genetic drift. Theoretical Population Biology 37, 192–212. doi:10.1016/0040-5809(90)90035-T.

Nait Saada, J., Kalantzis, G., Shyr, D., Cooper, F., Robinson, M., Gusev, A., Palamara, P.F., 2020. Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. Nature Communications 11, 6130. doi:10.1038/s41467-020-19588-x.

Nei, M., 1968. The frequency distribution of lethal chromosomes in finite populations. Proceedings of the National Academy of Sciences USA 60, 517–524. doi:10.1073/pnas.60.2.517.

Neuhauser, C., Krone, S.M., 1997. The genealogy of samples in models with selection. Genetics 145, 519–534. doi:10.1093/genetics/145.2.519.

Notohara, M., 1990. The coalescent and the genealogical process in geographically structured population. Journal of Mathematical Biology 29, 59–75. doi:10.1007/BF00173909.

O'Connor, T.D., Fu, W., NHLBI GO Exome Sequencing Project, ESP Population Genetics and Statistical Analysis Working Group, Turner, E., Mychaleckyj, J.C., Logsdon, B., Auer, P., Carlson, C.S., Leal, S.M., Smith, J.D., Rieder, M.J., Bamshad, M.J., Nickerson, D.A., Akey, J.M., 2015.

Rare variation facilitates inferences of fine-scale population structure in humans. Molecular Biology and Evolution 32, 653–660. doi:10.1093/molbev/msu326.

Platt, A., Pivirotto, A., Knoblauch, J., Hey, J., 2019. An estimator of first coalescent time reveals selection on young variants and large heterogeneity in rare allele ages among human populations. PLOS Genetics 15, 1–25. doi:10.1371/journal.pgen.1008340.

Reich, D.E., Lander, E.S., 2001. On the allelic spectrum of human disease. Trends in Genetics 17, 502–510. doi:10.1016/S0168-9525(01)02410-6.

Sawyer, S., 1983. A stability property of the Ewens sampling formula. Journal of Applied Probability 20, 449–459. doi:10.2307/3213883.

Schraiber, J.G., Evans, S.N., Slatkin, M., 2016. Bayesian inference of natural selection from allele frequency time series. Genetics 203, 493–511. doi:10.1534/genetics.116.187278.

Seplyarskiy, V., Koch, E.M., Lee, D.J., Lichtman, J.S., Luan, H.H., Sunyaev, S.R., 2023. A mutation rate model at the basepair resolution identifies the mutagenic effect of polymerase III transcription. Nature Genetics 55, 2235–2242. doi:10.1038/s41588-023-01562-0.

Seplyarskiy, V.B., Soldatov, R.A., Koch, E., McGinty, R.J., Goldmann, J.M., Hernandez, R.D., Barnes, K., Correa, A., Burchard, E.G., Ellinor, P.T., McGarvey, S.T., Mitchell, B.D., Vasan, R.S., Redline, S., Silverman, E., Weiss, S.T., Arnett, D.K., Blangero, J., Boerwinkle, E., He, J., Montgomery, C., Rao, D.C., Rotter, J.I., Taylor, K.D., Brody, J.A., Chen, Y.D.I., de las Fuentes, L., Hwu, C.M., Rich, S.S., Manichaikul, A.W., Mychaleckyj, J.C., Palmer, N.D., Smith, J.A., Kardia, S.L.R., Peyser, P.A., Bielak, L.F., O'Connor, T.D., Emery, L.S., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Population Genetics Working Group, Gilissen, C., Wong, W.S.W., Kharchenko, P.V., Sunyaev, S., 2021. Population sequencing data reveal a compendium of mutational processes in the human germ line. Science 373, 1030–1035. doi:10.1126/science.aba7408.

Shiga, T., 1988. Stepping stone models in population genetics and population dynamics, in: Albeverio, S., Blanchard, P., Hazewinkel, M., Streit, L. (Eds.), Stochastic Processes in Physics and Engineering. Springer Netherlands, Dordrecht, pp. 345–355. doi:10.1007/978-94-009-2893-0_18.

Slade, P.F., 2000a. Most recent common ancestor probability distributions in gene genealogies under selection. Theoretical Population Biology 58, 291–305. doi:10.1006/tpbi.2000.1488.

Slade, P.F., 2000b. Simulation of selected genealogies. Theoretical Population Biology 57, 35–49. doi:10.1006/tpbi.1999.1438.

Slater, L.J., 1960. Confluent Hypergeometric Functions. Cambridge University Press, Cambridge.

Slatkin, M., Rannala, B., 1997a. Estimating the age of alleles by use of intraallelic variability. The American Journal of Human Genetics 60, 447–458. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1712388/.

Slatkin, M., Rannala, B., 1997b. The sampling distribution of disease-associated alleles. Genetics 147, 1855–1861. doi:10.1093/genetics/147.4.1855.

Stephens, M., Donnelly, P., 2003. Ancestral inference in population genetics models with selection (with Discussion). Australian & New Zealand Journal of Statistics 45, 395–430. doi:10.1111/1467-842X.00295.

Stroock, D.W., Varadhan, S.R., 1979. Multidimensional Diffusion Processes. Springer Berlin, Heidelberg. doi:10.1007/3-540-28999-2.

Takahata, N., 1988. The coalescent in two partially isolated diffusion populations. Genetics Research 52, 213–222. doi:10.1017/S0016672300027683.

Tavaré, S., 2021. The magical Ewens sampling formula. Bulletin of the London Mathematical Society 53, 1563–1582. doi:10.1112/blms.12537.

Tricomi, F., Erdélyi, A., 1951. The asymptotic expansion of a ratio of gamma functions. Pacific Journal of Mathematics 1, 133–142. doi:10.2140/pjm.1951.1.133.

Vanyolos, A., Cho, M., Glasgow, S.A., 2014. Probability density of the CIR model. Available at SSRN 2508699 doi:10.2139/ssrn.250869.

Wakeley, J., 2008. Conditional gene genealogies under strong purifying selection. Molecular Biology and Evolution 25, 2615–2626. doi:10.1093/molbev/msn209.

Wakeley, J., Fan, W.T.L., Koch, E.M., Sunyaev, S.R., 2023. Recurrent mutation in the ancestry of a rare variant. Genetics doi:10.1093/genetics/iyad049. iyad049.

Watterson, G.A., 1974. The sampling theory of selectively neutral alleles. Advances in Applied Probability 6, 463–488. doi:10.2307/1426228.

Watterson, G.A., 1976. Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. Theoretical Population Biology 10, 239–253. doi:10.1016/0040-5809(76)90018-6.

Weghorn, D., Balick, D.J., Cassa, C., Kosmicki, J.A., Daly, M.J., Beier, D.R., Sunyaev, S.R., 2019. Applicability of the mutation-selection balance model to population genetics of heterozygous protein-truncating variants in humans. Molecular Biology and Evolution 36, 1701–1710. doi:10.1093/molbev/msz092.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159. URL: https://www.genetics.org/content/16/2/97.

Zaidi, A.A., Mathieson, I., 2020. Demographic history mediates the effect of stratification on polygenic scores. eLife 9, e61548. doi:10.7554/eLife.61548.