bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4/12/2016

The effects of population size histories on estimates of selection coefficients from time-series genetic data

Ethan M. Jewett[1,2], Matthias Steinrücken[3], Yun S. Song[1,2,4,5,6]

Departments of EECS[1], Statistics[2], and Integrative Biology[4], University of California, Berkeley, CA 94720, USA.
Department of Biostatistics and Epidemiology[3], University of Massachusetts, Amherst, MA 01003, USA.
Departments of Biology[5] and Mathematics[6], University of Pennsylvania, Philadelphia, PA 19104, USA.

Corresponding author: Yun S. Song (yss@berkeley.edu)

Keywords: selection, inference, time series, diffusion, Wright-Fisher

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Many approaches have been developed for inferring selection coefficients from time series data while accounting for genetic drift. However, the improvement in inference accuracy that can be attained by modeling drift is unknown. Here, by comparing maximum likelihood estimates of selection coefficients that account for the true population size history with estimates that ignore drift, we address the following questions: how much can modeling the population size history improve estimates of selection coefficients? How much can mis-inferred population sizes hurt inferences of selection coefficients? We conduct our analysis under the discrete Wright-Fisher model by deriving the exact probability of an allele frequency trajectory in a population of time-varying size and we replicate our results under the diffusion model by extending the exact probability of a frequency trajectory derived by Steinrücken *et al.* (2014) to the case of a piecewise constant population. For both the discrete Wright-Fisher and diffusion models, we find that ignoring drift leads to estimates of selection coefficients that are nearly as accurate as estimates that account for the true population history, even when population sizes are small and drift is high. In populations of time-varying size, estimates of selection coefficients that ignore drift are similar in accuracy to estimates that rely on crude, yet reasonable, estimates of the population history. These results are of interest because inference methods that ignore drift are widely used in evolutionary studies and can be many orders of magnitude faster than methods that account for population sizes.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

Methods for inferring the selection coefficient at a single genetic locus from time series data have been employed extensively in evolutionary studies of simple traits. Such methods track the frequency of an allele or Mendelian trait over multiple generations and infer the selection coefficient that best explains the observed frequency changes. Studies of selective pressures conducted using time series approaches have provided evidence for strong selective forces in natural populations and have helped to characterize the ways in which environmental factors influence evolution through selection (Clarke and Murray, 1962; Clark, 1979; Wall *et al.*, 1980; Lynch, 1987; Stine and Smith, 1990; Goudsmit *et al.*, 1996; Cowie and Jones, 1998; Harrigan *et al.*, 1998; Cook *et al.*, 1999; Haubruge and Arnaud, 2001; Bonhoeffer *et al.*, 2002; Reimchen and Nosil, 2002; Cook *et al.*, 2005; Labbé *et al.*, 2009).

Because random fluctuations in allele frequencies due to genetic drift are often small compared to changes due to selective pressures, it is common practice for studies to assume that allele frequencies change deterministically over time according to well-known deterministic formulas of Fisher (1922, p.424) and Haldane (1927, p.840) or related expressions (Gillespie, 2010; Hartl and Clark, 2007). However, because allele frequency trajectories can be heavily influenced by genetic drift when population sizes or selection coefficients are small, many methods have been developed to account for drift by explicitly modeling finite population sizes when inferring selection coefficients from observed allele frequency trajectories (Manly, 1985; O'Hara, 2005; Bollback *et al.*, 2008; Malaspinas *et al.*, 2012; Mathieson and McVean, 2013; Lacerda and Seoighe, 2014; Steinrücken *et al.*, 2014; Foll *et al.*, 2015; Ferrer-Admetlla *et al.*, 2015) and when testing hypotheses about selection versus drift (Fisher and Ford, 1947; Schaffer *et al.*, 1977; Wilson, 1980; Nishino, 2013; Feder *et al.*, 2014; Topa *et al.*, 2015).

Although estimates of selection coefficients are likely to be improved by accounting for population size histories, the expected amount of improvement is not well characterized. Even in relatively small populations, allele frequencies and other evolutionary processes behave almost deterministically if the selection coefficient or allele frequency is sufficiently high (Rouzine *et al.*, 2001), suggesting that methods that ignore drift might perform well under these conditions. Conversely, if drift is strong allele frequency trajectories can be noisy and the accuracy of methods that ignore drift may be comparable to that of methods that account for population size, as all methods are likely to perform poorly under these conditions (Gallet *et al.*, 2012).

If computationally fast methods that ignore drift are accurate, they could dramatically reduce the time required to infer selection coefficients in data sets with many loci. In addition to their computational efficiency, methods that ignore drift do not require estimates of effective population sizes, which can be difficult to obtain accurately. Moreover, ignoring drift can lead to simple formulas and inference procedures under complicated evolutionary scenarios (e.g., Illingworth *et al.*, 2012). Therefore, in light of the beneficial properties of methods that ignore drift and assume deterministic allele frequency trajectories, it is of interest to compare their accuracy to that of methods that account for population size histories.

The theoretical accuracy of methods for inferring selection coefficients can be difficult to derive analytically. Thus, to explore differences between methods that ignore or account for drift, one can take the approach of empirically comparing inferences made by the same estimator, either accounting for the true population size

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

history or ignoring the size history by assuming that populations are large and drift is negligible. This is the the approach we take here. For our analyses, we consider maximum likelihood estimators of selection coefficients because they are typically quite accurate and have desirable statistical properties. Moreover, the majority of recently-developed methods for inferring selection coefficients from time series data are maximum likelihood estimators, making them an important category of methods to evaluate.

To draw conclusions about the accuracy of maximum likelihood estimators, it is important to consider estimators based on exact likelihoods rather than approximations, so that differences in estimates can be attributed entirely to whether a method ignores or accounts for drift. Although several approximate approaches have been developed for computing the likelihood of a selection model given time series allele frequency data, only two existing methods compute probabilities that are exact under a widely accepted model. In particular, the methods of Bollback *et al.* (2008) and Steinrücken *et al.* (2014) compute exact probabilities under the diffusion approximation of the Wright-Fisher process. However, no method computes the exact probability of an allele frequency trajectory under the discrete Wright-Fisher model, as the matrix powers required for such a method are considered to be computationally inefficient. Moreover, no existing inference method based on the exact likelihood models time-varying population histories, making it difficult to explore the effects of accounting for demography on inference accuracy.

Here, we derive the exact probability of an allele frequency trajectory in a population of piecewise constant size under two classical models: the discrete Wright-Fisher model and the diffusion approximation of the Wright-Fisher process.

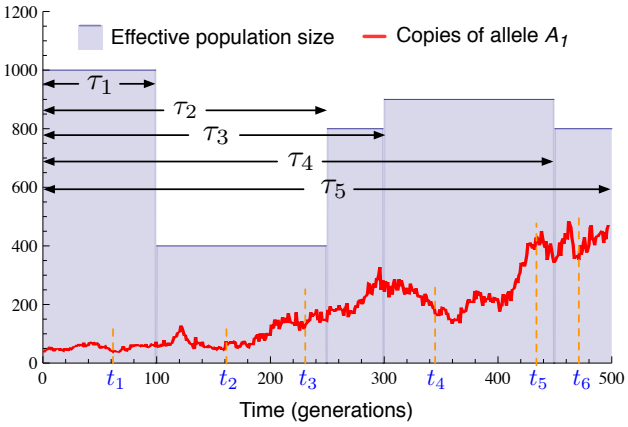

FIGURE 1. Diagram of the model. An allele at a single locus evolves in a population of piecewise constant size with $L = 5$ epochs spanning the time periods $[\tau_0, \tau_1], ..., [\tau_{L-1}, \tau_L]$, where $\tau_0 \equiv 0$. Samples of sizes $n_1, ..., n_k$ haplotypes are taken at times $t_1, ..., t_K$.

We then use maximum likelihood estimators obtained using these probabilities to explore how ignoring or accounting for the true population history affects estimates of selection coefficients.

RESULTS

To compare the performance of estimators that ignore or account for drift, we inferred selection coefficients from allele frequency trajectories simulated under a variety of population histories of time-varying size.

1.1. **The population model.** In all of our analyses, we considered a single biallelic locus with alleles labeled a and A evolving under selection and recurrent mutation in a panmictic population comprised of L different epochs $\ell = 1, ..., L$, each with constant size N_ℓ diploid individuals (Figure 1). Epoch ℓ corresponds to the time interval $[\tau_{\ell-1}, \tau_\ell]$, where time is measured continuously in units of generations and we define $\tau_0 \equiv 0$. By varying the population sizes N_ℓ across epochs, it

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is possible to model a variety of size-change patterns including exponential growth, bottlenecks, and rapidly oscillating population sizes.

Within epoch ℓ, all mutation and selection parameters are assumed to be constant. In particular, we assume that the per-generation probability that allele a mutates to allele A is $u_{aA}^{(\ell)}$ and the per-generation probability that allele A to a is $u_{Aa}^{(\ell)}$. The three possible genotypes, aa, aA, and AA, have relative fitnesses given by $w_{AA}^{(\ell)} = 1+s_\ell$, $w_{aA}^{(\ell)} = 1+h_\ell s_\ell$, and $w_{aa}^{(\ell)} = 1$ in epoch ℓ, where s_ℓ is the selection coefficient and h_ℓ is the dominance parameter.

We denote the collection of model parameters in epoch ℓ by Θ_ℓ and the set of parameters across all epochs by Θ. It will also be convenient to denote the value of the model parameters at time t by N_t, $u_{aA}^{(t)}$, $u_{Aa}^{(t)}$, s_t, and h_t, where t is measured continuously in units of generations. The epoch in which time t lies will be denoted by ℓ_t and the epoch in which sampling event k lies will be denoted by ℓ_k. It will be clear from the context whether the subscript on ℓ refers to a time or a sampling event.

We denote the population-wide number of copies of allele A in generation t by c_t and the population-wide frequency of allele A by y_t. In practice, we do not observe the true population count of allele A. Instead, the data consist of observed counts $o_1, ..., o_K$ of the number of times allele A is observed in K different samples of sizes $n_1, ..., n_K$ haplotypes, taken at times $t_1 < \cdots < t_K$. For simplicity, we assume that each sampling time t_k is an integer for $k = 1, ..., K$. The consecutive observed counts $(o_k, o_{k+1}, ..., o_{k'})$ will be denoted by $o_{[k:k']}$.

In general, we will denote random variables corresponding to observed quantities using capital letters (e.g., O_k, C_t, and Y_t). The goal is to compute the probability $\mathbb{P}_\Theta\{O_{[1:K]} = o_{[1:K]}\}$ of the observed data, conditional on the model parameters Θ.

1.2. Probabilities of frequency trajectories.

Several different evolutionary models can be used to describe stochastic allele frequency changes over time in a population. Discrete changes in allele frequency are often modeled using the Wright-Fisher and Moran processes, whereas continuous changes are often modeled using the diffusion approximation of the Wright-Fisher process (Karlin and Taylor, 1981; Ewens, 2004; Wakeley, 2008) or one of several approximations of the diffusion (e.g. Feder *et al.*, 2014; Lacerda and Seoighe, 2014).

Because it is unclear which model provides the most accurate description of biological evolutionary processes, we take the approach in this paper of deriving exact probabilities of allele frequency trajectories under two different evolutionary models: the discrete Wright-Fisher process and the continuous diffusion approximation.

Under the Wright-Fisher model, the probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$ of the observed allele counts can be obtained using the recursive formula developed in Section 3.1 (Procedure 1). Under the diffusion approximation, the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ can be obtained using the recursive formula developed in Section 3.2 (Procedure 2).

In Sections 3.4.1 and 3.4.2, we show that if drift is ignored and allele frequencies evolve deterministically, then the probabilities $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$ and $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ can be reduced to the simpler approximate probabilities $\mathbb{P}^{\infty}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$ and $\mathbb{P}^{\infty}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ which ignore the population history and which are computed using Procedures 3 and 4, respectively.

Different estimates of the model parameters Θ can be obtained using each of the different probabilities $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$, $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$, $\mathbb{P}^{\infty}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$, and $\mathbb{P}^{\infty}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ by finding the value of Θ that maximizes

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the given probability of the observed allele counts $o_{[1:K]}$. In our analyses we estimated the model parameters Θ separately using each of the different probabilities, yielding the estimators $\hat{s}_{\mathcal{W}}$, $\hat{s}_{\mathcal{D}}$, $\hat{s}^{\infty}_{\mathcal{W}}$, and $\hat{s}^{\infty}_{\mathcal{D}}$. The estimator $\hat{s}_{\mathcal{W}}$ accounts for drift under the discrete Wright-Fisher model, while drift in this model is ignored by the estimator $\hat{s}^{\infty}_{\mathcal{W}}$. Similarly, the estimator $\hat{s}_{\mathcal{D}}$ accounts for drift under the diffusion model, while drift in this model is ignored by the estimator $\hat{s}^{\infty}_{\mathcal{D}}$.

The degree to which accounting for drift can improve estimates of selection coefficients can be investigated by comparing $\hat{s}_{\mathcal{W}}$ to $\hat{s}^{\infty}_{\mathcal{W}}$ on trajectories simulated under the discrete Wright-Fisher model and by comparing $\hat{s}_{\mathcal{D}}$ to $\hat{s}^{\infty}_{\mathcal{D}}$ on trajectories simulated under the diffusion approximation.

1.3. **Overview of the experimental design.** We simulated allele frequency trajectories under a variety of selection strengths and piecewise constant population histories reflecting demographic patterns such as exponential growth, bottlenecks, rapid population size oscillations, and constant histories. We then compared the demography-aware estimates $\hat{s}_{\mathcal{W}}$ and $\hat{s}_{\mathcal{D}}$ with the estimates $\hat{s}^{\infty}_{\mathcal{W}}$ and $\hat{s}^{\infty}_{\mathcal{D}}$ that ignore drift to study the degree to which accounting for population size can improve the accuracy of inferences.

1.4. **Expected allele frequency trajectories.** Before comparing the accuracy of the different estimators, we first explored the degree to which trajectories that ignore drift differ from trajectories that account for the population size under different evolutionary scenarios. Figure 2 shows the expected frequency of allele A in a discrete Wright-Fisher population of constant size for several different initial allele frequencies, selection coefficients, and effective population sizes. Figure 2 illustrates that, for any starting frequency and selection coefficient, the mean allele frequency trajectory approaches the mean trajectory in a population without drift (e.g., in a population of infinite size), as the true population size increases. Moreover, if the initial frequency is sufficiently high, the expected trajectory is close to its deterministic limit even when the population size is small and drift is high.

The results presented in Figure 2 suggest that the allele frequency trajectory will differ substantially from the limiting trajectory without drift only when at least two of the three factors that influence stochasticity in the allele frequency trajectory (effective population size, selection coefficient, and initial allele frequency) are small. Moreover, for biological populations with sufficiently large effective sizes, the allele frequency trajectory is likely to match the deterministic trajectory, regardless of the selection coefficient and initial frequency.

From Figure 2 it can also be seen that an effective population size of several thousand individuals is often sufficiently large to guarantee deterministic behavior, even when the selection coefficient and initial allele frequency are small. Thus, selection coefficient inference methods that ignore drift are likely to be accurate for a broad range of population sizes and selection coefficients. As we will see, methods that ignore drift can be almost as accurate as methods that account for drift, even within the small-parameter-value regime.

1.5. **Inference accuracy, accounting for constant population size.** To explore how accounting for drift affects inference accuracy, we first considered the accuracy of inferring selection coefficients in a population of constant finite size. Figure 3 shows the maximum likelihood estimate (MLE) of the selection coefficient for three different effective population sizes ($N = 100, 500, 1000$), three selection coefficients ($s = 0.01, 0.05, 0.1$), and two initial allele frequencies

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Procedure 1 Computing $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$

1: Define the quantities $\mathbf{d}_0 = (\mathbb{P}\{C_0 = 0\}, \mathbb{P}\{C_0 = 1\}, ..., \mathbb{P}\{C_0 = 2N_{t_0}\})$ and $\boldsymbol{\gamma}(o_1)$, where $\boldsymbol{\gamma}(o_k) = (\gamma_0(o_k), \gamma_1(o_k), ..., \gamma_{2N_{t_k}}(o_k))$ with $\gamma_i(o_k) = \binom{n_k}{o_k}(i/2N_{t_k})^{o_k}(1 - i/2N_{t_k})^{n_k - o_k}$.

2: Initialize $\mathbf{v}_1 = \mathbf{d}_0 \left[\prod_{t=1}^{t_1} \mathbf{T}_{t-1,t}\right] \text{diag}\{\boldsymbol{\gamma}(o_1)\}$.

3: For $k = 2 : K$, compute

$$\mathbf{v}_k = \mathbf{v}_{k-1} \left[\prod_{t=t_{k-1}+1}^{t_k} \mathbf{T}_{t-1,t}\right] \text{diag}\{\boldsymbol{\gamma}(o_k)\}.$$

4: Compute $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\} = \sum_{i=0}^{2N_{t_K}} v_{K,i}$.

Procedure 2 Computing $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]}\}$

1: For an initial starting frequency x initialize

$$\mathbf{b}_0 = \mathbf{C}_{\ell_1}^{-1} \mathbf{B}_{\ell_1}(x),$$

where $\mathbf{B}_\ell(x)$ is the vector of eigenfunctions of the diffusion operator given in Equation (A.14) and $\mathbf{C}_\ell = \text{diag}\{\langle B_{\ell,i}, B_{\ell,i} \rangle\}_{i=0}^{\infty}$ is given in Equation (A.18).

2: For $k = 1 : K$, compute

$$\mathbf{a}_k = \begin{cases} \mathbf{b}_{k-1} \mathbf{E}_{\ell_k}(t_k - t_{k-1}) & \text{if } \ell_{k-1} = \ell_k, \\ \mathbf{b}_{k-1} \mathbf{F}(t_{k-1}, t_k; \boldsymbol{\zeta}) & \text{otherwise,} \end{cases}$$

and

$$\mathbf{b}_k = \mathbf{a}_k \mathbf{W}_{\ell_k} \mathbf{G}_{\ell_k}^{o_k} (1 - \mathbf{G}_{\ell_k})^{n_k - o_k} \mathbf{W}_{\ell_k}^{-1},$$

where the matrices $\mathbf{E}_\ell(t)$, $\mathbf{F}(t_{k-1}, t_k; \boldsymbol{\zeta})$, $\mathbf{W}_\ell$, and $\mathbf{G}_\ell$ are given by Equations (A.17), (B.10), (A.15) and (A.11), respectively and $\boldsymbol{\zeta}$ is the set of Chebyshev nodes in the interval $[0, 1]$. The matrix inverse $\mathbf{W}_\ell^{-1} = \mathbf{D}_\ell \mathbf{W}_\ell^T \mathbf{C}_\ell^{-1}$ is computed easily using the diagonal matrices $\mathbf{C}_\ell$ and $\mathbf{D}_\ell$ in Equations (A.18) and (A.19).

3: Compute

$$\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\} = \frac{c_{\ell_K,0}}{B_{\ell_K,0}(0)} b_{K,0}, \tag{1}$$

where $c_{\ell_K,0} = \langle B_{\ell,0}, B_{\ell,0} \rangle = [C_{\ell_K}]_{0,0}$ is the $(0,0)$ element of matrix $\mathbf{C}_\ell$ in Equation (A.18) and $B_{\ell_K,0}(0)$ is the 0th element of the vector $\mathbf{B}_{\ell_K}(0)$ in Equation (A.14).

$(y_0 = 0.01, 0.1)$ for $h = 1/2$. In each panel, the violin plots summarize the maximum likelihood estimates for 100 different simulation replicates in which an allele frequency trajectory was simulated for 500 generations with samples of size $n = 50$ taken at generations $t =$50, 100, 150, 200, 250, 300, 350, 400, 450, and 500.

For the discrete Wright-Fisher model, allele frequency trajectories were simulated by sampling the allele frequency in each generation from the vector of transition probabilities, conditional on the frequency in the previous generation. Under the diffusion model, trajectories were sampled using the approach of Jenkins and Spanò (2015, personal communication). Maximum likelihood estimates were obtained for the Wright-Fisher trajectories using a grid search over the likelihoods computed using Procedures 1 and 3, and maximum likelihood estimates for the diffusion trajectories were obtained using the same grid

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Procedure 3 Computing $\mathbb{P}^{\infty}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$

1: Starting with $y_0^{\infty} = y_0$, for $t = 0, ..., t_K - 1$,
   (1) Compute $\tilde{y}_t^{\infty} = u_{aA}^{(t)} + (1 - u_{Aa}^{(t)} - u_{aA}^{(t)})y_t^{\infty}$.
   (2) Compute

$$y_{t+1}^{\infty} = u_{aA}^{(t)} + \left[\frac{(\tilde{y}_t^{\infty})^2(1+s_t) + \tilde{y}_t^{\infty}(1-\tilde{y}_t^{\infty})(1+h_t s_t)}{\bar{w}_t}\right](1 - u_{Aa}^{(t)} - u_{aA}^{(t)}),$$

   where $\bar{w}_t = (\tilde{y}_t^{\infty})^2(1+s_t) + 2\tilde{y}_t^{\infty}(1-\tilde{y}_t^{\infty})(1+h_t s_t) + (1-\tilde{y}_t^{\infty})^2$.

2: Compute

$$\mathbb{P}^{\infty}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\} = \prod_{k=1}^{K}\binom{n_k}{o_k}(y_{t_k}^{\infty})^{o_k}(1 - y_{t_k}^{\infty})^{n_k - o_k}.$$

Procedure 4 Computing $\mathbb{P}^{\infty}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$

1: Fix a large integer n and set $\Delta t = 1/n$.
2: Starting with $y_0^{\infty} = y_0$, for $j = 0, ..., nt_K - 1$, compute

$$y_{(j+1)\Delta t}^{\infty} = \Big\{u_{aA}^{(j\triangle t)} - (u_{aA}^{(j\triangle t)} + u_{Aa}^{(j\triangle t)})y_{j\Delta t}^{\infty} + y_{j\Delta t}^{\infty}(1 - y_{j\Delta t}^{\infty})[(1 - 2y_{j\triangle t}^{\infty})h_{j\triangle t}s_{j\triangle t} + y_{j\Delta t}^{\infty}s_{j\triangle t}]\Big\}\Delta t.$$

3: Compute

$$\mathbb{P}^{\infty}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\} = \prod_{k=1}^{K}\binom{n_k}{o_k}(y_{t_k}^{\infty})^{o_k}(1 - y_{t_k}^{\infty})^{n_k - o_k}.$$

search approach over the likelihoods computed using Procedures 2 and 4.

By comparing the estimates computed accounting for drift with the estimates obtained ignoring drift, it can be seen that all methods have similar accuracies. All methods perform well when the population size, selection coefficient, and initial frequency are sufficiently large (e.g., Figure 3I for the case $y_0 = 0.01$ and Panels 3K through 3R for the case $y_0 = 0.1$), and all methods perform poorly, otherwise. Figure 3 suggests that the parameter range in which selection coefficients can be inferred accurately by maximum likelihood corresponds with the range in which the assumption $N \approx \infty$ yields accurate inferences. Put another way: the regime in which selective pressures are strong enough to measure accurately corresponds to the regime in which allele frequency change is quasi-deterministic. Thus, methods that ignore or account for drift are likely to produce estimates of similar accuracy.

1.6. **Inference accuracy in populations of piecewise constant size.** We next explored the degree to which accounting for more complicated population histories can improve maximum likelihood estimates, focusing on three scenarios, a population with a bottleneck, a population undergoing exponential growth, and a population undergoing rapid oscillations in size. Under each scenario, we simulated 100 allele frequency trajectories for an allele with selection coefficient $s = 0.05$, dominance parameter $h = 1/2$, and initial frequency $y_0 = 0.1$ under either the Wright-Fisher or diffusion models. The parameter values

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

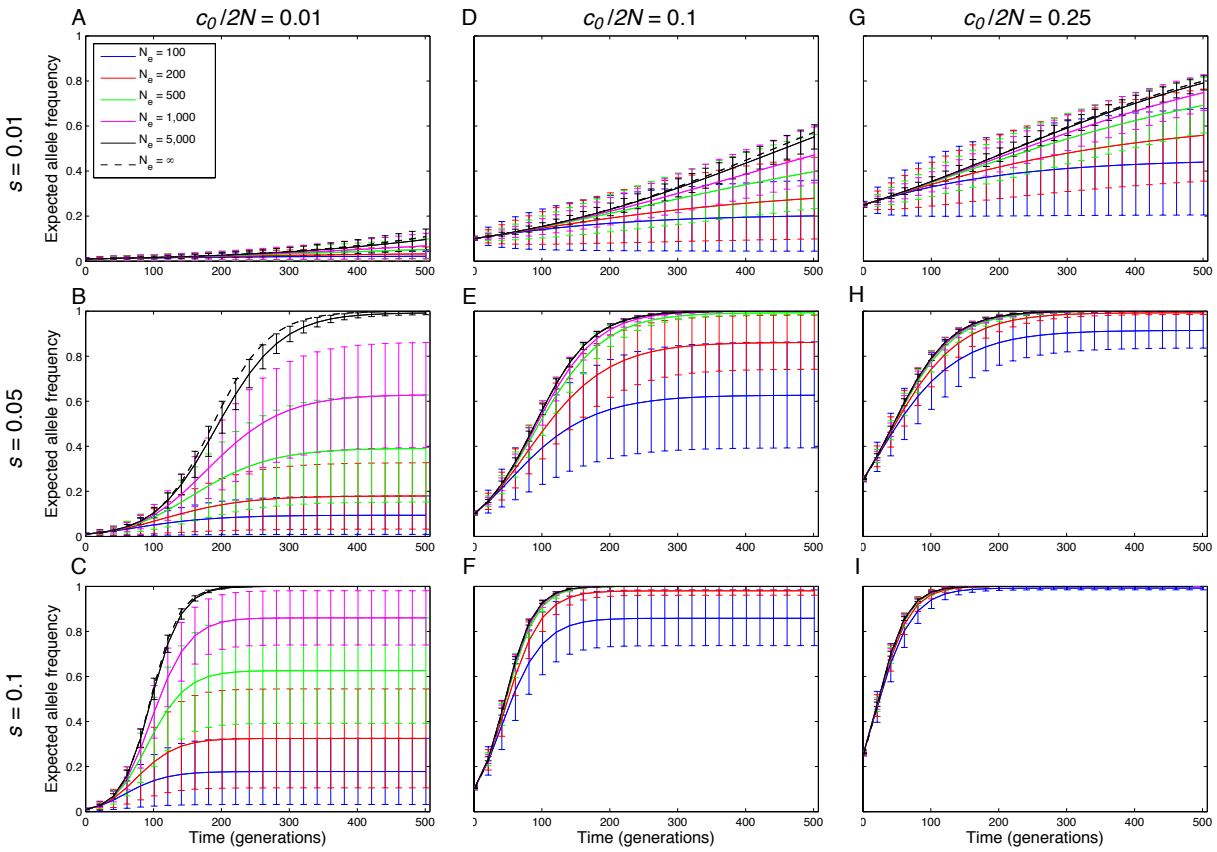

FIGURE 2. Expected Wright-Fisher trajectories of allele A for different initial starting counts c_0, selection coefficients s, and effective population sizes N. Columns correspond to different initial starting frequencies $c_0/2N$ with $c_0/2N = 0.01, 0.1$, and 0.25. In all panels, the error bars show the mean deviation on either side of the expected trajectory ($\mathbb{E}[\max\{0, c_t - \mathbb{E}[c_t]\}]/2N$ for the upper bar and $\mathbb{E}[\max\{0, \mathbb{E}[c_t] - c_t\}]/2N$ for the lower bar). The dominance parameter is set to $h = 1/2$ in all panels. Because the effects of mutation are negligible during the time periods we consider, we set $u_{Aa} = u_{aA} = 0$.

in these simulations were chosen so that drift would be strong enough to affect allele frequency trajectories, but not strong enough to prevent accurate inferences of selection coefficients.

To investigate the effect on accuracy of using crude, yet reasonable estimates of the population history, we also inferred selection coefficients using likelihoods computed using variants of Procedures 1 and 2 in which the population was assumed to consist of a single epoch of constant size equal to the Watterson estimate (Hein *et al.*, 2005, p.62) of the effective population size. The Watterson estimate was obtained by computing the expected site frequency spectrum (SFS) for the multi-epoch model for a sample size of 20 alleles, and then inferring the effective size of a single epoch using Watterson's estimate. The discrete Wright-Fisher and diffusion estimators based on the Watterson estimate of effective size are denoted by $\hat{s}_{\mathcal{W}}^{N_e}$ and $\hat{s}_{\mathcal{D}}^{N_e}$, respectively.

1.6.1. *The case of a bottleneck.* To model populations with bottlenecks, we considered populations composed of three epochs, each of length 100 generations, with sizes N_1, N_2, and N_3 satisfying $N_1 = N_3 = 5N_2$. Samples of size 50 were taken at times $50, 100, 150, 200, 250$, and 300. Figures 4A

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

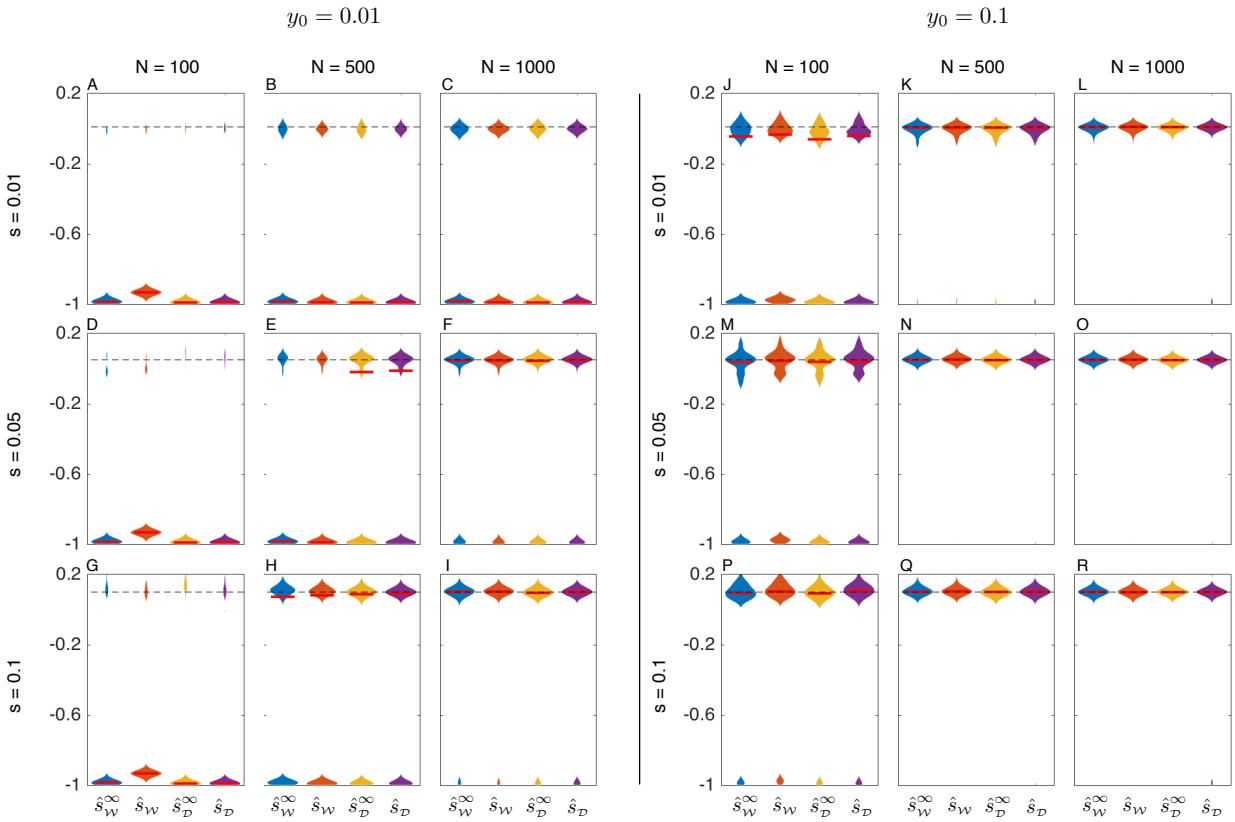

FIGURE 3. Maximum likelihood estimates of the selection coefficient s in populations of constant size. For each of three different selection coefficients ($s = 0.01, 0.05, 0.1$) and effective population sizes ($N = 100, 500, 1000$), 100 allele frequency trajectories were simulated for 500 generations under the either the Wright-Fisher or diffusion models. Samples of 50 alleles were taken at times 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500 generations. Bimodal violin plots are due to the fact that allele frequency trajectories typically fall into one of two categories: trajectories in which allele A is lost quickly, resulting in a strong negative estimate of the selection coefficient, and trajectories in which allele A remains segregating long enough to allow a more accurate estimate of the selection coefficient. Red bars indicate medians. The maximum width of each violin plot is scaled to the same value for all estimators.

and 4B show results for two different populations; in the population in Figure 4A, we set $N_1 = 500$ and in the population in Figure 4B we set $N_1 = 2500$.

From Figures 4A and 4B, it can be seen that all methods performed similarly. However, the methods that relied on the Watterson estimate of the effective population size were more biased than the other two methods when the effective population size was small, suggesting that methods that ignore drift entirely can produce more accurate estimates than methods that rely on rough estimates of the population history for the bottleneck model. Note that, despite the tight bottleneck in Figure 4A, inferences were still relatively accurate due to the larger sizes of epochs 1 and 3.

1.6.2. *The case of exponential growth.* To model exponential growth, we considered populations composed of five epochs, each of length 100 generations, with effective population sizes chosen to represent five-fold exponential growth across

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

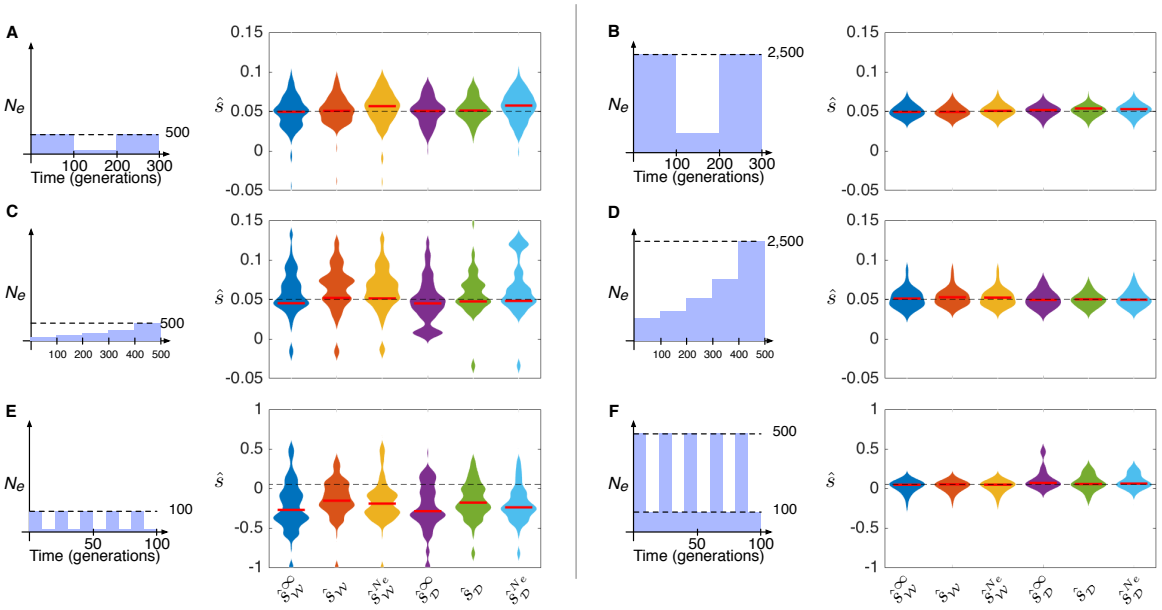

FIGURE 4. Maximum likelihood estimates of the selection coefficient s in populations with a bottleneck, exponential growth, or rapidly oscillating population size. In each panel, the trajectory of an allele with selection coefficient $s = 0.05$, dominance parameter $h = 1/2$, and starting frequency $y_0 = 0.1$ was simulated 100 times under the Wright-Fisher and diffusion models. Red bars indicate medians. The maximum width of each violin plot is scaled to the same value for all estimators.

all five epochs. Specifically, the size in epoch ℓ was set to $N_\ell = N_1 e^{-\eta\tau_{\ell-1}}$, where the growth constant η was chosen such that $e^{-\eta\tau_5} = 1/5$. Samples of size 50 were taken in generations 100, 200, 300, 400, and 500. From the results in Figures 4C and 4D, it can be seen that all methods performed with similar accuracy in the growth scenario.

1.6.3. *The case of rapidly oscillating population size.* Figures 4E and 4F show inferences of the selection coefficient in a population with rapidly oscillating size. Such demographic histories, which are often seen in insect populations like *Drosophila*, have moderate arithmetic mean sizes, but small harmonic mean sizes and experience episodes of extreme drift.

In the simulations shown in Figure 4E, the population size oscillates rapidly between 10 and 100 diploids every five generations. In the simulations shown in Figure 4F, the population size oscillates between 100 and 500 diploids every five generations. From Figure 4 it can be seen that the methods that ignore drift have similar accuracy to the methods that account for drift, although the methods that account for drift are slightly less biased when the population size oscillates between very small values (Figure 4E).

1.7. **Conditioning on segregation in the final sample.** It is sometimes of interest to infer the selection coefficient of an allele, conditional on the event that the allele is segregating in the most recent sample. Such conditional inferences are useful if alleles are ascertained in present-day samples and their historical trajectories are subsequently investigated.

Conditioning on segregation in the final sample is also useful for estimating weak positive selection coefficients when initial allele frequencies are low. This is because a large fraction of weakly selected alleles with low initial frequencies will

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

drift out of the population quickly resulting in large negative estimates of their selection coefficients. However, more accurate estimates can be obtained for the subset of alleles that are not lost quickly, which can be seen in Figures 3B, 3C, and 3E, in which the part of the density corresponding to alleles that are not lost quickly from the population is localized around the true selection coefficient.

Considering only alleles that are segregating in the final sample can lead to biased estimates of selection coefficients if likelihood methods do not properly condition on segregation. For example, weakly selected alleles typically drift out of small populations quickly. Thus, weakly selected alleles that escape loss and ultimately fix generally exhibit faster-than-expected increases in frequency that are similar to the unconditional trajectories of alleles under stronger selection. Thus, if a likelihood method does not properly account for conditioning, weakly selected alleles that are segregating in the final sample will have inflated inferred selection coefficients.

Estimators that ignore drift cannot be modified to condition on the event of segregation in the final sample because they implicitly assume that alleles follow fixed trajectories whose long-term behavior in the absence of mutation is entirely determined by the selection coefficient: fixation for positively selected alleles and loss for negatively selected alleles. Thus, estimators that ignore drift are expected to produce biased estimates of selection coefficients when applied to conditioned trajectories.

In contrast, the allele frequency trajectories in likelihood methods that account for the population size are modeled stochastically, allowing likelihoods to be modified to condition on segregation in the final sample. It is expected that methods that account for the true population size can be modified to produce accurate estimates of selection coefficients, whereas methods that ignore drift will necessarily produce biased estimates.

1.7.1. *Simulations conditioning on segregation.* To investigate the degree to which accounting for drift can improve estimates of selection coefficients when allele frequency trajectories are conditioned on segregation in the final sample, we modified the discrete Wright-Fisher probability in Section 1 to compute the likelihood conditional on segregation in the final sample using results derived in Section 3.3. Under a grid search, this modified likelihood yields the conditional maximum likelihood estimator $\hat{s}_{\mathcal{W}|S_K}$. We compared the estimates computed using the exact conditional estimator $\hat{s}_{\mathcal{W}|S_K}$ with estimates computed using the approximate estimator $\hat{s}^{\infty}_{\mathcal{W}}$ that ignores drift and cannot be modified to account for conditional allele frequency trajectories.

The effect of failing to account for conditioning is evident in the blue violin plots in Figure 5A-I, which correspond to the unconditional approximate maximum likelihood estimates $\hat{s}^{\infty}_{\mathcal{W}}$. As expected, when the true selection coefficient is small ($s \leq 0.01$), the estimates $\hat{s}^{\infty}_{\mathcal{W}}$ are biased upward. Conversely, when the selection coefficient is larger ($s \geq 0.05$), the approximate estimator $\hat{s}^{\infty}_{\mathcal{W}}$ produces negatively biased estimates because alleles under strong positive selection that remain segregating in the final sample show slower-than-expected increases in frequency. In contrast to the estimator $\hat{s}^{\infty}_{\mathcal{W}}$, the bias is negligible in the estimator $\hat{s}_{\mathcal{W}|S_K}$, which accounts for drift and properly conditions on segregation in the final sample (orange violin plots).

The results shown in Figure 5A-I suggest that methods that account for drift are capable of significantly improving the accuracy of estimates of

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

selection coefficients when allele frequency trajectories are conditioned on segregation. The differences in accuracy between methods that ignore or account for drift are visible for a range of selection coefficients and population sizes. However, the differences in accuracy between the methods diminish as the selection coefficient becomes weaker or the population size becomes larger.

1.7.2. *Simulations conditioning on segregation or fixation.* The magnitude of the bias in the estimates $\hat{s}^{\infty}_{\mathcal{W}}$ is due in part to the event on which trajectories are conditioned. In cases involving positive selection in populations of moderate or large size, most alleles will be fixed in the final sample (e.g., $> 80\%$ fixation within 10 generations when $s = 0.1$, $h = 1/2$, and $y_0 = 0.01$, and $N = 1000$). Thus, it may sometimes be more natural to condition on the event F_K that a selected allele is found (segregating or fixed) in the final sample. Under this conditioning scheme, the approximate estimator $\hat{s}^{\infty}_{\mathcal{W}}$ will not generally produce negatively biased estimates of selection coefficients because allele frequency trajectories will not be constrained to those which exhibit slower-than-expected increases in allele frequency.

In light of these considerations, we repeated the analysis shown in Figure 5A-I, simulating allele frequency trajectories conditional on the event that the allele was segregating or fixed in the final sample. To compare the estimates $\hat{s}^{\infty}_{\mathcal{W}}$ with maximum likelihood estimates that fully account for drift and the proper conditioning, we also modified the probability in $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$ computed in Procedure 1 to condition on the event F_K of segregation or fixation in the final sample, yielding the conditional probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|F_K\}$ (Equation 19) with the associated estimator $\hat{s}_{\mathcal{W}|F_K}$.

By comparing Figure 5J-R with Figure 5A-I, it can be seen that the estimator $\hat{s}^{\infty}_{\mathcal{W}}$ has considerably less bias when conditioning on the event F_K than when conditioning on S_K. Although the bias is still high when the population size is small ($N \approx 100$), it decreases quickly as the population size increases and becomes comparable to the bias in the properly conditioned, demography-aware estimator $\hat{s}_{\mathcal{W}|F_K}$ when the population size is greater than approximately $N = 500$ diploids. In contrast to Figure 5E-I, the bias in $\hat{s}^{\infty}_{\mathcal{W}}$ observed in Figure 5M-R is positive because the trajectories on which these estimates are based exclude those in which the allele is lost; thus, they exhibit faster-than-expected growth on average. The results in Figure 5J-R suggest that under certain conditioning schemes, methods that ignore drift can produce similar estimates to methods that account for drift.

1.8. **The effect of sample size on accuracy.** When the sample size is small, the variance in estimates arising from sampling noise will tend to obscure small differences between estimators that ignore or account for population size. Thus, when comparing methods, it is important to sample a sufficiently large number of alleles to ensure that the differences between the methods due to ignoring or accounting for drift are visible.

To evaluate the effects of sample size on inference accuracy, we inferred the selection coefficient for a range of sample sizes for several different combinations of the population size and selection coefficient. Figure 6 shows a plot of the variance in selection coefficients inferred using Procedures 1 and 3 for sample sizes ranging from $n = 2$ to $n = 50$. For each combination of N_e, s, and n, the trajectories of 100 alleles were simulated under the Wright-Fisher process with an initial allele frequency of $y_0 = 0.1$. Samples were taken in generations 50 and 100.

The plots in Figure 6 suggest that variability due to small sample sizes has a strong effect on the variability in estimates only for sample sizes

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

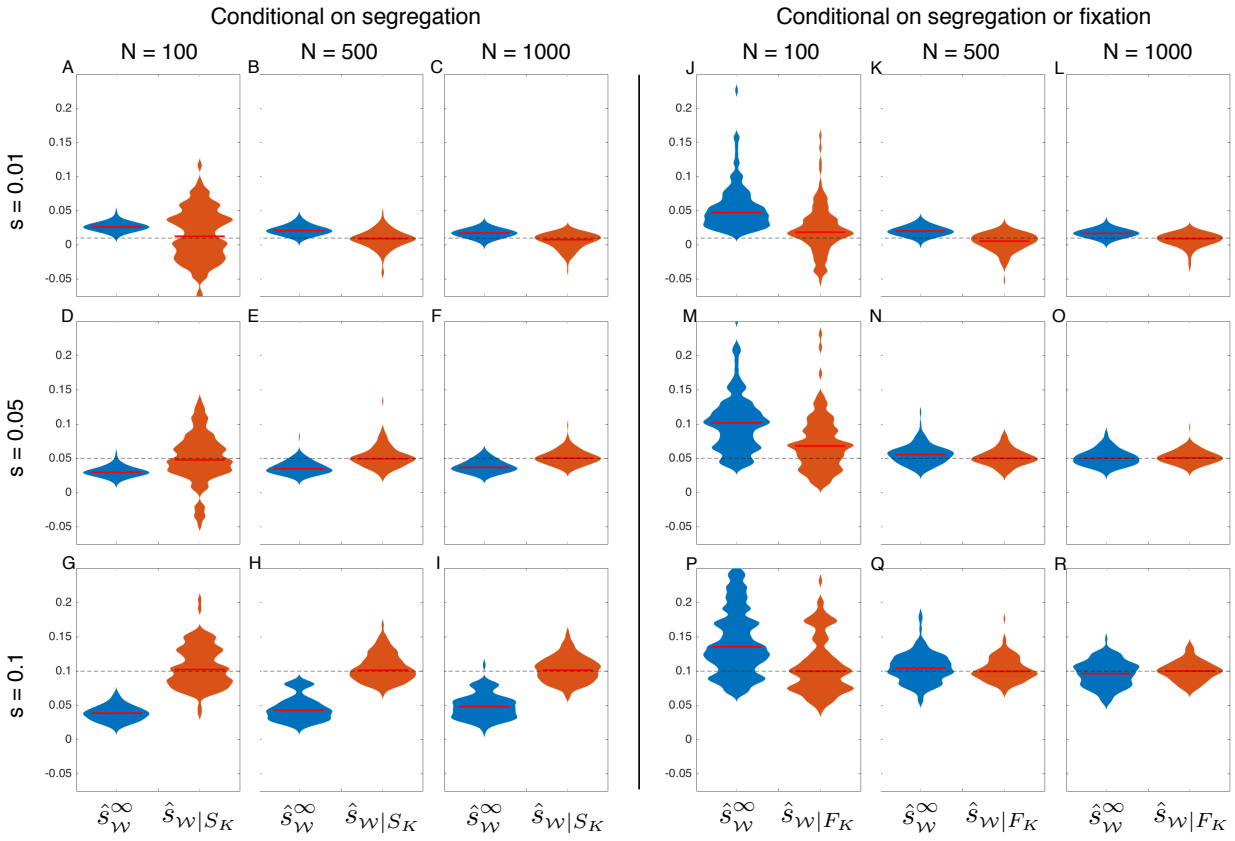

FIGURE 5. Estimates of selection coefficients, conditional on segregation. Each violin plot was computed using 100 frequency trajectories sampled over 500 generations for an allele with selection coefficient $s = 0.01$ and initial frequency $y_0 = 0.01$. As in Figure 3, samples of size $n = 50$ were taken in generations 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500. In Panels A-I, trajectories were sampled conditional on the event that the selected allele was segregating in the final sample. In Panels J-R, trajectories were sampled conditional on the event that the selected allele was either segregating or fixed in the final sample. Red bars indicate medians. The maximum width of each violin plot is scaled to the same value for both estimators.

smaller than 10 alleles. Thus, in all of our simulations we have used a sample size of $n = 50$ alleles so that differences between estimators are not likely to be obscured by the variance in estimates due to small sample sizes.

1.9. **Computational efficiency of methods.** As we have noted, methods that assume that allele frequency trajectories are deterministic can be considerably faster than methods that account for population size histories. Table 1 shows the average runtimes of the estimators $\hat{s}^{\infty}_{\mathcal{W}}$, $\hat{s}_{\mathcal{W}}$, $\hat{s}_{\mathcal{D}}$,

TABLE 1. Mean runtimes of the methods in Figure 4A-I (seconds).

N_e	s	$\hat{s}^{\infty}_{\mathcal{W}}$	$\hat{s}_{\mathcal{W}}$	$\hat{s}^{\infty}_{\mathcal{D}}$	$\hat{s}_{\mathcal{D}}$
100	0.01	0.01	2.30	4.74	197.25
	0.05	0.01	2.36	4.23	204.66
	0.1	0.01	2.29	3.98	217.34
500	0.01	0.02	134.07	4.41	185.18
	0.05	0.01	132.45	4.41	496.83
	0.1	0.02	126.35	4.46	531.15
1000	0.01	0.02	175.27	4.64	196.90
	0.05	0.02	191.53	4.78	815.27
	0.1	0.02	199.32	4.67	1950.59

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

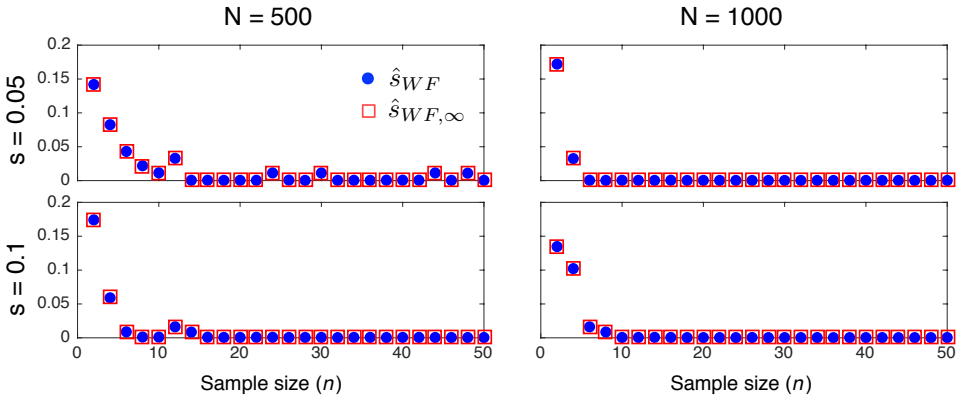

FIGURE 6. The effect of sample size on inference accuracy. The variance of the estimates produced by the methods in Procedures 1 and 3 are shown for a range of sample sizes.

and $\hat{s}_{\mathcal{D}}^{\infty}$ for the computations used to produce Figure 4A-I.

From the table, it can be seen that the runtimes are considerably faster for the estimators based on deterministic trajectories ($\hat{s}_{\mathcal{W}}^{\infty}$ and $\hat{s}_{\mathcal{D}}^{\infty}$). Moreover, the runtimes for $\hat{s}_{\mathcal{W}}^{\infty}$ and $\hat{s}_{\mathcal{D}}^{\infty}$ do not depend on the population size or selection coefficient. In comparison, the runtimes for the estimators $\hat{s}_{\mathcal{W}}$ and $\hat{s}_{\mathcal{D}}$ increase with increasing N_e and s because these methods depend on eigenvalue decompositions or sparse matrix products, which require larger matrices or greater precision when N_e or s is large.

2. Discussion

The results of our analyses suggest that accurate estimates of selection coefficients from allele frequency time series data can be obtained by assuming that alleles evolve without drift in a population of infinite size. In the majority of our simulations, the estimates obtained using deterministic approximations were nearly as accurate as estimates obtained by explicitly modeling the true population history and they were sometimes more accurate than estimates obtained using crude but reasonable estimates of the population history. Surprisingly, estimates made under the deterministic approximation were generally as accurate as estimates that accounted for drift, due to the fact that the exact maximum likelihood methods had low accuracy when drift was strong.

Accounting for the true population history only resulted in significantly improved estimates of selection coefficients when conditioning on the event that the target allele was segregating in the final sample. Methods that modeled the true population history were more accurate in this case because they could be modified to model conditional trajectories, whereas methods that assumed infinite population sizes could not. These results suggest that methods that account for drift are likely to be preferable under circumstances in which conditioning on segregation is desirable. However, it is important to note that deterministic methods can perform well when population sizes are moderately large if allele frequencies are conditioned on a slightly different event: the event that an allele is found (segregating or fixed) in the final sample.

The idea that ignoring drift can lead to accurate estimates of selection coefficients is not new. In fact, inference methods based on deterministic allele frequency trajectories capitalize on exactly this idea. However, our comparison with estimators based on exact likelihoods makes it possible

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to characterize the relative loss in accuracy that is incurred when drift is ignored, as well as the demographic, evolutionary, and sampling scenarios under which accounting for drift is likely to be important.

The comparatively accurate estimates achieved by methods that assume deterministic allele frequency trajectories are encouraging for three primary reasons. First, a large number of studies have relied on the assumption that alleles evolve deterministically in order to infer selection coefficients from biological time series data. Our results suggest that these estimates are likely to be nearly as accurate as those obtained using the exact likelihood accounting for drift. Second, estimators based on deterministic trajectories can be considerably faster than estimators that account for drift, making them useful for inferring selection coefficients at large numbers of loci. Third, it may be easier to obtain analytical results under the assumption that allele frequencies change deterministically, simplifying the development of inference methods for inferring selection coefficients under more complicated scenarios; for example, inferring coefficients at linked loci (Illingworth *et al.*, 2012). The ability to model factors such as linkage between alleles under selection may ultimately be more important than modeling drift, as these factors can have a strong effect on evolutionary dynamics (Burke, 2012; Long *et al.*, 2015). Finally, the ability to ignore the population size is useful in situations in which the true population history is unknown or difficult to infer.

In addition to characterizing the degree to which accounting for drift can improve estimates of selection coefficients, our results shed light on the accuracy of exact maximum likelihood methods for inferring selection coefficients from allele frequency trajectories. In accordance with previous work (Schaffer *et al.*, 1977; Gallet *et al.*, 2012), our findings suggest that very small selection coefficients ($s \leq 0.01$) are difficult to infer if the initial allele frequency and population size are not large. Moreover, even if the population size is large, the accurate inference of a small selection coefficient may require samples taken over hundreds of generations, during which time the selection coefficient could change considerably (Felsenstein, 1976; Siepielski *et al.*, 2009).

Despite the difficulties of inferring weak selection coefficients when the population size is small, coefficients of one percent or lower can be inferred accurately if the initial allele frequency is sufficiently high. It is important to note that the selection coefficient need not be high at the time of the very first sampling event, as long as the allele has reached a sufficiently high frequency at one of the intermediate sampling events, leading to quasi-deterministic behavior between some sampling time points that can be exploited by the maximum likelihood estimator. Thus, one need not restrict analyses to cases of selection on standing variation to obtain accurate inferences.

Although we have only considered positively selected alleles in our analyses, our results apply equally well to negatively selected alleles, as it is arbitrary whether we choose to track the trajectory of the allele with higher or lower fitness. We have also considered only low initial allele frequencies ($y_0 \leq 0.1$) for selected alleles; however, it is clear from Figure 2 that allele frequency trajectories become increasingly deterministic as the initial allele frequency increases. Thus, the accuracy of a method that assumes a deterministic trajectory will become more similar to that of a method that accounts for drift as the initial allele frequency increases. Conversely, for negatively selected alleles, the accuracy of the deterministic method will approach that of the exact likelihood as the initial allele frequency decreases. Thus,

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

our analyses provide a characterization of inference accuracy for both positively and negatively selected alleles for the full range of starting frequencies.

At first glance, our finding that the population size does not strongly influence estimates of selection coefficients might appear to be at odds with the fact that population size histories can be inferred from allele frequency time series data (O'Hara, 2005; Bollback *et al.*, 2008; Ferrer-Admetlla *et al.*, 2015). However, this is not the case. Methods for inferring the population size capitalize on information in the short term fluctuations of the allele frequency around its expected value, arising from drift; conversely, estimators of selection coefficients capitalize on the long-term changes in allele frequency due to selection, effectively averaging over the short-term fluctuations due to drift. Our results suggest that allele frequencies often change quasi-deterministically, even in small populations. Thus, deviations around the expected trajectory can be distinguished from long-term changes, allowing effective population sizes to be inferred accurately even in small populations.

We have conducted our analyses under two different models of evolution: the discrete Wright-Fisher model and the continuous diffusion model. Although the diffusion model was developed as an approximation to the Wright-Fisher process, it also captures the limiting behavior of a large class of evolutionary models, including the Wright-Fisher process, as the population size grows to infinity and mutation and selection parameters are scaled accordingly. Thus, it is reasonable to believe that our findings will generalize to maximum likelihood estimators derived under a wide range of evolutionary models.

Taken together, our results help to characterize the properties of maximum likelihood methods for inferring selection coefficients from time series data. Because of the accuracy and beneficial properties of maximum likelihood methods, it is reasonable to believe that our results provide insight into the accuracy with which it is possible to infer selection coefficients from biological data, and the degree to which accounting for the true population history can improve these estimates. Our results also provide justification for the use of fast inference methods based on the assumption that allele frequencies evolve deterministically. Such methods can be applied to infer selection coefficients efficiently on large genomic data sets with many loci. Finally, our results provide further justification for the use of deterministic approximations in the development of statistical approaches for studying time series data.

3. Methods

In this section, we compute the exact probability of an allele frequency trajectory in a population of piecewise-constant size under the discrete Wright-Fisher model and under the diffusion approximation. We also describe how drift can be ignored in these probabilities, yielding approximate estimators of selection coefficients that are similar to commonly-used approaches that assume deterministic allele frequency trajectories.

3.1. Computing $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$ under the discrete Wright-Fisher model.

To compute the probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$ under the discrete Wright-Fisher model, we make use of a hidden Markov model (HMM) similar to that presented in Steinrücken *et al.* (2014). However, the hidden state in our discrete model is the count c_t of the number of (unobserved) copies of allele A in the population at time t, rather than the continuous allele frequency y_t.

In our model, the count c_t of allele A evolves according to a Wright-Fisher process in which

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mutation occurs, followed by random mating, selection, and drift. Given that the count of allele A in generation t is $c_t = i$, let $f^t_{A|i}$ be the frequency of allele A in the gamete pool after mutation. Then

$$f^t_{A|i} = \left(\frac{i}{2N_t}\right)(1-\mu^{(t)}_{Aa}) + \left(1-\frac{i}{2N_t}\right)\mu^{(t)}_{aA} = \mu^{(t)}_{aA} + (1-\mu^{(t)}_{Aa}-\mu^{(t)}_{aA})\left(\frac{i}{2N_t}\right). \quad (2)$$

After random mating, the fraction of zygotes with each of the genotypes AA, Aa, and aa is $(f^t_{A|i})^2$, $2f^t_{A|i}(1-f^t_{A|i})$, and $(1-f^t_{A|i})^2$, from which it follows that the fraction of genotypes of each kind remaining after selection is given by

$$p^t_{AA|i} = \frac{(f^t_{A|i})^2(1+s_t)}{\bar{w}_{t,i}},$$
$$p^t_{Aa|i} = \frac{2f^t_{A|i}(1-f^t_{A|i})(1+h_t s_t)}{\bar{w}_{t,i}},$$
$$p^t_{aa|i} = \frac{(1-f^t_{A|i})^2}{\bar{w}_{t,i}}, \quad (3)$$

where $\bar{w}_{t,i} = (f^t_{A|i})^2(1+s_t) + 2f^t_{A|i}(1-f^t_{A|i})(1+h_t s_t) + (1-f^t_{A|i})^2$ is the mean fitness of the population.

Immediately after selection and before drift occurs, the probability that a randomly chosen allele is of type A is given by $p^t_{A|i} = p^t_{AA|i} + \frac{1}{2}p^t_{Aa|i}$. Then, as the result of drift, the count of allele A in generation $t+1$ is binomially distributed with mean $p^t_{A|i}$. Thus, the probability that allele A has count j in generation $t+1$, given that it had count i in generation t is

$$\mathbb{P}_{\Theta,\mathcal{W}}\{C_{t+1} = j | C_t = i\} = \binom{2N_{t+1}}{j}(p^t_{A|i})^j(1-p^t_{A|i})^{2N_{t+1}-j}. \quad (4)$$

The Wright-Fisher transition matrix $\mathbf{T}_{t,t+1}$ from generation t to generation $t+1$ is the $(2N_t+1)\times(2N_{t+1}+1)$ matrix with entry i,j given by

$$[\mathbf{T}_{t,t+1}]_{ij} = \mathbb{P}_{\Theta,\mathcal{W}}\{C_{t+1} = j | C_t = i\}, \quad (5)$$

which can be used to obtain the allele frequency distribution at each discrete generation t given the initial distribution at some time $s < t$. In particular, define $\mathbf{d}_t = (\mathbb{P}\{c_t = 0\}, \mathbb{P}\{c_t = 1\}, ..., \mathbb{P}\{c_t = 2N_t\})$, to be the distribution of the count of allele A in generation t. Using Equation (5), $\mathbf{d}_t$ can be computed recursively as $\mathbf{d}_t = \mathbf{d}_s\left[\prod_{g=s+1}^{t}\mathbf{T}_{g-1,g}\right]$ for $s < t$.

3.1.1. *Computing the probability* $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}\}$. The probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:k]}\}$ of the observed data is computed using the forward procedure for hidden Markov models. In particular, we define the vector $\mathbf{v}_k$ whose ith entry $v_{k,i}$ is the joint probability of the population-wide count of allele A at the kth sampling event and the observed sample allele counts up to sample k:

$$v_{k,i} = \mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:k]}, C_{t_k} = i\}. \quad (6)$$

To simplify calculations, we also define the conditional "emission probability"

$$\gamma_i(o_k) = \mathbb{P}_{\Theta}\{O_k = o_k | C_{t_k} = i\} = \binom{n_k}{o_k}(i/2N_{t_k})^{o_k}(1-i/2N_{t_k})^{n_k-o_k} \quad (7)$$

of the observed allele count, conditional on the population allele count. The probability in Equation (7) comes from the fact that the observed allele count at time t_k is a binomial random variable with sample size n_k and probability c_{t_k}. We then construct the emission probability vector

$$\boldsymbol{\gamma}(o_k) = (\gamma_0(o_k), \gamma_1(o_k), ..., \gamma_{2N_{t_k}}(o_k)). \quad (8)$$

The probability of the data is then given by the forward procedure (Rabiner, 1989), outlined in Procedure 1. In Procedure 1, the formula for $\mathbf{v}_1$ comes from the fact that

$$\mathbf{v}_1 = (\mathbb{P}_{\Theta,\mathcal{W}}\{O_1 = o_1, C_{t_1} = 0\}, ..., \mathbb{P}_{\Theta,\mathcal{W}}\{O_1 = o_1, C_{t_1} = 2N_{\ell_{t_1}}\}) = (\gamma_0(o_1)\mathbb{P}_{\Theta,\mathcal{W}}\{C_{t_1} = 0\}, ...,$$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\begin{aligned}
&\gamma_{2N_{\ell_{t_1}}}(o_1)\mathbb{P}_{\Theta,\mathcal{W}}\{C_{t_1} = 2N_{\ell_{t_1}}\}) \\
&= \mathbf{d}_{t_1}\mathrm{diag}\{\boldsymbol{\gamma}(o_1)\} \\
&= \mathbf{d}_0\left[\prod_{t=1}^{t_1}\mathbf{T}_{t-1,t}\right]\mathrm{diag}\{\boldsymbol{\gamma}(o_1)\}, \qquad (9)
\end{aligned}$$

where $\mathrm{diag}(\boldsymbol{\gamma})$ denotes the square matrix whose diagonal entries are given by $\boldsymbol{\gamma}$.

It has been noted by several authors that computing powers of the transition matrix is computationally prohibitive, providing one motivating factor for the use of approximations of the Wright-Fisher process, such as the diffusion and Gaussian approximations (Ewens, 1963; Feder *et al.*, 2014; Lacerda and Seoighe, 2014). However, the products $\prod_{t=t_{k-1}+1}^{t_k}\mathbf{T}_{t-1,t}$ in Procedure 1 do not require products of the transition matrix $\mathbf{T}_{t-1,t}$ because it suffices to repeatedly compute vector-matrix products instead of multiplying full matrices together. In practice, this can be done very quickly, even for large population sizes. A similar fast procedure was carried out by Zhao *et al.* (2014) to simulate trajectories under the Wright-Fisher model.

3.2. **Computing $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ under the diffusion approximation.** The diffusion approximation models the evolution of the continuous population frequency Y_t of allele A, rather than its count C_t. The time-evolution of the random frequency Y_t is governed by the diffusion transition density $p_\Theta(s,t;x,y)$ given by

$$\begin{aligned}
&p_\Theta(s,t;x,y)dy = \\
&\qquad \mathbb{P}_{\Theta,\mathcal{D}}\{y \le Y_t < y + dy | Y_s = x\}, \qquad (10)
\end{aligned}$$

for an infinitesimal increment dy. The quantity $p_\Theta(s,t;x,y)$ specifies the density of the allele frequency at time t, conditional on the value of the allele frequency at an earlier time s. For more details about the transition density function of the diffusion approximation, see Appendix A.

Using the diffusion transition density $p_\Theta(s,t;x,y)$ Steinrücken *et al.* (2014) developed an HMM to compute the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ of the data in a single epoch of constant size by efficiently integrating over the hidden allele frequencies $\{y_{t_1}, ..., y_{t_K}\}$ at the set of sampling times. Here, we extend this HMM to the case of piecewise-constant population size.

To compute the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ efficiently, Steinrücken *et al.* (2014) define the quantities $f_k(y)$ and $g_k(y)$ satisfying

$$\begin{aligned}
&f_k(y)dy := \\
&\qquad \mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:k]} = o_{[1:k]}, y \le Y_{t_k} < y + dy\}, \qquad (11)
\end{aligned}$$

and

$$\begin{aligned}
&g_k(y)dy := \\
&\qquad \mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:k-1]} = o_{[1:k-1]}, y \le Y_{t_k} < y + dy\} \qquad (12)
\end{aligned}$$

for an infinitesimal increment dy. The quantity $f_k(y)$ is the joint density of the allele frequency at time t_k and the observed counts up to sampling event k. The quantity $g_k(y)$ is the joint density of the allele frequency at time t_k and the observed counts up to sampling event $k-1$.

It follows from the definition of $f_k(y)$ that the probability of the data is given by

$$\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\} = \int_{y=0}^{1} f_K(y)dy. \qquad (13)$$

The quantity $f_K(y)$ can be obtained efficiently by recursion using the relationships

$$f_k(y) = g_k(y)\binom{n_k}{o_k}y^{o_k}(1-y)^{n_k-o_k}, \qquad (14)$$

and

$$g_k(y) = \int_{z=0}^{1} f_{k-1}(z)p_\Theta(t_{k-1},t_k;z,y)dz. \qquad (15)$$

Equation (14) follows from the fact that the observed number of copies of allele A at sampling event k is binomially distributed with count n_k and probability y_{t_k} and Equation (15) follows from the law of total probability integrating over $Y_{t_{k-1}}$.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Let $B_{\ell,i}(y)$ $(i = 0, 1, ...)$ be the ith eigenfunction of the backward diffusion operator $\mathcal{L}_\ell$ and let $\pi_\ell(y)$ be the speed density of $\mathcal{L}_\ell$ (Appendix A). Steinrücken *et al.* (2014) demonstrated that the recursive formulas in Equations (14) and (15) can be evaluated efficiently by expressing $f_k(y)$ and $g_k(y)$ as series of the form

$$f_k(y) = \sum_{i=0}^{\infty} b_{k,i}\pi_{\ell_k}(y)B_{\ell_k,i}(y) = \mathbf{b}_k\pi_{\ell_k}(y)\mathbf{B}_{\ell_k}(y) \tag{16}$$

and

$$g_k(y) = \sum_{i=0}^{\infty} a_{k,i}\pi_{\ell_k}(y)B_{\ell_k,i}(y) = \mathbf{a}_k\pi_{\ell_k}(y)\mathbf{B}_{\ell_k}(y), \tag{17}$$

where $\mathbf{B}_\ell(y) = (B_{\ell,0}(y),\ B_{\ell,1}(y),\ ...)$ and where $\mathbf{b}_k = (b_{k,0}, b_{k,1}, ...)$ and $\mathbf{a}_k = (a_{k,0}, a_{k,1}, ...)$ are vectors of constants that encode the densities $f_k(y)$ and $g_k(y)$ at the beginning of the epoch. In Appendix B, we extend the results of Steinrücken *et al.* (2014) to derive recursive formulas for the coefficients $\mathbf{a}_k$ and $\mathbf{b}_k$, resulting in Procedure 2, which computes the probability of an allele frequency trajectory under the diffusion approximation in a population of piecewise constant size.

3.3. **Conditional probabilities.** Sometimes it is desirable to compute the probability of the observed allele counts conditional on the event S_K that allele A is segregating in the final sample. In this section, we provide formulas for these conditional probabilities under the Wright-Fisher and diffusion models.

3.3.1. *Computing* $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|S_K\}$. In Section 1.7.1, we consider the probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|S_K\}$ of the data conditional on the event S_K that allele A is segregating in the final sample. In Appendix C, we show that in the case of the discrete Wright-Fisher model,

$$\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|S_K\} = \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\}} \sum_{i=0}^{2N_{t_K}} v_{K,i}, \tag{18}$$

where $v_{k,i}$ is defined in Equation (6) and $\mathbb{P}\{S_K|O_K = o_K\} = 1$ if $1 \le o_K < n_K$, or 0 otherwise. The probability $\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\}$ is given in Equation (C.3). Thus, if we wish to compute conditional probabilities under the Wright-Fisher model, we carry out Procedure 1, replacing step 3 with Equation (18).

3.3.2. *Computing* $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|F_K\}$. Similarly, for the event F_K that allele A is segregating or fixed in the final sample, we show in Appendix C that

$$\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|F_K\} = \frac{\mathbb{P}\{F_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{W}}\{F_K\}} \sum_{i=0}^{2N_{t_K}} v_{K,i}, \tag{19}$$

where $v_{k,i}$ is defined in Equation (6) and $\mathbb{P}\{F_K|O_K = o_K\} = 1$ if $1 \le o_K \le n_K$, or 0 otherwise. The probability $\mathbb{P}_{\Theta,\mathcal{W}}\{F_K\}$ is given in Equation (C.6). If we wish to compute conditional probabilities under the Wright-Fisher model, we carry out Procedure 1, replacing step 3 with Equation (19).

3.3.3. *Computing* $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}|S_K\}$. In the case of the diffusion approximation, we show in Appendix D that the conditional probability of the data given S_K can be computed as

$$\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}|S_K\} = \frac{\mathbb{P}\{S_K|O_K = o_K\}c_{\ell_K,0}b_{K,0}(t_K)}{B_{\ell_K,0}(0) - c_{\ell_K,0}\tilde{b}_{K,0}(0) - c_{\ell_K,0}\tilde{b}_{K,0}(n_K)}, \tag{20}$$

where $\mathbb{P}\{S_K|O_K = o_K\} = 1$ if $1 \le o_K < n_K$ or 0 otherwise, and

$\tilde{\mathbf{b}}_K(j) =$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\begin{cases} \mathbf{b}_0\mathbf{E}_{\ell_1}(t_K)\mathbf{W}_{\ell_K}\mathbf{G}_{\ell_K}^{j}(1-\mathbf{G}_{\ell_K})^{n_K-j}\mathbf{W}_{\ell_K}^{-1}, \\ \text{if } \ell_{t_K}=1, \\ \mathbf{b}_0\mathbf{F}(0,t_K;\boldsymbol{\zeta})\mathbf{W}_{\ell_K}\mathbf{G}_{\ell_K}^{j}(1-\mathbf{G}_{\ell_K})^{n_K-j}\mathbf{W}_{\ell_K}^{-1}, \\ \text{otherwise.} \end{cases} \tag{21}$$

Thus, if we are interested in conditional probabilities under the diffusion model, we carry out Procedure 2, replacing step 3 with Equation (20).

3.4. **The probability in the absence of genetic drift.** If we ignore genetic drift, the allele frequency changes deterministically over time. Here, we obtain versions of Procedures 1 and 2 in the case when the changes in allele frequency arising from genetic drift are negligible relative to the changes due to selection and recurrent mutation.

3.4.1. *Deterministic allele frequency trajectories under the Wright-Fisher model.* If there is no contribution to the change in allele frequency arising from genetic drift, the allele frequency in a given generation is equal to its expectation after mutation, random mating, and selection, conditional on its value in the previous generation. Because the expectation is not necessarily integer-valued, we no longer consider discrete integer allele counts c_t. Instead, we track the expected allele frequency, which we denote by $y_t^\infty \equiv \mathbb{E}_\infty[Y_t]$, where the subscript ∞ denotes the expectation when drift is negligible.

The expected frequency y_t^∞ is obtained by combining Equations (2) and (3), ignoring the drift step in Equation (4), yielding

$$y_{t+1}^\infty = \left[\frac{(\tilde{y}_t^\infty)^2(1+s_t)+\tilde{y}_t^\infty(1-\tilde{y}_t^\infty)(1+h_ts_t)}{\bar{w}_t}\right], \tag{22}$$

where

$$\tilde{y}_t^\infty = u_{aA}^{(t)} + (1-u_{Aa}^{(t)}-u_{aA}^{(t)})y_t^\infty \tag{23}$$

and $\bar{w}_t = (\tilde{y}_t^\infty)^2(1+s_t)+2\tilde{y}_t^\infty(1-\tilde{y}_t^\infty)(1+h_ts_t)+(1-\tilde{y}_t^\infty)^2$. Equations (22) and (23) are iterated to find the allele frequency in any generation $t>0$.

3.4.2. *Deterministic allele frequency trajectories under the diffusion model.* Under the diffusion model in an Epoch ℓ of constant size, the allele frequency Y_t obeys the stochastic differential equation (SDE)

$$dY_t = \mathcal{M}_\ell(Y_t)dt + \sqrt{Y_t(1-Y_t)}dB_t, \quad t\in[\tau_{\ell-1},\tau_\ell], \tag{24}$$

with the initial condition $Y_{\tau_{\ell-1}} = y_{\tau_{\ell-1}}$, where time is measured in units of generations and $\tau_{\ell-1}$ is the time at which Epoch ℓ begins (Durrett, 2008, Section 7.2). The quantity $\sqrt{Y_t(1-Y_t)}$ in Equation (24) controls random fluctuations due to drift whereas the quantity $\mathcal{M}_\ell(y)$ describes the deterministic change in the mean frequency of the allele over time due to mutation and selection and is given by

$$\mathcal{M}_\ell(y) = u_{aA}^{(\ell)} - (u_{aA}^{(\ell)}+u_{Aa}^{(\ell)})y + y(1-y)[h_\ell s_\ell(1-2y)+s_\ell y], \tag{25}$$

In Equation (25) we have rescaled the usual form of $\mathcal{M}_\ell$ so that time is measured continuously in units of generations.

If the drift term in Equation (24) is negligible compared with $\mathcal{M}_\ell(Y_t)$, then Equation (24) can be approximated by the ordinary differential equation

$$\frac{dy_t^\infty}{dt} = \mathcal{M}_\ell(y_t^\infty), \tag{26}$$

where we may write y_t^∞ instead of Y_t because the evolution of the allele frequency is deterministic and follows its expectation in the absence of drift.

We can also suppress the explicit dependence on the epoch ℓ by defining $\mathcal{M}_t(y_t^\infty) \equiv \mathcal{M}_{\ell_t}(y_t^\infty)$, yielding

$$\frac{dy_t^\infty}{dt} = \mathcal{M}_t(y_t^\infty), \quad y_0^\infty = y_0,\ t\in[0,\tau_L], \tag{27}$$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which holds for the full population history across all epochs $\ell = 1, ..., L$. Equation (27) can be solved numerically, for instance by choosing a sufficiently small time step $\triangle t$ and iteratively computing $y^{\infty}_{t+\triangle t} = \mathcal{M}_t(y^{\infty}_t)\triangle t$.

3.4.3. *Sample probabilities based on deterministic allele frequency trajectories.* To compute the probability $\mathbb{P}^{\infty}_{\Theta}\{O_{[1:K]} = o_{[1:K]}\}$ under either the discrete Wright-Fisher or diffusion models when drift is negligible, we note that the observations $(O_1, ..., O_K)$ are conditionally independent of one another, given the underlying allele frequencies. Thus, in the absence of drift we have

$$\mathbb{P}^{\infty}_{\Theta}\{O_{[1:K]} = o_{[1:K]}\} = \prod_{k=1}^{K} \mathbb{P}_{\Theta}\{O_k = o_k | Y_{t_k} = y^{\infty}_{t_k}\} \qquad (28)$$

for both the diffusion and Wright-Fisher models, where $y^{\infty}_{t_k}$ is the deterministic allele frequency at time t_k, for $k = 1, ..., K$. Using Equations (22) and (28), the probability of the data under the Wright-Fisher model in a population without drift can be obtained using Procedure 3. Similarly, using Equations (27) and (28), the probability of the data in the case of the diffusion model is given by Procedure 4.

Acknowledgments

This research was supported by the National Institutes of Health (grant number R01-GM094402) and by a Packard Fellowship for Science and Engineering.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendix A. Diffusion transition densities: background

The equations in Section 3.2 were derived under a model in which the selected allele A evolves under the diffusion approximation in a population of piecewise constant size. Given that allele A has frequency x at a fixed time s, the density at a later time t is given by the transition density of the diffusion approximation (Equation 10). Steinrücken *et al.* (2014) derived a formula for the density for the case of a single population of constant size. Here, we review this derivation to provide background and notation for the derivation of the diffusion model probability computed in Procedure 2.

A.1. The diffusion approximation in a population of constant size.

Let $p_\ell(s,t;x,y)$ denote the transition density restricted to a specific epoch ℓ of constant size with $s,t \in \ell$. The density $p_\ell(s,t;x,y)$ is the unique solution of the Kolmogorov backward equation,

$$\frac{\partial p_\ell(s,t;x,y)}{\partial t} = \frac{1}{2N_\ell}\mathcal{L}_\ell p_\ell(s,t;x,y) \qquad \text{(A.1)}$$

satisfying the terminal condition $\rho_s(y) = \delta(y-x)$, where $\delta(\cdot)$ is the Dirac delta distribution and $\mathcal{L}_\ell$ is the Kolmogorov backward operator in the epoch defined in Equation (A.2). The factor $1/2N_\ell$ in Equation (A.1), is introduced so that the time-scaling is the same in all epochs, and time is measured continuously in units of generations.

The Kolmogorov backward operator is defined in terms of the scaled mutation and selection parameters $\beta_\ell = 4N_\ell u_{aA}^{(\ell)}$, $\alpha_\ell = 4N_\ell u_{Aa}^{(\ell)}$, and $\sigma_\ell = N_\ell s_\ell$ as

$$\mathcal{L}_\ell = \frac{1}{2}\xi^2(x)\frac{\partial^2}{\partial x^2} + \mu_\ell(x)\frac{\partial}{\partial x}, \qquad \text{(A.2)}$$

where the quantity

$$\xi^2(x) = x(1-x) \qquad \text{(A.3)}$$

captures the contribution to the change in allele frequency arising from genetic drift and

$$\mu_\ell(x) = \frac{1}{2}[\beta_\ell - (\beta_\ell + \alpha_\ell)x] + 2x(1-x)[h_\ell\sigma_\ell(1-2x) + \sigma_\ell x] \qquad \text{(A.4)}$$

captures the contribution from recurrent mutation and selection.

Song and Steinrücken (2012) showed that $p_\ell(s,t;x,y)$ can be expressed as an expansion in the eigenfunctions of $\mathcal{L}_\ell$ of the form

$$p_\ell(s,t;x,y) = \sum_{n=0}^{\infty} e^{-\lambda_{\ell,n}(t-s)/2N_\ell}\frac{\pi_\ell(y)B_{\ell,n}(x)B_{\ell,n}(y)}{\langle B_{\ell,n}, B_{\ell,n}\rangle_{\pi_\ell}}, \qquad \text{(A.5)}$$

where $\{B_{\ell,n}(x)\}_{n=0}^\infty$ are the eigenfunctions of $\mathcal{L}_\ell$ with associated eigenvalues $\{\lambda_{\ell,n}\}_{n=0}^\infty$ and the function $\pi_\ell(y)$ is given by

$$\pi_\ell(y) = e^{\bar{\sigma}_\ell(y)}y^{\beta_\ell - 1}(1-y)^{\alpha_\ell - 1}, \qquad \text{(A.6)}$$

where $\bar{\sigma}_\ell(y) = 4h_\ell\sigma_\ell y(1-y) + 2\sigma_\ell y^2$. The inner product $\langle f, g\rangle_\omega$ with respect to a weight function $\omega(x)$ in Equation (A.5) is defined for two functions f and g on an interval $[a,b]$ by

$$\langle f, g\rangle_\omega = \int_a^b f(x)g(x)\omega(x)dx. \qquad \text{(A.7)}$$

In Equation (A.5), the inner product $\langle\cdot,\cdot\rangle_{\pi_\ell}$ is taken over the interval $[0,1]$ with respect to $\pi_\ell(y)$.

A.2. Expressions for the quantities in Equation (A.5).

Expressions for the eigenvalues $\{\lambda_{\ell,n}\}_{n=0}^\infty$, eigenfunctions $\{B_{\ell,n}(y)\}_{n=0}^\infty$, and inner products $\{\langle B_{\ell,n}, B_{\ell,n}\rangle\}_{n=0}^\infty$ in Equation (A.5) can be obtained using a matrix formulation developed by Steinrücken *et al.* (2014). In particular, the eigenfunctions $\{B_{\ell,n}(y)\}_{n=0}^\infty$ can be expressed as

$$B_{\ell,n}(y) = \sum_{m=0}^{\infty} w_{\ell,n,m}e^{-\bar{\sigma}_\ell(y)/2}R_m^{(\beta_\ell,\alpha_\ell)}(y), \qquad \text{(A.8)}$$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $R_m^{(\alpha,\beta)}(y) = p_m^{(\beta-1,\alpha-1)}(2y-1)$ and $p_m^{(a,b)}(y)$ is the mth classical Jacobi polynomial (Abramowitz and Stegun, 1972, Chapter 22). The vector $\mathbf{w}_{\ell,n} = (w_{\ell,n,0}, w_{\ell,n,1}, \ldots)$ is the nth left eigenvector of the infinite-dimensional matrix

$$\mathbf{M}_\ell := -\left(\boldsymbol{\Upsilon}^{(\alpha_\ell,\beta_\ell)} + \sum_{r=0}^{4} q_{\ell,r}\mathbf{G}_\ell^r\right) \tag{A.9}$$

corresponding to the nth eigenvalue $\lambda_{\ell,n}$, where $\boldsymbol{\Upsilon}^{(\alpha,\beta)} = \mathrm{diag}(\upsilon_0^{(\alpha,\beta)}, \upsilon_1^{(\alpha,\beta)}, \ldots)$ is the diagonal matrix with elements given by $\upsilon_n^{(\alpha,\beta)} = \frac{1}{2}n(n+\alpha+\beta-1)$ and the quantities $q_{\ell,r}$ are given by

$$\begin{aligned}
q_{\ell,0} &= \alpha_k h_\ell \sigma_\ell,\\
q_{\ell,1} &= -(2+3\alpha_\ell+\beta_\ell-2h_\ell\sigma_\ell)h_\ell\sigma_\ell + (1+\alpha_\ell)\sigma_\ell,\\
q_{\ell,2} &= (2+2\alpha_\ell+2\beta_\ell+4\sigma_\ell-10h_\ell\sigma_\ell)h_\ell\sigma_\ell - \\
&\qquad (1+\alpha_\ell+\beta_\ell)\sigma_\ell,\\
q_{\ell,3} &= 16h_\ell^2\sigma_\ell^2 + 2\sigma_\ell^2(1-6h_\ell),\\
q_{\ell,4} &= -2\sigma_\ell^2(1-2h_\ell)^2.
\end{aligned} \tag{A.10}$$

The matrix $\mathbf{G}_\ell^r$ in Equation (A.9) has elements given by

$$[\mathbf{G}_\ell]_{n,m} = \begin{cases} \frac{(n+\alpha_\ell-1)(n+\beta_\ell-1)}{(2n+\alpha_\ell+\beta_\ell-1)(2n+\alpha_\ell+\beta_\ell-2)}, \\ \quad \text{if } m = n-1 \text{ and } n > 0, \\ \frac{1}{2} - \frac{\beta_\ell^2-\alpha_\ell^2-2(\beta_\ell-\alpha_\ell)}{2(2n+\alpha_\ell+\beta_\ell)(2n+\alpha_\ell+\beta_\ell-2)}, \\ \quad \text{if } m = n \text{ and } n \geq 0, \\ \frac{(n+1)(n+\alpha_\ell+\beta_\ell-1)}{2(2n+\alpha_\ell+\beta_\ell)(2n+\alpha_\ell+\beta_\ell-1)}, \\ \quad \text{if } m = n+1 \text{ and } n \geq 0, \\ 0, \\ \quad \text{otherwise}, \end{cases} \tag{A.11}$$

which correspond to the coefficients of the three-term recurrence relation satisfied by the Jacobi Polynomials.

A.3. **Matrix expressions for the transition density.** It is computationally and notationally simpler to express the eigenfunctions of $\mathcal{L}_\ell$ and the transition density as products of matrices. In particular, we can express Equation (A.8) as

$$B_{\ell,n}(y) = e^{-\bar{\sigma}_\ell(y)/2}\mathbf{w}_{\ell,n}\mathbf{R}^{(\alpha_\ell,\beta_\ell)}(y), \tag{A.12}$$

where

$$\mathbf{R}^{(\alpha,\beta)}(y) = (R_0^{(\alpha,\beta)}(y), R_1^{(\alpha,\beta)}(y), \ldots)^T \tag{A.13}$$

and we can express the vector $\mathbf{B}_\ell(y)$ of eigenfunctions as

$$\begin{aligned}
\mathbf{B}_\ell(y) &= (B_{\ell,0}(y), B_{\ell,1}(y), \ldots)^T \\
&= e^{-\bar{\sigma}_\ell(y)/2}\mathbf{W}_\ell\mathbf{R}^{(\alpha_\ell,\beta_\ell)}(y),
\end{aligned} \tag{A.14}$$

where

$$\mathbf{W}_\ell = \begin{bmatrix} \mathbf{w}_{\ell,0} \\ \mathbf{w}_{\ell,1} \\ \vdots \end{bmatrix} \tag{A.15}$$

is the matrix whose rows are the left eigenvectors of the matrix $\mathbf{M}_\ell$ in Equation (A.9).

Using Equations (A.5) and (A.14), the transition density in a single epoch ℓ can then be expressed as the matrix product

$$p_\ell(s,t;x,y) = \pi_\ell(y)\mathbf{B}_\ell^T(x)\mathbf{C}_\ell^{-1}\mathbf{E}_\ell(t-s)\mathbf{B}_\ell(y), \tag{A.16}$$

where

$$\mathbf{E}_\ell(t) = \mathrm{diag}\{e^{-\lambda_{\ell,0}t/2N_\ell}, e^{-\lambda_{\ell,1}t/2N_\ell}, \ldots\} \tag{A.17}$$

and $\mathbf{C}_\ell = \mathrm{diag}\{\langle B_{\ell,n}, B_{\ell,n}\rangle_{\pi_\ell}\}_{n=0}^\infty$. Steinrücken *et al.* (2014) showed that the matrix $\mathbf{C}_\ell$ in Equation (A.16) can be expressed as

$$\mathbf{C}_\ell = \mathbf{W}_\ell\mathbf{D}_\ell\mathbf{W}_\ell^T, \tag{A.18}$$

where

$$\mathbf{D}_\ell = \mathrm{diag}\{d_0^{(\alpha_\ell,\beta_\ell)}, d_1^{(\alpha_\ell,\beta_\ell)}, \ldots\} \tag{A.19}$$

and

$$d_i^{(\alpha_\ell,\beta_\ell)} = \frac{\Gamma(i+\alpha_\ell)\Gamma(i+\beta_\ell)}{(2i+\alpha_\ell+\beta_\ell-1)\Gamma(i+\alpha_\ell+\beta_\ell-1)\Gamma(i+1)}. \tag{A.20}$$

Thus, the transition density in a single epoch can be computed by constructing matrix $\mathbf{M}_\ell$,

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

computing its eigenvectors $\mathbf{W}_\ell$ and eigenvalues $(\lambda_{\ell,0}, \lambda_{\ell,1}, ...)$, and plugging these into the components of Equation (A.16). In practice, because the matrix $\mathbf{M}_\ell$ has infinite dimension, we approximate it by truncating its dimensions at some large integer M yielding approximate eigenvectors $\{\tilde{\mathbf{w}}_{\ell,n}\}_{n=0}^{M}$ and eigenvalues $\{\tilde{\lambda}_{\ell,n}\}_{n=0}^{M}$. We also truncate the length of the vector $\mathbf{B}_\ell(y)$ at a large integer N. Although these truncations lead to approximate values of the transition density, the approximation can be made arbitrarily precise by taking $N \leq M$ to be sufficiently large.

Appendix B. Recursions for the coefficients $\mathbf{a}_k$ and $\mathbf{b}_k$.

B.1. **Discussion of the problem.** Here, we extend the HMM of Steinrücken *et al.* (2014) to accommodate populations of piecewise constant size. As we noted in Section 3.2, the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ of the data under the diffusion model can be obtained using the equation

$$\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\} = \int_{y=0}^{1} f_K(y)dy, \quad \text{(B.1)}$$

where the quantity $f_K(y)$ is obtained by recursively evaluating Equations (14) and (15). Because $f_k(y)$ and $g_k(y)$ can be expressed as the series $f_k(y) = \pi_{\ell_k}(y)\mathbf{b}_k\mathbf{B}_{\ell_k}(y)$ and $g_k(y) = \pi_{\ell_k}(y)\mathbf{a}_k\mathbf{B}_{\ell_k}(y)$ (Equations 16 and 17), determining $f_k(y)$ and $g_k(y)$ amounts to determining the coefficients $\mathbf{a}_k$ and $\mathbf{b}_k$. Thus, it is useful to develop analogs of the recursions (13) and (14) that apply to the coefficients themselves.

B.2. **Equations for propagating coefficients.** From Equation (14), it can be seen that obtaining $f_k(y)$ from $g_k(y)$ involves only multiplication by a polynomial in y. Thus, the formula for obtaining the coefficients $\mathbf{b}_k$ from the coefficients $\mathbf{a}_k$ does not depend on the population history and, therefore, it can be obtained from results in Steinrücken *et al.* (2014) who derived formulas for the recursion for the case of a population of constant size. However, the formula for obtaining $g_k(y)$ from $f_{k-1}(y)$ (Equation 15) involves the transition probability $p_\Theta(t_{k-1}, t_k; z, y)$, which depends on the population parameters Θ. Thus, it is necessary to account for the population history when computing the coefficients $\mathbf{a}_k$ from the coefficients $\mathbf{b}_{k-1}$.

To obtain $\mathbf{a}_k$ from $\mathbf{b}_{k-1}$, we first consider the more general problem of obtaining the generalized vector of coefficients $\mathbf{a}_k(t)$ from $\mathbf{b}_{k-1}$, where $\mathbf{a}_k(t)$ is defined as the vector of coefficients of the expansion of the generalized density $g_k(y,t)$ defined by

$$\begin{aligned} &g_k(y,t)dy \\ &:= \mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:k-1]} = o_{[1:k-1]}, y \leq Y_t < y + dy\} \\ &= \pi_{\ell_t}(y)\mathbf{a}_k(t)\mathbf{B}_{\ell_t}(y), \end{aligned} \quad \text{(B.2)}$$

i.e., the joint density of the observed data up to sample $k-1$ and the allele frequency at time t, where we assume $t_{k-1} \leq t$ so that the time t at which $g_k(y,t)$ is evaluated occurs later than the time t_{k-1} at which $f_k(y)$ is evaluated. The generalized density $g_k(y,t)$ is related to the density $g_k(y)$ defined in Equation (12) by $g_k(y) = g_k(y, t_k)$.

To obtain $\mathbf{a}_k(t)$ from $\mathbf{b}_{k-1}$, there are two scenarios to consider: the case in which both t_{k-1} and t lie within the same epoch ℓ and the case in which t_{k-1} and t lie within distinct epochs. Our derivations of these separate cases provide the results necessary for step 2 of Procedure 2.

B.2.1. *The case $\ell_{t_{k-1}} = \ell_t = \ell$.* If both t_{k-1} and t lie within the same epoch ℓ, then the transition density is given by Equation (A.16) and we have

$$\begin{aligned} &\pi_\ell(y)\mathbf{a}_k(t)\mathbf{B}_\ell(y) \\ &\quad = g_k(y,t) \\ &\quad = \int_0^1 f_{k-1}(z)p_\ell(t_{k-1}, t; z, y)dz \end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$= \int_0^1 \pi_\ell(z)\mathbf{b}_{k-1}\mathbf{B}_\ell(z)\pi_\ell(y)\mathbf{B}_\ell^T(z)\mathbf{C}_\ell^{-1}\times \mathbf{E}_\ell(t-t_{k-1})\mathbf{B}_\ell(y)dz$$
$$= \pi_\ell(y)\mathbf{b}_{k-1}\left[\int_0^1 \pi_\ell(z)\mathbf{B}_\ell(z)\mathbf{B}_\ell^T(z)dz\right]\times \mathbf{C}_\ell^{-1}\mathbf{E}_\ell(t-t_{k-1})\mathbf{B}_\ell(y)$$
$$= \pi_\ell(y)\mathbf{b}_{k-1}\mathbf{E}_\ell(t-t_{k-1})\mathbf{B}_\ell(y), \quad \text{(B.3)}$$

where the second equality follows from Equation (15) and where we have used the fact that $\int_0^1 \pi_\ell(y)\mathbf{B}_\ell(y)\mathbf{B}_\ell^T(y)dy = \mathbf{C}_\ell$. Because the eigenfunctions $\{B_{\ell,n}(y)\}_{n=0}^\infty$ form a complete basis of the Hilbert space defined with respect to the inner product $\langle\cdot,\cdot\rangle_{\pi_\ell}$, the coefficients in the expansion on the left-hand side of Equation (B.3) must equal those on the right-hand side. Thus,

$$\mathbf{a}_k(t) = \mathbf{b}_{k-1}\mathbf{E}_\ell(t-t_{k-1}), \quad \text{if } \ell_{t_{k-1}} = \ell_t. \quad \text{(B.4)}$$

B.2.2. *The case when* $\ell_{t_{k-1}} \neq \ell_t$. If the times t_{k-1} and t lie in different epochs, $\ell_{t_{k-1}}$ and ℓ_t, then the transition density is no longer given by Equation (A.16). Instead, we must use a formula for the transition density across multiple epochs of different sizes. Steinrücken *et al.* (2015) showed that if the allele frequency density $\rho_{\ell,s}(y)$ at time s in epoch ℓ is given by the expansion

$$\rho_{\ell,s}(y) = \pi_\ell(y)\mathbf{r}_{\ell,s}\mathbf{B}_\ell(y), \quad \text{(B.5)}$$

where $\mathbf{r}_{\ell,s} = (r_{\ell,s,0}, r_{\ell,s,1}, ...)$ are the coefficients encoding the density at time s in the basis of the eigenfunctions $\{\mathbf{B}_{\ell,n}(y)\}_{n=0}^\infty$, then at time t in epoch $\ell+1$, the allele frequency density is given by $\rho_{\ell+1,t}(y) = \pi_{\ell+1}(y)\mathbf{r}_{\ell+1,t}\mathbf{B}_{\ell+1}(y)$, where the coefficients $\mathbf{r}_{\ell+1,t}$ are given by

$$\mathbf{r}_{\ell+1,t} = \mathbf{r}_{\ell,s}\mathbf{Z}_\ell(\tau_\ell - s;\boldsymbol{\zeta})\mathbf{E}_{\ell+1}(t-\tau_\ell), \quad \text{(B.6)}$$

where τ_ℓ is the time of the terminating boundary of epoch ℓ, and

$$\mathbf{Z}_\ell(\tau;\boldsymbol{\zeta}) = \mathbf{E}_\ell(\tau)\mathbf{W}_\ell\mathbf{R}_\ell(\boldsymbol{\zeta})\mathbf{H}_{\ell,\ell+1}(\boldsymbol{\zeta})\mathbf{R}_{\ell+1}^{-1}(\boldsymbol{\zeta})\mathbf{W}_{\ell+1}^{-1}. \quad \text{(B.7)}$$

In Equation (B.7), $\mathbf{R}_\ell(\boldsymbol{\zeta})$ and $\mathbf{H}_{\ell,\ell+1}(\boldsymbol{\zeta})$ are given by

$$\mathbf{R}_\ell(\boldsymbol{\zeta}) = \left[\mathbf{R}^{(\alpha_\ell,\beta_\ell)}(\zeta_0), \mathbf{R}^{(\alpha_\ell,\beta_\ell)}(\zeta_1), ...\right], \quad \text{(B.8)}$$

where $\mathbf{R}^{\alpha,\beta}(y)$ is defined in Equation (A.13) and

$$\mathbf{H}_{\ell,\ell+1}(\boldsymbol{\zeta}) = \mathrm{diag}\left\{\frac{\pi_\ell(\zeta_0)e^{-\bar{\sigma}_\ell(\zeta_0)/2}}{\pi_{\ell+1}(\zeta_0)e^{-\bar{\sigma}_{\ell+1}(\zeta_0)/2}}, \frac{\pi_\ell(\zeta_1)e^{-\bar{\sigma}_\ell(\zeta_1)/2}}{\pi_{\ell+1}(\zeta_1)e^{-\bar{\sigma}_{\ell+1}(\zeta_1)/2}}, ...\right\}, \quad \text{(B.9)}$$

for an arbitrary collection of distinct values $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, ...) \in [0,1]$. In practice, we take $\boldsymbol{\zeta}$ to be the Chebyshev nodes (Steinrücken *et al.*, 2015).

By repeated application of Equation (B.6), it follows that if the coefficients $\mathbf{r}_{\ell_s,s}$ encode the density $\rho_s(y)$ at time s in epoch ℓ_s, then the coefficients $\mathbf{r}_{\ell_t,t}$ encoding the density $\rho_t(y)$ at time t in epoch $\ell_t > \ell_s$ are given by $\mathbf{r}_{\ell_t,t} = \mathbf{r}_{\ell_s,s}\mathbf{F}(s,t;\boldsymbol{\zeta})$, where

$$\mathbf{F}(s,t;\boldsymbol{\zeta}) = \mathbf{Z}_{\ell_s}(\tau_{\ell_s} - s;\boldsymbol{\zeta})\left[\prod_{i=\ell_s+1}^{\ell_t-1}\mathbf{Z}_i(\tau_i - \tau_{i-1};\boldsymbol{\zeta})\right]\times \mathbf{E}_{\ell_t}(t-\tau_{\ell_t-1}). \quad \text{(B.10)}$$

Moreover, if we define $\mathbf{r}_{\ell_s,s}(x)$ to be the vector of coefficients encoding the density $\rho(y) = \delta(y-x)$, then it follows from Equation (B.10) that the transition density $p_\Theta(s,t;x,y)$ for times $s < t$ lying in distinct epochs $\ell_s < \ell_t$ is given by

$$p_\Theta(s,t;x,y) = \pi_{\ell_t}(y)\mathbf{r}_{\ell_s,s}(x)\mathbf{F}(s,t;\boldsymbol{\zeta})\mathbf{B}_{\ell_t}(y), \quad \text{if } \ell_s < \ell_t. \quad \text{(B.11)}$$

For the initial condition $\rho_{\ell,s}(y) = \delta(y-x)$, it was shown in Proposition 1 of Steinrücken *et al.*

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2014) that the coefficients $\mathbf{r}_{\ell_s,s}(x)$ are given by

$$\mathbf{r}_{\ell_s,s}(x) = \left(\frac{B_{\ell_s,0}(x)}{\langle B_{\ell_s,0}, B_{\ell_s,0} \rangle_{\pi_{\ell_s}}}, \frac{B_{\ell_s,1}(x)}{\langle B_{\ell_s,1}, B_{\ell_s,1} \rangle_{\pi_{\ell_s}}}, \ldots \right) = \mathbf{B}_{\ell_s}(x)^T \mathbf{C}_{\ell_s}^{-1}, \tag{B.12}$$

yielding

$$p_\Theta(s,t;x,y) = \pi_{\ell_t}(y) \mathbf{B}_{\ell_s}(x)^T \mathbf{C}_{\ell_s}^{-1} \mathbf{F}(s,t;\boldsymbol{\zeta}) \mathbf{B}_{\ell_t}(y), \quad \text{if } \ell_s < \ell_t, \tag{B.13}$$

which is obtained by plugging Equation (B.12) into Equation (B.11).

We can now plug Equation (B.13) into Equation (15) to obtain a relationship between $\mathbf{a}_k(t)$ and $\mathbf{b}_{k-1}$ when times t_{k-1} and t lie in different epochs:

$$\begin{aligned}
&\pi_{\ell_t}(y) \mathbf{a}_k(t) \mathbf{B}_{\ell_t}(y) \\
&\quad = g_k(y,t) \\
&\quad = \int_0^1 f_{k-1}(z) p_\Theta(t_{k-1}, t; z, y) dz \\
&\quad = \int_0^1 \pi_{\ell_{k-1}}(z) \mathbf{b}_{k-1} \mathbf{B}_{\ell_{k-1}}(z) \pi_{\ell_t}(y) \mathbf{B}_{\ell_{k-1}}(z)^T \\
&\qquad\qquad \mathbf{C}_{\ell_{k-1}}^{-1} \mathbf{F}(t_{k-1}, t; \boldsymbol{\zeta}) \mathbf{B}_{\ell_t}(y) dz \\
&\quad = \pi_{\ell_t}(y) \mathbf{b}_{k-1} \Big[\int_0^1 \pi_{\ell_{k-1}}(z) \mathbf{B}_{\ell_{k-1}}(z) \\
&\qquad\qquad \mathbf{B}_{\ell_{k-1}}(z)^T dz \Big] \mathbf{C}_{\ell_{k-1}}^{-1} \mathbf{F}(t_{k-1}, t; \boldsymbol{\zeta}) \mathbf{B}_{\ell_t}(y) \\
&\quad = \pi_{\ell_t}(y) \mathbf{b}_{k-1} \mathbf{F}(t_{k-1}, t; \boldsymbol{\zeta}) \mathbf{B}_{\ell_t}(y),
\end{aligned} \tag{B.14}$$

where we have again used the fact that $\int_0^1 \pi_\ell(z) \mathbf{B}_\ell(z) \mathbf{B}_\ell(x)^T dy = \mathbf{C}_\ell$. Finally, by the uniqueness of expansions in the Hilbert basis $\{B_{\ell_t,n}\}_{n=0}^\infty$, we have

$$\mathbf{a}_k(t) = \mathbf{b}_{k-1} \mathbf{F}(t_{k-1}, t; \boldsymbol{\zeta}), \quad \text{if } \ell_{t_{k-1}} \neq \ell_t. \tag{B.15}$$

The results derived in Section B.2.2 provide the machinery necessary to propagate the coefficients $\mathbf{a}_k$ and $\mathbf{b}_k$ in the HMM over time. These results can now be used to compute the probability of observing a set of sampled allele frequencies under the diffusion model.

B.3. **Derivation of lemmas necessary for Procedure 2.** We now obtain three lemmas that provide the steps in Procedure 2.

Lemma B.3.1. *If the initial frequency density $\rho_0(y)$ at time $t_0 = 0$ is $\rho_0(y) = \delta(y - x)$, then the value of the initial vector $\mathbf{b}_0$ encoding the quantity $f_0(y)$ is given by*

$$\mathbf{b}_0 = \left(\frac{B_{\ell_1,0}(x)}{c_{\ell_1,0}}, \frac{B_{\ell_1,1}(x)}{c_{\ell_1,1}}, \ldots \right) = \mathbf{C}_{\ell_1}^{-1} \mathbf{B}_{\ell_1}(x), \tag{B.16}$$

where $\mathbf{B}_\ell(x)$ is given in Equation (A.14) *and $\mathbf{C}_\ell$ is the diagonal matrix given in Equation* (A.18).

Proof. Because $\mathbf{b}_0$ depends only on the parameters Θ_{ℓ_1} in the first epoch, the proof of Lemma B.3.1 is the same whether we consider a population composed of a single epoch, or a population composed of multiple epochs. The equation for $f_k(y)$ (Equation 16) is the same as Equation 2.14 of Steinrücken *et al.* (2014). Thus, the coefficients $\mathbf{b}_k$ in this paper correspond to the coefficients $\mathbf{b}_k$ in Steinrücken *et al.* (2014) who proved Lemma B.3.1 for the case of a population of constant size. Thus, the first equality in Lemma B.3.1 follows directly from Proposition 1 of Steinrücken *et al.* (2014). The matrix representation in the second equality follows directly from the definitions of $\mathbf{C}_\ell$ and $\mathbf{B}_\ell(x)$. □

Lemma B.3.2. *Let $\mathbf{G}_\ell$, $\mathbf{W}_\ell$, $\mathbf{E}_\ell(t)$, and $\mathbf{F}(s,t;\boldsymbol{\zeta})$ denote the matrices defined in Equations* (A.11), (A.15), (A.17), *and* (B.10), *respectively, where $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \ldots)$ is a set of distinct values arbitrarily chosen such that $\{\zeta_0, \zeta_1, \ldots\} \in [0,1]$. Then the coefficient vectors $\mathbf{a}_k$ and $\mathbf{b}_k$ satisfy the recursive relationships*

$$\mathbf{b}_k = \mathbf{a}_k \mathbf{W}_{\ell_k} \mathbf{G}_{\ell_k}^{o_k} (1 - \mathbf{G}_{\ell_k})^{n_k - o_k} \mathbf{W}_{\ell_k}^{-1}, \tag{B.17}$$

$$\mathbf{a}_k = \begin{cases} \mathbf{b}_{k-1} \mathbf{E}_{\ell_k}(t_k - t_{k-1}) & \text{if } \ell_{k-1} = \ell_k, \\ \mathbf{b}_{k-1} \mathbf{F}(t_{k-1}, t_k; \boldsymbol{\zeta}) & \text{otherwise}, \end{cases} \tag{B.18}$$

where $\mathbf{W}_\ell^{-1} = \mathbf{D}_\ell \mathbf{W}_\ell^T \mathbf{C}_\ell^{-1}$.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Proof. The relationship in Equation (B.18) is obtained immediately by setting $t = t_k$ in Equations (B.4) and (B.15), which follows because $\mathbf{a}_k(t_k) = \mathbf{a}_k$. The relationship in Equation (B.17) does not depend on the population parameters Θ; therefore, Equation (B.17) is the same as that derived in Steinrücken *et al.* (2014), who considered a population of constant size (see Steinrücken *et al.* (2014), Theorem 2). □

Lemma B.3.3. *The probability* $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\}$ *of observing the allele counts* $o_{[1:K]}$*, given the population parameters* Θ *is*

$$\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\} = \frac{c_{\ell_K,0}}{B_{\ell_K,0}(0)} b_{K,0}, \quad \text{(B.19)}$$

where $c_{\ell,0} = [\mathbf{C}_\ell]_{0,0}$ *is element* $0,0$ *of the matrix* $\mathbf{C}_\ell$ *in Equation* (A.18) *and*

$$B_{\ell,0}(0) = \sum_{m=0}^{\infty} (-1)^m [\mathbf{W}_\ell]_{0,m} \frac{\Gamma(m+\alpha_\ell)}{\Gamma(m+1)\Gamma(\alpha_\ell)}. \quad \text{(B.20)}$$

The quantity $[\mathbf{W}_\ell]_{i,j}$ *in Equation* (B.20) *is element* i,j *of the matrix* $\mathbf{W}_\ell$ *given in Equation* (A.15).

Proof. Equation (B.19) can be obtained by integrating over the joint density $f_K(y)$ of the data $O_{[1:K]}$ and the allele frequency Y_{t_K} at the final sampling time:

$$\begin{aligned}
&\mathbb{P}_\Theta\{O_{[1:K]} = o_{[1:K]}\} \\
&\quad = \int_0^1 f_K(y)dy \\
&\quad = \int_0^1 \sum_{n=0}^{\infty} b_{K,n} \pi_{\ell_K}(y) B_{\ell_K,n}(y) dy \\
&\quad = \sum_{n=0}^{\infty} b_{K,n} \int_0^1 \pi_{\ell_K}(y) B_{\ell_K,n}(y) dy \\
&\quad = \sum_{n=0}^{\infty} b_{K,n} \int_0^1 \pi_{\ell_K}(y) B_{\ell_K,n}(y) \frac{B_{\ell_K,0}(y)}{B_{\ell_K,0}(0)} dy \\
&\quad = b_{K,0} \frac{c_{\ell_K,0}}{B_{\ell_K,0}(0)}, \quad \text{(B.21)}
\end{aligned}$$

where $c_{\ell_K,0} = [\mathbf{C}_{\ell_K}]_{0,0} \equiv \langle B_{\ell_K,0}, B_{\ell_K,0} \rangle_{\pi_{\ell_K}}$. In the fourth equality we have used the fact that $B_{\ell,0}(y) = B_{\ell,0}(0)$ is a constant function in y. To see why $B_{\ell,0}(y)$ is constant, note that the eigenvalues $\lambda_{\ell,0}, \lambda_{\ell,1}, ...$ are non-negative and strictly increasing. Thus, all terms in Equation (A.5) must vanish in the limit $s \to -\infty$, except possibly the term $n = 0$. Because $p_\ell(s,t;x,y)$ approaches the stationary density in the limit $s \to -\infty$, it must be the case that $\lambda_{\ell,0} = 0$, so at least one term does not vanish. Thus, we have

$$\lim_{s \to -\infty} p_\theta(s,t;x,y) = \pi_\ell(y) \frac{B_{\ell,0}(y)}{\langle B_{\ell,0}, B_{\ell,0} \rangle_{\pi_\ell}} \propto \pi_\ell(y), \quad \text{(B.22)}$$

where we have used the fact that $\pi_\ell(y)$ is proportional to the stationary density of the diffusion equation in Epoch ℓ. It follows from Equation (B.22) that $B_{\ell,0}(y)$ is constant. Thus, we obtain the result, proving Equation (B.19). Equation (B.20) follows directly from the proof of Proposition 3 in Steinrücken *et al.* (2014). □

Appendix C. Conditional probabilities: the Wright-Fisher model

Under the Wright-Fisher model, the probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]} | S_K\}$ of the observed allele counts, conditional on the event S_K that allele A is segregating in the final sample can be computed using the fact that

$$\begin{aligned}
&\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]} | S_K\} \\
&\quad = \sum_{j=0}^{2N_{t_K}} \mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}, c_{t_K} = j | S_K\} \\
&\quad = \sum_{j=0}^{2N_{t_K}} \frac{\mathbb{P}_{\Theta,\mathcal{W}}\{S_K | O_{[1:K]} = o_{[1:K]}, c_{t_K} = j\}}{\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\}} \times \\
&\qquad\qquad \mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}, c_{t_K} = j\}
\end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$= \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\}} \times \sum_{j=0}^{2N_{\ell_{t_K}}} \mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}, c_{t_K} = j\}$$

$$= \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\}} \sum_{i=0}^{2N_{t_K}} \mathbf{v}_{K,i}, \tag{C.1}$$

where the third equality in Equation (C.1) follows from the fact that the conditional probability $\mathbb{P}_{\Theta,\mathcal{W}}\{S_K|O_{[1:K]} = o_{[1:K]}, c_{t_K} = j\}$ depends only on the allele count o_K and the final equality in Equation (C.1) follows from the definition of $\mathbf{v}_k$. The probability $\mathbb{P}\{S_K|O_K = o_K\}$ in Equation (C.1) is given by

$$\mathbb{P}\{S_K|O_K = o_K\} = \begin{cases} 1, & \text{if } 1 \le o_K < n_K, \\ 0, & \text{otherwise} \end{cases} \tag{C.2}$$

and the probability $\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\}$ is given by

$$\begin{aligned}
&\mathbb{P}_{\Theta,\mathcal{W}}\{S_K\} \\
&= \sum_{i=0}^{2N_{t_K}} \mathbb{P}\{S_K|C_{t_K} = i\}\mathbb{P}_{\Theta,\mathcal{W}}\{C_{t_K} = i\} \\
&= \sum_{i=0}^{2N_{t_K}} [1 - \mathbb{P}\{O_K = 0|C_{t_K} = i\} - \\
&\qquad \mathbb{P}\{O_K = n_K|C_{t_K} = i\}] \times \\
&\qquad\qquad \mathbb{P}_{\Theta,\mathcal{W}}\{C_{t_K} = i\} \\
&= \sum_{i=0}^{2N_{t_K}} \left[1 - \left(1 - \frac{i}{2N_{t_K}}\right)^{n_K} - \left(\frac{i}{2N_{t_K}}\right)^{n_K}\right] \times \\
&\qquad\qquad \mathbb{P}_{\Theta,\mathcal{W}}\{c_{t_K} = i\},
\end{aligned} \tag{C.3}$$

where, as before, $\mathbb{P}_{\Theta,\mathcal{W}}\{C_{t_K} = i\}$ is given by the ith element of $\mathbf{d}_t = \mathbf{d}_0 \prod_{t=1}^{t_K} T_{t-1,t}$.

Note that it is easy to condition on other configurations of the final sample using a procedure similar to that used to derive Equation (C.1). For example, for the event F_K that allele A is segregating or fixed in the final sample, which we consider in Section 1.7.2, the probability $\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|F_K\}$ is given by

$$\mathbb{P}_{\Theta,\mathcal{W}}\{O_{[1:K]} = o_{[1:K]}|F_K\} = \frac{\mathbb{P}\{F_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{W}}\{F_K\}} \sum_{i=0}^{2N_{t_K}} \mathbf{v}_{K,i}, \tag{C.4}$$

where

$$\mathbb{P}\{F_K|O_K = o_K\} = \begin{cases} 1, & \text{if } 1 \le o_K \le n_K, \\ 0, & \text{otherwise} \end{cases} \tag{C.5}$$

and

$$\begin{aligned}
&\mathbb{P}_{\Theta,\mathcal{W}}\{F_K\} \\
&= \sum_{i=0}^{2N_{t_K}} [1 - \mathbb{P}\{O_K = 0|C_{t_K} = i\}]\,\mathbb{P}_{\Theta,\mathcal{W}}\{C_{t_K} = i\} \\
&= \sum_{i=0}^{2N_{t_K}} \left[1 - \left(1 - \frac{i}{2N_{t_K}}\right)^{n_K}\right] \mathbb{P}_{\Theta,\mathcal{W}}\{c_{t_K} = i\}.
\end{aligned} \tag{C.6}$$

Other probabilities can be obtained in a similar fashion.

Appendix D. Conditional probabilities: diffusion model

Under the diffusion approximation, the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}|S_K\}$ of the observed allele counts conditional on the event S_K that allele A is segregating in the final sample can be computed using the fact that

$$\begin{aligned}
&\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}|S_K\} \\
&= \int_{y=0}^{1} \mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}, Y_{t_K} = y|S_K\}dy \\
&= \int_{y=0}^{1} \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{D}}\{S_K\}} f_K(y)dy \\
&= \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{D}}\{S_K\}} \int_0^1 f_K(y)dy \\
&= \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{D}}\{S_K\}} \mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}\} \\
&= \frac{\mathbb{P}\{S_K|O_K = o_K\}}{\mathbb{P}_{\Theta,\mathcal{D}}\{S_K\}} \frac{c_{\ell_K,0}}{B_{\ell_K,0}(0)} b_{K,0},
\end{aligned} \tag{D.1}$$

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where the second equality follows from the fact that the conditional probability $\mathbb{P}\{S_K|O_K = o_K, Y_{t_K} = y\}$ depends only on the allele count o_K in the final sample, and the final equality follows from Equation (B.19).

The probability $\mathbb{P}_{\Theta,\mathcal{D}}\{S_K\}$ can be computed as

$$\begin{aligned}&\mathbb{P}_{\Theta,\mathcal{D}}\{S_K\}\\ &= 1 - \mathbb{P}_{\Theta,\mathcal{D}}\{O_K = 0\} - \mathbb{P}_{\Theta,\mathcal{D}}\{O_K = n_K\}.\end{aligned} \tag{D.2}$$

In Equation (D.2), the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_K = j\}$ can be found easily by noting that if the only sampling time is t_K, at which $O_K = j$ lineages are observed, then the probability computed using Procedure 2 is precisely the probability $\mathbb{P}_{\Theta,\mathcal{D}}\{O_K = j\}$.

Consider the problem in which the only sampling occurs at time t_K and denote the coefficient vectors for this related problem by $\tilde{\mathbf{a}}_k$ and $\tilde{\mathbf{b}}_k$. Then, by Equation (B.19), we see that

$$\mathbb{P}_{\Theta,\mathcal{D}}\{O_K = j\} = \frac{c_{\ell_K,0}}{B_{\ell_K,0}(0)}\tilde{b}_{K,0}(j), \tag{D.3}$$

where $\tilde{b}_K(j)$ is is obtained by computing the steps in Procedure 2. In Step 1, we compute

$$\tilde{\mathbf{b}}_0 = \mathbf{b}_0, \tag{D.4}$$

which follows because the initial vector $\mathbf{b}_0$ depends only on the initial frequency. In Step 2, we compute

$$\tilde{\mathbf{a}}_K = \begin{cases} \tilde{\mathbf{b}}_0\mathbf{E}_{\ell_1}(t_K), & \text{if } \ell_{t_K} = 1,\\ \tilde{\mathbf{b}}_0\mathbf{F}(0, t_K; \boldsymbol{\zeta}), & \text{otherwise,}\end{cases} \tag{D.5}$$

which follows because the coefficients are propagated directly from time $t_0 = 0$ to time t_K. Finally, in Step 3 we have

$$\tilde{b}_K(j) = \tilde{\mathbf{a}}_K\mathbf{W}_{\ell_K}\mathbf{G}^j_{\ell_K}(1 - \mathbf{G}_{\ell_K})^{n_K-j}\mathbf{W}^{-1}_{\ell_K}. \tag{D.6}$$

Combined together, Equations (D.4), (D.5), and (D.6) yield

$$\begin{aligned}&\tilde{\mathbf{b}}_K(j)\\ &= \begin{cases} \mathbf{b}_0\mathbf{E}_{\ell_1}(t_K)\mathbf{W}_{\ell_K}\mathbf{G}^j_{\ell_K}(1 - \mathbf{G}_{\ell_K})^{n_K-j}\mathbf{W}^{-1}_{\ell_K}, \\ \quad \text{if } \ell_{t_K} = 1,\\ \mathbf{b}_0\mathbf{F}(0, t_K; \boldsymbol{\zeta})\mathbf{W}_{\ell_K}\mathbf{G}^j_{\ell_K}(1 - \mathbf{G}_{\ell_K})^{n_K-j}\mathbf{W}^{-1}_{\ell_K}, \\ \quad \text{otherwise.}\end{cases}\end{aligned} \tag{D.7}$$

Plugging Equations (D.2) and (D.3) into Equation (D.1) gives

$$\begin{aligned}&\mathbb{P}_{\Theta,\mathcal{D}}\{O_{[1:K]} = o_{[1:K]}|S_K\}\\ &= \frac{\mathbb{P}\{S_K|O_K = o_K\}c_{\ell_K,0}b_{K,0}(t_K)}{B_{\ell_K,0}(0) - c_{\ell_K,0}\tilde{b}_{K,0}(0) - c_{\ell_K,0}\tilde{b}_{K,0}(n_K)},\end{aligned} \tag{D.8}$$

where

$$\mathbb{P}\{S_K|O_K = o_K\} = \begin{cases} 1, & \text{if } 1 \le o_K < n_K,\\ 0, & \text{otherwise.}\end{cases} \tag{D.9}$$

Note that it is easy to condition on other configurations of the final sample by computing the probabilities $\mathbb{P}\{V_K|O_K = o_k\}$ and $\mathbb{P}_{\Theta,\mathcal{W}}\{V_K\}$ for some other event V_K.

Literature Cited

Abramowitz, M. and Stegun, I. A., editors 1972. *Handbook of mathematical functions with formulas, graphs, and mathematical tables, 9th printing*. Dover, New York.

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

Bonhoeffer, S., Barbour, A. D., and De Boer, R. J. 2002. Procedures for reliable estimation of viral fitness from time-series data. *Proc. R. Soc. Lond. B.*, 269: 1887–1893.

Burke, M. 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc. Roy. Soc. Lond. B*, page rspb20120799.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Clark, A. 1979. The effects of interspecific competition on the dynamics of a polymorphism in an experimental population of Drosophila melanogaster. *Genetics*, 92: 1315–1328.

Clarke, B. and Murray, J. 1962. Changes in gene-frequency in Cepaea nemoralis (L.): the estimation of selective values. *Heredity*, 17: 467–476.

Cook, L. M., Cowie, R. H., and Jones, J. S. 1999. Change in morph frequency in the snail Cepaea nemoralis on the Marlborough Downs. *Heredity*, 82: 336–342.

Cook, L. M., Sutton, S. L., and Crawford, T. J. 2005. Melanic moth frequencies in Yorkshire, an old English industrial hot spot. *Journal of Heredity*, 96: 522–528.

Cowie, R. H. and Jones, J. S. 1998. Gene frequency changes in Cepaea snails on the Marlborough Downs over 25 years. *Biological journal of the Linnean Society*, 65: 233–255.

Durrett, R. 2008. *Probability models for DNA sequence evolution*. Springer Science & Business Media.

Ewens, W. J. 1963. Numerical results and diffusion approximations in a genetic process. *Biometrika*, 50: 241–249.

Ewens, W. J. 2004. *Mathematical Population Genetics: I, 2nd ed.* Springer.

Feder, A. F., Kryazhimskiy, S., and Plotkin, J. B. 2014. Identifying signatures of selection in genetic time series. *Genetics*, 196: 509–522.

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, 10: 253–280.

Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D., and Wegmann, D. 2015. An approximate markov model for the wright-fisher diffusion. *Genetics. doi:10.1534*.

Fisher, R. A. 1922. On the dominance ratio. *Proceedings of the royal society of Edinburgh*, 42: 321–341.

Fisher, R. A. and Ford, E. B. 1947. *The spread of a gene in natural conditions in a colony of the moth Panaxia dominula L.* Oliver & Boyd.

Foll, M., Shim, H., and Jensen, J. D. 2015. WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, 15: 87–98.

Gallet, R., Cooper, T. F., Elena, S. F., and Lenormand, T. 2012. Measuring selection coefficients below 10^{-3}: method, questions, and prospects. *Genetics*, 190(1): 175–186.

Gillespie, J. H. 2010. *Population genetics: a concise guide*. JHU Press.

Goudsmit, J., De Ronde, A., Ho, D. D., and Perelson, A. S. 1996. Human Immunodeficiency Virus fitness in vivo: calculations based on a single zidovudine resistance mutation at codon 215 of reverse transcriptase. *Journal of virology*, 70: 5662–5664.

Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection, Part V: selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23: 838–844.

Harrigan, P. R., Bloor, S., and Larder, B. A. 1998. Relative replicative fitness of zidovudine-resistant Human Immunodeficiency Virus Type 1 isolates in vitro. *Journal of Virology*, 72: 3773–3778.

Hartl, D. L. and Clark, A. G. 2007. *Principles of Population Genetics, 4th ed.* Sinauer Associates.

Haubruge, E. and Arnaud, L. 2001. Fitness consequences of malathion-specific resistance in Red Flour Beetle (Coleoptera: Tenebrionidae) and selection for resistance in the absence of malathion. *Journal of economic entomology*, 94(2): 552–557.

Hein, J., Schierup, M. H., and Wiuf, C. 2005. *Gene Genealogies, Variation and Evolution*.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Oxford University Press, Milton Keynes, U.K.

Illingworth, C. J. R., Parts, L., Schiffels, S., Liti, G., and Mustonen, V. 2012. Quantifying selection acting on a complex trait using allele frequency time series data. *Mol. Biol. Evol.*, 29: 1187–1197.

Jenkins, P. A. and Spanò, D. 2015. Exact simulation of the Wright-Fisher diffusion. arXiv:1506.06998, http://arxiv.org/abs/1506.06998.

Karlin, S. and Taylor, H. 1981. *A second course in stochastic processes, Second Ed.* Academic Press.

Labbé, P., Sidos, N., Raymond, M., and Lenormand, T. 2009. Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long-term cline series. *Genetics*, 182: 303–312.

Lacerda, M. and Seoighe, C. 2014. Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*, 198: 1237–1250.

Long, A., Liti, G., Luptak, A., and Tenaillon, O. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat. Rev. Genet.*, 16: 567–582.

Lynch, M. 1987. The consequences of fluctuating selection for isozyme polymorphisms in daphnia. *Genetics*, 115: 657–669.

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

Manly, B. F. 1985. *The statistics of natural selection*. Chapman & Hall.

Mathieson, I. and McVean, G. 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, 193: 973–984.

Nishino, J. 2013. Detecting selection using time-series data of allele frequencies with multiple independent reference loci. *G3*, 3: 2151–2161.

O'Hara, R. B. 2005. Comparing the effects of genetic drift and fluctuating selection on genotype frequency changes in the scarlet tiger moth. *Proc. Roy. Soc. Lond. B*, 272: 211–217.

Rabiner, L. R. 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. IEEE*, 77: 257–286.

Reimchen, T. E. and Nosil, P. 2002. Temporal variation in divergent selection on spine number in Threespine Stickleback. *Evolution*, 56: 2472–2483.

Rouzine, I. M., Rodrigo, A., and Coffin, J. M. 2001. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiology and molecular biology reviews*, 65: 151–185.

Schaffer, H. E., Yardley, D., and Anderson, W. W. 1977. Drift or selection: a statistical test of gene frequency variation over generations. *Genetics*, 87: 371–379.

Siepielski, A. M., DiBattista, J. D., and Carlson, S. M. 2009. Its about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, 12(11): 1261–1276.

Song, Y. S. and Steinrücken, M. 2012. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*, 190(3): 1117–1129.

Steinrücken, M., Bhaskar, A., and Song, Y. S. 2014. A novel spectral method for inferring general diploid selection from time series genetic data. *Annals of Applied Statistics*, 8(4): 2203–2222.

Steinrücken, M., Jewett, E. M., and Song, Y. S. 2015. Spectraltdf: transition densities of diffusion processes with time-varying selection parameters, mutation rates and effective population sizes. *Bioinformatics*, page btv627.

bioRxiv preprint doi: https://doi.org/10.1101/048355; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stine, O. C. and Smith, K. D. 1990. The estimation of selection coefficients in afrikaners: Huntington disease, porphyria variegata, and lipoid proteinosis. *Am. J. Hum. Genet.*, 46: 452–458.

Topa, H., Jónás, Á., Kofler, R., Kosiol, C., and Honkela, A. 2015. Gaussian process test for high-throughput sequencing time series: application to experimental evolution. *Bioinformatics*, 31: 1762–1770.

Wakeley, J. 2008. *Coalescent theory: An introduction*. Roberts & Company Publishers, Greenwood Village, CO.

Wall, S., Carter, M. A., and Clarke, B. 1980. Temporal changes of gene frequencies in cepaea hortensis. *Biological Journal of the Linnean Society*, 14(3-4): 303–317.

Wilson, S. R. 1980. Analyzing gene-frequency data when the effective population size is finite. *Genetics*, 95: 489–502.

Zhao, L., Lascoux, M., and Waxman, D. 2014. Exact simulation of conditioned wright–fisher models. *Journal of theoretical biology*, 363: 419–426.