A numerical framework for genetic hitchhiking in populations of variable size

Eric Friedlander[1] and Matthias Steinrücken[1,2]

[1]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA
[2]Department of Human Genetics, University of Chicago, Chicago, IL, USA

Abstract

Natural selection on beneficial or deleterious alleles results in an increase or decrease, respectively, of its frequency within the population. Due to chromosomal linkage, the dynamics of the selected site affect the genetic variation at nearby neutral loci in a process commonly referred to as genetic hitchhiking. Changes in population size, however, can yield patterns in genomic data that mimic the effects of selection. Accurately modeling these dynamics is thus crucial to understanding how selection and past population size changes impact observed patterns of genetic variation.

Here, we model the evolution of haplotype frequencies with the Wright-Fisher diffusion to study the impact of selection on linked neutral variation. Explicit solutions are not known for the dynamics of this diffusion when selection and recombination act simultaneously. Thus, we present a method for numerically evaluating the Wright-Fisher diffusion dynamics of two linked loci separated by a certain recombination distance when selection is acting. We can account for arbitrary population size histories explicitly using this approach. A key step in the method is to express the moments of the associated transition density, or sampling probabilities, as solutions to ordinary differential equations. Numerically solving these differential equations relies on a novel accurate and numerically efficient technique to estimate higher order moments from lower order moments.

We demonstrate how this numerical framework can be used to quantify the reduction and recovery of genetic diversity around a selected locus over time and elucidate distortions in the site-frequency-spectra of neutral variation linked to loci under selection in various demographic settings. The method can be readily extended to more general modes of selection and applied in likelihood frameworks to detect loci under selection and infer the strength of the selective pressure.

1 Introduction

Natural selection acting on an allele alters its frequency in a population over time. Beneficial alleles tend to increase in frequency, whereas deleterious alleles tend to decrease in frequency. More general forms of diploid selection like dominance, recessive, heterozygote advantage, or underdominance yield more intricate behavior. Due to chromosomal linkage, the dynamics of the allele under selection impacts the allele frequency trajectories at nearby neutral loci, thus each selection regime results in a characteristic pattern of observed genetic variation. For example, in the case of strong positive selection, the beneficial allele can quickly sweep to fixation. This will cause closely linked neutral sites to increase in frequency as well, despite the alleles being neutral, in a processes referred to as genetic hitchhiking (Maynard Smith and Haigh, 1974; Kaplan et al., 1989). Typically, hitchhiking results in a decrease in genetic diversity around the selected locus. This can be quantified, for example, by examining the heterozygosity around a selected locus, or inspecting the local site-frequency-spectrum (SFS), see for example (Kim and Stephan, 2002, Fig. 2) or (Fay and Wu, 2000, Fig. 1). Numerous successful tools that detect genetic variation under selection in genome-wide sequence data have been implemented to scan for these characteristic patterns, see for example Pavlidis and Alachiotis (2017) or Hejase et al. (2020) for a review.

While selective sweeps leave very distinct patterns, different scenarios of selection may lead to more subtle signals. For example, balancing selection tends to maintain the selected allele at intermediate frequencies over

long evolutionary times (Charlesworth, 2006), whereas selection on polygenic traits is thought to act through concerted small adjustments of frequencies at many loci that affect a trait (Pritchard et al., 2010). In addition to the mode of selection, the demographic history of the population also impacts patterns of genetic diversity and in some cases leads to changes in allele frequencies that mimic the impact of selection (Williamson et al., 2005). Thus, developing methods capable of capturing the dynamics of the interplay of selection, variable population size, and chromosomal linkage is vital to interpreting observed genetic variation and elucidating the underlying population genetic and evolutionary processes.

To this end, we present a numerical method for studying the dynamics of haplotype frequencies at two linked loci in a population of changing size. The loci are separated by a given recombination distance and subject to selective pressure. Our approach can be used to accurately and efficiently compute quantities of interest, including expected allele or haplotype frequency trajectories, and expected heterozygosity. Moreover, it can be used to study statistics of linkage disequilibrium (LD), fundamental to many population genetic tools, for example, to characterize linkage blocks in human genetic data (Berisa and Pickrell, 2015), infer recombination rates (McVean et al., 2002; Spence and Song, 2019), infer complex demographic histories (Ragsdale and Gutenkunst, 2017; Ragsdale and Gravel, 2019), or investigate the polygenic and pleiotropic architecture of complex traits (Bulik-Sullivan et al., 2015; Zabad et al., 2020). Furthermore, our method enables exploring the effects of linked and background selection on the site-frequency-spectrum (SFS), another statistic central to many population genetic analysis aimed at either inferring population genetic parameters from neutral site frequency spectra (Gutenkunst et al., 2009; Kamm et al., 2020) or characterizing selection directly (Boyko et al., 2008; Živković and Stephan, 2011). Many of these and other LD- or SFS-based tools operate under the assumptions of neutrality, and thus it is vital for the correct interpretation of the results to understand how direct and linked selection affects these statistics, while correctly accounting for underlying complex demographic histories. In addition, our methodology provides a tractable way of calibrating existing methods for detecting selection, for example, by determining the size of the genomic footprint for different modes of selection under a given demographic history.

Our method is based on the two-locus Wright-Fisher model, and its diffusion approximation, which has found widespread application in population genetics (see e.g. Ewens, 2010; Durrett, 2008). The Wright-Fisher diffusion is a stochastic process that describes the random dynamics of allele or haplotype frequencies in a population under genetic drift, and can explicitly account for selection, recombination, mutation, and arbitrary population size histories. The dynamics of the Wright-Fisher diffusion can be described by a partial differential operator, or infinitesimal generator. The generator defines a pair of associated partial differential equations (PDEs) known as the Kolmogorov forward and backward equations. The solutions to these equations yield the transition density of the Wright-Fisher diffusion, which in turn gives the likelihoods of allele frequency changes over time. However, no analytic solutions are known to these equations, even in the single-locus case. Instead, numerical approaches have been applied to compute this transition density (Kimura, 1955b,a; Song and Steinrücken, 2012; Bergman et al., 2018; Zhao et al., 2013; He et al., 2020a) or to compute likelihoods of observed genetic data (Jenkins and Song, 2012; Steinrücken et al., 2014; Schraiber et al., 2016; He et al., 2020b).

To motivate our approach, we note in particular the framework used by Williamson et al. (2005) and Gutenkunst et al. (2009), who, in the single locus case, first solved the PDE using a finite difference scheme to obtain the transition density of the Wright-Fisher diffusion. The authors then integrated the resulting solution numerically to obtain the expected SFS, which in turn was used for likelihood-based inference of demographic parameters and selection coefficients. This approach was later extended by Ragsdale and Gutenkunst (2017) to the two-locus case under neutrality. The primary challenge with this approach, however, is that it can become prohibitively computationally expensive and inaccurate, especially in the two-locus case, due to the high dimensionality required for the solution scheme and accumulation of errors in the subsequent numerical integration. Evans et al. (2007) introduced a moment-based approach that circumvents this two-step procedure. The authors derived a system of ordinary differential equations (ODEs) that can be solved more efficiently to directly compute the moments of the diffusion, which can be used to obtain the expected SFS or the probabilities of obtaining given sample configurations. Jouganous et al. (2017) extended this approach to compute the joint SFS for multiple populations. While their approach can in principle include the action of selection, they focus on the inference of complex demographic histories. One hurdle that arises with the inclusion of selection is that the system of ODEs does not "close." That is, the dynamics of the moments of a certain order n depends on moments of order $n + 1$. The authors

address this issue through a Jackknife approximation in which the higher-order moments are approximated from those of lower order.

A straightforward extension of this technique to the two-locus case also results in equations that do not close, due to recombination and selection. However, one can consider alternative representations of the moments which close under recombination but not selection. To this end, Ragsdale and Gravel (2019) use an extension of the moments presented by Hill and Robertson (1968) to derive a framework for the estimation of complex demographic histories using multi-population two-locus summary statistics. Kamm et al. (2016) use a related set of moments, derived in a coalescent framework, to perform fine-scale recombination rate estimation. Although these alternative moments do close the system of ODEs under recombination, they also increase its dimension. Since these alternative moments do not close under selection, one would still need to rely on a Jackknife technique to approximate higher order moments to capture the selection dynamics. For the approximation to be accurate, moments of sufficiently high order have to be considered. However, due to the high dimensionality, increasing the order of these alternative moments can become prohibitive, making it computationally expensive to accurately include selection in this approach.

In this work, we introduce a numerical framework to compute the trajectories of the two-locus moments of the haplotype frequencies, or haplotype sampling probabilities, under the Wright-Fisher diffusion subject to genetic drift, mutation, recombination, and selection. This method can simultaneously capture the impact of selection, recombination, and changing population size on higher order moments, or sampling probabilities, making it a valuable tool for studying patterns of genetic variation arising in populations under selection. To circumvent the high dimensionality required for alternative moment representations, we instead use the regular moments (defined below) in our implementation, combined with a novel efficient approximation of higher order moments to "close" the dynamics for recombination and selection. A similar approach, including selection, is outlined in Section S1.3 of the Appendix in Ragsdale and Gravel (2019). Here, we give an alternative derivation which relies on Dynkin's formula. Moreover, the authors focus on the neutral case in their study, whereas we explore the dynamics of a selected locus and it's effect on linked neutral variation. We explore these dynamics under complex demographic histories. Specifically, we investigate a scenario with a population bottleneck followed by exponential population growth, relevant to the Out-of-Africa bottleneck and subsequent population expansion in human evolution. By comparing to simulations, we show that our method can accurately and efficiently compute expected allele and haplotype frequency trajectories, heterozygosity, linkage disequilibrium, and the folded local SFS in a window around a selected locus. Thus, the outlined method provides the means to further our understanding of patterns observed in genomic data from populations subject to selective pressure, in which the history of changing population sizes may obscure or confound the signal of selection.

This paper is organized as follows. In Section 2, we derive the system of ODEs for the moments of the two-locus diffusion by applying Dynkin's formula to the infinitesimal generator of the Wright-Fisher diffusion. We present our novel moment closure technique in Section 3, and demonstrate its accuracy by comparing with several alternative approaches in Section 4.1.1. In Section 4, we show that our method accurately computes the expected trajectories of allele frequencies, haplotype frequencies, heterozygosity and linkage disequilibrium across a broad range of parameters by comparing it to simulated trajectories. In addition, in Section 5, we show that the method accurately captures the impact of selection at a given locus on the local SFS in a window around the selected site. Finally, in Section 6 we discuss possible applications of our method in composite likelihood frameworks to detect and infer the strength of selection, as well as other strategies for how our method can improve existing inference techniques.

2 Moment Equations

In this section we first give a description of the Wright-Fisher diffusion, which is the mathematical framework we use to develop our method. This is followed by a derivation of the system of ODEs for the moment dynamics, as well as a description of the moment-closure technique essential to the method.

2.1 Wright-Fisher Diffusion

We first introduce the two-locus Wright-Fisher diffusion. Consider a panmictic, diploid population of size N_t ($2N_t$ gametes), at time $t \geq 0$, where time is measured in $2N_0$ generations, where N_0 is the size of the reference population. We model the dynamics of two linked loci, locus A and locus B. Let $E^A = \{1, \ldots, K_A\}$ and $E^B = \{1, \ldots, K_B\}$ represent the set of alleles at locus A and locus B, respectively. It follows that the set of possible two-locus haplotypes is given by $E = E^A \times E^B$, with $K = |E|$. Any haplotype $i \in E$ can be represented by the 2-tuple (i_A, i_B), where i_A and i_B are the alleles at locus A and B, respectively. Let $\boldsymbol{X}(t) = \{X_k(t)\}_{k \in E}$ be the vector of frequencies of the haplotypes in the population at time t, where $\boldsymbol{X}(t)$ takes values in the $(K-1)$-dimensional simplex,

$$\Delta_K = \left\{ \boldsymbol{x} \in \mathbb{R}^K \middle| 0 \leq x_i \leq 1 \, \forall \, i \in E, \sum_{i \in E} x_i = 1 \right\}.$$

For moderate to large population sizes, the dynamics of $\boldsymbol{X}(t)$ can be modeled using the Wright-Fisher diffusion, see for example (Durrett, 2008, Ch. 7) or (Ewens, 2010, Ch. 4). Let $\mathcal{L}_t$ be the infinitesimal generator of the Wright-Fisher diffusion, which we define in more detail below. In general, this generator depends on the time t, allowing us to model changing population sizes. We refer the reader to (Karlin and Taylor, 1981, Ch. 15.2) or (Ethier and Kurtz, 2009, Ch. 10) for a more detailed introduction on the concept of infinitesimal generators for diffusions. The transition density of the Wright-Fisher diffusion, $p_t(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} := \mathbb{P}\{\boldsymbol{X}(t) \in d\boldsymbol{y} | \boldsymbol{X}(0) = \boldsymbol{x}\}$ for $\boldsymbol{x}, \boldsymbol{y} \in \Delta_K$, can be obtained as the solution to the corresponding backward equation,

$$\frac{\partial}{\partial t} p_t(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{L}_t \big\{ p_t(\cdot, \boldsymbol{y}) \big\}(\boldsymbol{x}), \tag{2.1}$$

for $\boldsymbol{y}$ fixed and $\mathcal{L}_t$ acting on $p_t(\boldsymbol{x}, \boldsymbol{y})$ as a function of $\boldsymbol{x}$. Alternatively, the density can be characterized through the corresponding forward equation where $\boldsymbol{x}$ is fixed and the adjoint of $\mathcal{L}_t$ is applied to $p_t(\boldsymbol{x}, \boldsymbol{y})$ as a function of $\boldsymbol{y}$.

In this work, we model the dynamics in a population subject to genetic drift, mutation, natural selection, and recombination. The Wright-Fisher diffusion has the property that the infinitesimal-generator is linear in these population genetic processes. Namely, the generator can be decomposed as follows,

$$\mathcal{L}_t = \mathcal{L}_t^{\text{drift}} + \mathcal{L}^{\text{mut}} + \mathcal{L}^{\text{sel}} + \mathcal{L}^{\text{rec}}$$

where $\mathcal{L}_t^{\text{drift}}$, $\mathcal{L}^{\text{mut}}$, $\mathcal{L}^{\text{sel}}$, and $\mathcal{L}^{\text{rec}}$ represent the contribution to the generator from genetic drift, mutation, selection, and recombination, respectively. These four operators are given as follows:

Genetic Drift Operator: Setting $\lambda_t := \frac{N_0}{N_t}$, the operator representing genetic drift is

$$\mathcal{L}_t^{\text{drift}} \big\{ f(\boldsymbol{x}) \big\} := \frac{1}{2} \sum_{i,j \in E} \lambda_t x_i (\delta_{i,j} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{x}), \tag{2.2}$$

where $\delta_{i,j}$ is the Kronecker delta, see for example (Durrett, 2008, Ch. 8.1). Here the dependence of λ_t on time t allows the model to capture changing population size by varying the strength of genetic drift.

Mutation Operator: Let $m_{i,j}$ for $i, j \in E$ be the per generation probability that a gamete carrying haplotype i mutates to haplotype j. Let $\theta_{i,j} \doteq 4N_0 m_{i,j}$ be the population scaled mutation rate from haplotype i to j and let $\theta_{i,i} = -\sum_{\substack{j \in E \\ j \neq i}} \theta_{i,j}$. The operator corresponding to mutation can then be written as,

$$\begin{aligned} \mathcal{L}^{\text{mut}} \big\{ f(\boldsymbol{x}) \big\} &: = \frac{1}{2} \sum_{i \in E} \sum_{\substack{j \in E \\ j \neq i}} \Big\{ x_j \theta_{j,i} - x_i \theta_{i,j} \Big\} \frac{\partial}{\partial x_i} f(\boldsymbol{x}) \\ &= \frac{1}{2} \sum_{i \in E} \sum_{j \in E} x_j \theta_{j,i} \frac{\partial}{\partial x_i} f(\boldsymbol{x}), \end{aligned} \tag{2.3}$$

see for example (Durrett, 2008, Ch. 8.1). This formulation allows a very general mutation model, where in principle, the alleles at both loci could mutate at the same time. If mutation rates are low, and only one locus mutates at a time, we could use the special case $\theta_{i,j} = \theta_{i_A,j_A}$, if $i_B = j_B$ and $\theta_{i,j} = \theta_{i_B,j_B}$, if $i_A = j_A$.

Selection Operator: We assume a model of general diploid selection. To this end, let $s_{i,j}$ for $i, j \in E$ be the fitness advantage of a diploid individual carrying the haplotypes i and j, that is, the probability that this individual is the parent of an offspring individual in the next generation is proportional to $1 + s_{i,j}$. Furthermore, let $\sigma_{i,j} = 4N_0 s_{i,j}$ be the population-scaled relative fitness of an individual with haplotypes (i, j). Without loss of generality, we can set $s_{K,K} = 0 = \sigma_{K,K}$. Define the marginal fitness of haplotype $i \in E$ in a population with haplotype frequencies $\boldsymbol{x}$ as,

$$\sigma_i(\boldsymbol{x}) := \sum_{j \in E} \sigma_{i,j} x_j$$

and the mean fitness of the population $\bar{\sigma}$ as,

$$\bar{\sigma}(\boldsymbol{x}) := \sum_{i,j \in E} \sigma_{i,j} x_i x_j = \sum_{i \in E} x_i \sigma_i(\boldsymbol{x}).$$

Using this notation, the selection operator can be expressed as,

$$\mathcal{L}^{\text{sel}}\big\{f(\boldsymbol{x})\big\} := \sum_{i \in E} x_i \big[\sigma_i(x) - \bar{\sigma}(x)\big] \frac{\partial}{\partial x_i} f(\boldsymbol{x}), \tag{2.4}$$

see for example (Steinrücken et al., 2013).

Recombination Operator: Let r be the per generation probability of recombination between locus A and B. For allele $i_A \in E^A$ at locus A and allele $i_B \in E^B$ at locus B, we define the marginal allele frequencies as,

$$x_{(i_A,\cdot)} := \sum_{\ell=1}^{K_B} x_{(i_A,\ell)} \quad \text{and} \quad x_{(\cdot,i_B)} := \sum_{\ell=1}^{K_A} x_{(\ell,i_B)}.$$

The recombination operator is then given by,

$$\mathcal{L}^{\text{rec}}\big\{f(\boldsymbol{x})\big\} := \frac{1}{2} \sum_{i=(i_A,i_B) \in E} \rho(x_{(i_A,\cdot)} x_{(\cdot,i_B)} - x_i) \frac{\partial}{\partial x_i} f(\boldsymbol{x}) \tag{2.5}$$

where $\rho \doteq 4N_t r$ is the population scaled recombination rate between A and B, see (Jenkins and Song, 2012).

2.2 Dynkin's Formula and Moment Dynamics

The transition density $p_t(\boldsymbol{x}, \boldsymbol{y})$ of the Wright-Fisher diffusion describes the dynamics of the haplotype frequencies at two linked loci. Ideally, one would obtain $p_t(\cdot, \cdot)$ analytically buy solving the Kolmogorov forward or backward equations, however, no general analytical solution is known for these PDEs in general. Numerically solving equation (2.1), for example by using finite difference methods, is computationally expensive and challenging in many high-dimensional settings (Gutenkunst et al., 2009; Ragsdale and Gutenkunst, 2017). In order to circumvent solving the PDE of the forward equation (2.1), an alternative technique presented by Evans et al. (2007) provides a less computationally intensive method of directly computing the moments of the transition density. These moments also yield the probability of obtaining a given sample of haplotypes from the population, and thus are also referred to as sampling probabilities. Conveniently, in many applications these moments are the desired quantities, and computing the transition density is just an intermediate step to obtain them (Gutenkunst et al., 2009; Ragsdale and Gutenkunst, 2017; Williamson et al., 2005). The moments can also be used to compute the expected values of many statistics obtained from population genetic data, or for likelihood-based inference of parameters.

Suppose that at time $t \geq 0$ one obtains a sample of haplotypes $\mathbf{n} = \{n_i\}_{i \in E}$ from the population, where n_i denotes the number of haplotypes of type $i \in E$ in the sample. Setting $n = |\mathbf{n}| \doteq \sum_{i \in E} n_i$, $\boldsymbol{y}^{\mathbf{n}} \doteq \prod_{i \in E} y_i^{n_i}$,

and denoting by $\boldsymbol{y}_0$ the haplotype frequencies at time 0, the probability of obtaining such a sample is given by,

$$M_{\mathbf{n}}(t) = C(\mathbf{n}) \int_{\Delta_K} \boldsymbol{y}^{\mathbf{n}} p_t(\boldsymbol{y}_0, \boldsymbol{y})\, d\boldsymbol{y} \tag{2.6}$$

where $C(\mathbf{n}) \doteq \binom{n}{n_1,\ldots,n_K}$ is the corresponding multinomial coefficient. We will frequently refer to $M_{\mathbf{n}}$ as defined above as a "moment of order n" and find it helpful to define the vector of all moments of order n as follows. For each $n \in \mathbb{Z}_+$, define the set $\Gamma_n = \{\boldsymbol{\eta} \in \mathbb{Z}_+^K | \sum_{i\in E} \eta_i = n\}$, and define the vector $\mathbf{M}_n = \{M_{\mathbf{n}}(t)\}_{\mathbf{n}\in\Gamma_n}$ as the collection of all moments of order n. The evolution of $\mathbf{M}_n(t)$ is given as the solution to a system of ODEs derived from the Wright-Fisher diffusion. In previous works, these ODEs have been derived using integration-by-parts (Evans et al., 2007) or the transition probabilities in a finite population (Jouganous et al., 2017; Ragsdale and Gravel, 2019). We give an alternate, more convenient, derivation using Dynkin's formula, see for example (Karlin and Taylor, 1981, Ch. 15) or (Øksendal, 2003, Lemma 7.3.2).

Theorem 2.1 (Dynkin's Formula)**.** *Let $\boldsymbol{X}$ be a diffusion with generator $\mathcal{A}$ and let f be a compactly-supported and twice differentiable function with continuous second derivative. Then if $X(0) = x$,*

$$\mathbb{E}\big[f\big(\boldsymbol{X}(t)\big)\big] = f(x) + \mathbb{E}\left[\int_0^t \mathcal{A}_t f\big(\boldsymbol{X}(s)\big) ds\right].$$

To obtain the requisite system of ODEs describing the evolution of $\mathbf{M}_n(t)$, we apply Dynkin's formula to functions f of the form of the integrand in equation (2.6):

$$f(\boldsymbol{y}) = C(\mathbf{n})\boldsymbol{y}^{\mathbf{n}}.$$

This yields,

$$\begin{aligned} M_{\mathbf{n}}(t) &= M_{\mathbf{n}}(0) + C(\mathbf{n}) \int_{\Delta_K} \int_0^t \mathcal{L}_s \left\{\boldsymbol{y}^{\mathbf{n}}\right\} ds\, p_t(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} \\ &= M_{\mathbf{n}}(0) + C(\mathbf{n}) \int_0^t \int_{\Delta_K} \mathcal{L}_s \left\{\boldsymbol{y}^{\mathbf{n}}\right\} p_s(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y}\, ds, \end{aligned}$$

where the second equality follows from Fubini's Theorem. Taking the derivative with respect to time t on both sides then gives,

$$\dot{M}_{\mathbf{n}}(t) = C(\mathbf{n}) \int_{\Delta_K} \mathcal{L}\left\{\boldsymbol{y}^{\mathbf{n}}\right\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y}.$$

The following theorem is the key representation used in our numerical method.

Theorem 2.2. *The derivative of the moment-vector of order n, $\dot{\mathbf{M}}_n$, can be written as a linear combination of moments of order n, $n+1$, and $n+2$ in the form,*

$$\dot{\mathbf{M}}_n(t) = \mathcal{Q}^n_{drift}\mathbf{M}_n + \mathcal{Q}^n_{mut}\mathbf{M}_n + \mathcal{Q}^n_{sel}\mathbf{M}_{n+2} + \mathcal{Q}^n_{rec}\mathbf{M}_{n+1} \tag{2.7}$$

where $\dot{\mathbf{M}}_n$ is the vector of derivatives for each entry of $\mathbf{M}_n$. Furthermore, when only genic selection is considered $\mathcal{Q}^n_{sel}\mathbf{M}_{n+2} = \mathcal{Q}^n_{genic}\mathbf{M}_{n+1}$ and the order $n+2$ moments are no longer needed.

The proof of this theorem relies on carefully rearranging terms for each individual moment $\dot{M}_{\mathbf{n}}(t)$ and aggregating across moments (see Appendix A for details) to show that,

$$\dot{M}_{\mathbf{n}}(t) = q^{\mathbf{n}}_{\text{drift}}\mathbf{M}_n + q^{\mathbf{n}}_{\text{mut}}\mathbf{M}_n + q^{\mathbf{n}}_{\text{sel}}\mathbf{M}_{n+2} + q^{\mathbf{n}}_{\text{rec}}\mathbf{M}_{n+1}$$

where each $q^{\mathbf{n}}$ is a row vector of appropriate dimension corresponding to the four population genetic processes of interest in this work. The row vectors $q^{\mathbf{n}}$ are then aggregated into matrices $\mathcal{Q}^n$. A detailed proof is provided in Appendix A. With this representation, we can obtain the moments of the transition density by solving this system of ODEs, which is numerically more tractable than solving the multi-dimensional PDE of the Wright-Fisher diffusion. However, note that moments of order higher than n appear on the right-hand side

of equation (2.7). The system of ODEs does not "close" in that the derivative in t of the moments of order n depends on moments of order higher than n. The lack of closure is due to the action of selection and recombination. Therefore, in order to solve these ODEs, one would need to know all moments of all orders. We detail a novel alternative solution to this problem in Section 3, inspired by the approach of Ragsdale and Gutenkunst (2017), in which higher order moments are estimated from those of lower order. Using this approximation effectively "closes" the ODEs and thus makes it tractable to numerically solve the ODEs for a given order n.

3 Moment Closure

In this section, we detail our novel approach to approximate higher order moments and close the ODEs described in the previous section. As we will compare our approach to other approaches in Section 4.1.1, it will be convenient to refer to this method as *logit-linear* throughout the paper. Our approach is motivated by the fact that properly rescaled moments converge to the underlying distribution. The following convergence lemma makes this statement more precise in the one-dimensional case:

Lemma 3.1. *Let μ be a probability measure on $[0,1]$ that has a density $f : [0,1] \to \mathbb{R}_{\geq 0}$. For the sequence of approximating measures,*

$$\mu^n := \bar{f}_n(y)dy$$

with density,

$$\bar{f}_n(y) := \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) (n+1) \mathbb{1}_{[\frac{i-0.5}{n}, \frac{i+0.5}{n})}(y) \tag{3.1}$$

where,

$$b_i^n(z) = \binom{n}{i} z^i (1-z)^{n-i}$$

are the Bernstein polynomials, we have that

$$\mu^n \to \mu$$

converges weakly as $n \to \infty$.

Here $\mathbb{1}_A(y)$ is the indicator function which is equal to 1 if $y \in A$, and equal to 0 if $y \notin A$. We provide a proof of this lemma in Appendix B. Note that the factor $n+1$ in equation (3.1) compensates for the fact that the individual moments computed by the integral converge to 0. This factor does not need to be exactly $n+1$, every factor that grows at same rate as n would lead to the same results. We choose $n+1$ for convenience.

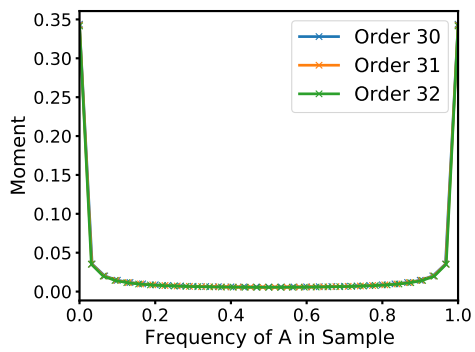
To illustrate Lemma 3.1 in practice, consider a single locus with two alleles A and a. Suppose that the population frequency of the A allele follows a Beta distribution with parameters $\alpha = \beta = 0.1$. The Beta distribution is the stationary distribution under the mutation model considered here. Our aim is to approximate moments of order $n+m$ from those of order n. To illustrate this, we depict in Figure 1(a) the vectors of moments of order 30, 31, and 32. The x-coordinate for the i-th moment of order n is $\frac{i}{n}$, and we interpolate linearly between these values. Since the moments also represent probabilities of obtaining a certain sample configuration, the fraction $\frac{i}{n}$ is also the frequency of the A allele in the sample. Note that in Lemma 3.1, the density $\bar{f}_n$ is defined as a piece-wise constant function, whereas we use linear interpolation here. However, the result from Lemma 3.1 would hold for any suitable interpolation scheme, so we proceed with the linear interpolation for illustration purposes.

We observe that the shapes of the curves for different n are similar, and the values are reasonably close to each other. However, as can be seen in Figure 1(b), when focusing on the values for intermediate frequencies, the exact values indeed differ and decrease as the order n increases. As Lemma 3.1 suggests, the discrepancy between moments of different orders can be compensated for by rescaling the moments appropriately. Indeed, Figure 1(c) shows the same curves, but the moments of order $n = 30$ and $n+1 = 31$ are multiplied by $\frac{n+1}{n+3}$ and $\frac{n+2}{n+3}$, respectively. The resulting curves align well. However, note that the moments are small, and will

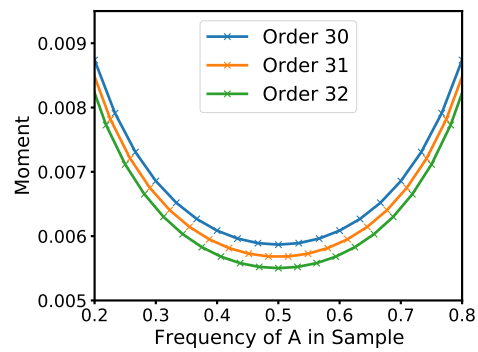
further decrease with increasing n. Moreover, the moments also represent probabilities and the values are in the interval $[0, 1]$. Thus, we can apply the `logit` transformation

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

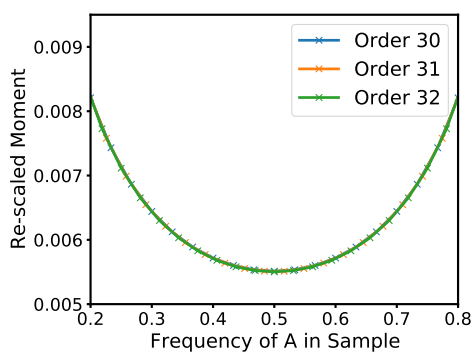
to improve the resolution for small values and further improve the fit. Figure 1(d) shows the resulting transformed values, and suggests approximating higher order moments by interpolating the `logit`-transformed rescaled values.



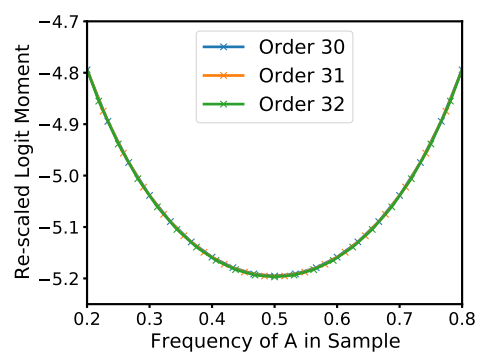
(a) Moments of the Beta distribution with $\alpha = \beta = 0.1$ for order $n = 30$, 31, and 32. The x-axis shows respective frequency of A alleles in the sample represented by the respective moment.



(b) Same moments as Figure 1(a), but only frequencies between 0.2 and 0.8 are shown to highlight the differences and the decreasing trend.



(c) The moments for order $n = 30$ and $n = 31$ are rescaled by $\frac{31}{33}$ and $\frac{32}{33}$, respectively and approximate the moments for $n = 32$ well.



(d) Performing approximations in the `logit`-space increases the resolution near zero.

Figure 1: Graphical representation of the *linear-logit* approximation scheme in the single-locus case. Properly rescaled moments of a certain order can be used to approximate moments of higher order.

For the two-locus case, we require a multi-dimensional version of Lemma 3.1:

Conjecture 3.2. *Let μ be a probability measure on Δ_K with density $f : \Delta_K \to \mathbb{R}_{\geq 0}$. Define the measures μ_n on Δ_K with densities*

$$\bar{f}_n(\boldsymbol{y}) := \sum_{\mathbf{n} \in \Gamma_n} \left(\int_{\Delta_K} C(\mathbf{n}) \boldsymbol{z}^{\mathbf{n}} f(\boldsymbol{z})\, d\boldsymbol{z} \right) (n+1)^{K-1} \left(\prod_{j=1}^{K-1} \mathbb{1}_{\left[\frac{n_j - 0.5}{n}, \frac{n_j + 0.5}{n}\right)}(y_j) \right). \tag{3.2}$$

Then

$$\mu_n \to \mu$$

converges weakly.

This conjecture extends the results of Lemma 3.1 to multiple dimensions, but we do not provide an explicit proof for it here.

By substituting the transition density of the Wight-Fisher diffusion $p_t(\boldsymbol{x}, \boldsymbol{y})$ for the density $f(\boldsymbol{y})$ in equation (3.2), we observe that when the moments of the Wright-Fisher diffusion are properly rescaled they approximate its transition density as the moment order increases. Again, the factor $(n+1)^{K-1}$ reflects the fact that individual moments converge to 0, since the total probability mass is spread out over an increasing number of configurations as n increases. The factor compensates for this effect.

We use Conjecture 3.2 to motivate our moment approximation technique as follows. Convergence of the rescaled moments to the transition density implies that the elements of the approximating sequence are increasingly close to the limit. The individual elements of this sequence of rescaled moments can be used as approximations to each other. Thus we can accurately approximate $M_{\mathbf{n}'}(t)$ for a given configuration $\mathbf{n}'$ with $|\mathbf{n}'| = n+1$ by interpolating the values $M_{\mathbf{n}}(t)$ for certain configurations $\mathbf{n}$ with $|\mathbf{n}| = n$ "close" to $\mathbf{n}'$. Here, we use linear interpolation of the `logit` of the values of $M_{\mathbf{n}}(t)$. We now explicitly define our *logit-linear* approximation in the general case.

Logit-linear Interpolation: To define our moment approximation scheme in the two-locus case, we restrict ourselves to the setting in which both loci are bi-allelic for convenience, that is $|E^A| = |E^B| = 2$ and $K = 4$. Then, the moment vectors can be represented as three dimensional vectors in the 3-simplex Δ_4. As suggested by Conjecture 3.2 and outlined in the single-locus example, we will (1) rescale the moments so that the magnitude is comparable across orders, (2) apply the `logit` transformation to the rescaled moments, (3) perform linear interpolation, and (4) invert the `logit` transformation to obtain the approximate moments of higher order. For numerical stability, we will also (5) rescale the approximate moments to sum to 1.

Suppose we have moments of order n, call them $\mathbf{M}_n$ as above, suppressing the dependence on time, and we wish to estimate moments of order $n+m$, $\mathbf{M}_{n+m}$. In the case of genic selection and recombination (analyzed in Section 4) we will only need to increase the moment order by 1, that is $m = 1$, while for general models of selection $m = 2$ is needed. In order to rescale $\mathbf{M}_n$ to a comparable magnitude as $\mathbf{M}_{n+m}$ we will multiply $\mathbf{M}_n$ by,

$$\kappa(n, m) \doteq \frac{(n+1)^3}{(n+m+1)^3}.$$

Focusing on the details of the interpolation step, recall that the moment vectors $\mathbf{M}_n$ and $\mathbf{M}_{n+m}$ are indexed by the sets Γ_n and Γ_{n+m}, respectively. For any $\tilde{n} \in \mathbb{N}$, define sets $\tilde{\Gamma}_{\tilde{n}} \doteq \{\boldsymbol{\eta} \in \mathbb{R}_+^K | \tilde{n} \times \boldsymbol{\eta} \in \Gamma_{\tilde{n}}\}$. Namely, these are the same as the sets Γ_n and Γ_{n+m} with the elements rescaled so that they are elements of Δ_4, the 3-dimensional simplex. Whereas Γ_n is the set of all moment powers that add up to n, $\tilde{\Gamma}_n$ represents the corresponding configurations in the frequency space Δ_4. For every $n \in \mathbb{N}$ let $\iota_n : \tilde{\Gamma}_n \mapsto \Gamma_n$ map the elements in $\tilde{\Gamma}_n$ to the corresponding configuration in Γ_n. We call this the "sample frequency representation" of a moment since it represents the frequency of each haplotype in a sample of size n. Define the function $f_{n,m} : \Delta_4 \mapsto \mathbb{R}$ where for every $\tilde{\mathbf{n}} \in \tilde{\Gamma}_n$,

$$f_{n,m}(\tilde{\mathbf{n}}) \doteq \text{logit}\left(\kappa(n, m) M_{\iota_n(\tilde{\mathbf{n}})}\right).$$

Specifically $f_{n,m}$ maps the scaled indices $\tilde{\Gamma}_n$ to their corresponding elements of $\mathbf{M}_n$, rescaled by $\kappa(n, m)$ and with the `logit` transformation applied. Given $\boldsymbol{l} \in \Gamma_{n+m}$, $M_{\boldsymbol{l}}$ is estimated by first computing the linear interpolation of $f_{n,m}(\tilde{\mathbf{n}}_1), \ldots, f_{n,m}(\tilde{\mathbf{n}}_4)$, where $\tilde{\mathbf{n}}_1, \ldots, \tilde{\mathbf{n}}_4$ are the four elements of $\tilde{\Gamma}_n$ closest to $\iota_{n+m}^{-1}(\boldsymbol{l})$, and then applying the inverse of the `logit` transformation. We perform this process to obtain each element of $\mathbf{M}_{n+m}$. The final step is to re-normalize the estimated moments $\mathbf{M}_{n+m}$ to sum to one. We found that this scheme increased accuracy and stability when numerically solving the moment ODEs. In Section 4.1.1, we will provide comparison to the following alternative estimation techniques.

3.1 Alternative Techniques

In this section we give a brief outline of other moment approximation techniques that we considered. While these methods all provide viable means of estimating higher order moments, the *logit-linear* method showed the best performance, as we will demonstrate in Section 4.1.1.

Linear Interpolation (no logit): The first alternative considered, referred to as *linear*, is the same as *logit-linear* without applying the `logit` transformation. Specifically, the steps for linear are (1) rescale the moments of order n, (2) perform linear interpolation on the rescaled moments, (3) rescale the resulting approximate moments of order $n+m$ to sum to 1.

"Jackknife" of Jouganous et al. (2017), Ragsdale and Gravel (2019): This moment closure technique was first introduced for the single-locus setting by Jouganous et al. (2017) following ideas presented by Gravel et. al. (2014). The basic approach is to approximate the distribution of the population allele frequency underlying the moments using a quadratic function $a_0 + a_1 x + a_2 x^2$ with coefficients $\boldsymbol{a} = [a_0, a_1, a_2]$, which results in expressions for the moments of any order in terms of $\boldsymbol{a}$. A given moment of order $n+m$, call it m_i, can be approximated as follows: Select the three moments of order n which are "closest" (in the sense of their sample frequency representation), say $m_{\hat{i}-1}, m_{\hat{i}}, m_{\hat{i}+1}$. Since these moments of order n are known quantities, one can derive a system of equations and solve for the parameters $\boldsymbol{a}$. In turn, these parameters can then be used to compute m_i. While assuming the distribution is quadratic may seem restrictive, note that since the $\boldsymbol{a}$'s are not fixed a priori and are recomputed for each m_i, this amounts to assuming the distribution is only locally quadratic, a much milder assumption. For more information we refer to reader to Jouganous et al. (2017).

In Appendix S1.3 of (Ragsdale and Gravel, 2019), the authors present a two-locus implementation of this scheme. Assuming bi-allelic loci, the quadratic function for the population allele frequency is then a three dimensional function on Δ_4 and requires 10 coefficients. Thus, in order to estimate a moment of order $n+m$ (indexed by (i,j,k)), one must select the 10 closest moments of order n (indexed by $(\hat{i},\hat{j},\hat{k})$). The authors further specify a constraint stating that three separate values for each coordinate (i.e. $\hat{i},\hat{j},\hat{k}$) must be included in the set used to estimate the quadratic coefficients. We will refer to this method as *jackknife* and refer the reader to (Ragsdale and Gravel, 2019, Appendix S1.3.5) for a more detailed explanation. In addition, we consider the estimator without the added constraint that each coordinate takes three values which we will call *jackknife-unconstrained*.

Least Squares Projection: Lastly, we also considered a method that we refer to as *least-squares*. To this end, we say that $\mathbf{M}_{n+m}$ and $\mathbf{M}_n$ are "parsimonious" if $\mathbf{M}_n$ is equal to the moments of order n computed from the larger samples in $\mathbf{M}_{n+m}$. Note that if we view the moments as sampling probabilities then $\mathbf{M}_n$ can be computed by downsampling from $\mathbf{M}_{n+m}$. In this case the two vectors will satisfy a set of identities relating them to one another, see Section C.2 for a more detailed description. One can then define a matrix $\mathcal{D}_{n+m,n}$ representing these relations such that $\mathcal{D}_{n+m,n}\mathbf{M}_{n+m} = \mathbf{M}_n$ holds if they are parsimonious. In the *least-squares* estimator we select the vector of moments of order $m+n$ which is parsimonious with the moments of order n and has the smallest sum of squares. In particular we take the estimate of $\mathbf{M}_{m+n}$ to be the solution to the following optimization problem:

$$\min \|\mathbf{M}_{n+m}\|_2 \qquad \text{such that} \quad \mathcal{D}_{n+m,n}\mathbf{M}_{n+m} = \mathbf{M}_n.$$

This is a common optimization problem for which closed form solutions and efficient solvers exist. We refer the reader to (Boyd and Vandenberghe, 2018, Ch. 16) for a detailed introduction.

4 Simulation Results

In this section, to verify the efficiency and accuracy of our novel method, we compare moments obtained using our numerical solutions of the ODEs with the same quantities obtained from simulations conducted using `SimuPOP` (Peng and Kimmel, 2005; Peng and Amos, 2008). Details on our numerical implementation of a step-wise scheme to solve the ODEs can be found in Appendix C.1. We note that in order to make the

simulations computationally feasible, we follow a standard approach to decrease the population size by a factor of 10 and rescaled all other parameters accordingly. Section 4.1 is devoted to validating the model when the genetic variation at the pair of loci separated by a given recombination distance is initialized with fixed haplotype frequencies. We also demonstrate that our novel moment approximation technique outperforms the other candidates outlined in Section 3.1 in the simulated scenarios. In Section 4.2, we investigate scenarios where the genetic variation at the neutral sites is initialized from the stationary distribution, and vary the recombination distance to exhibit linkage effects as they change along the genome. In both sections we consider the case in which all loci are bi-allelic (i.e. four haplotypes at a pair of loci) and only consider the case of genic selection. We consider a recurrent mutation model in which every allele mutates at rate m, that is haplotypes which share the same allele at one locus but differ at the other mutate into each other at rate m. We expect similar accuracy if the mutation probabilities would be locus-specific. The demographic model considered here is an idealized approximation of a European population in the Out-of-Africa model (e.g. Jouganous et al., 2017; Gutenkunst et al., 2009). The ancestral population size is $10,000$, which is reduced to $2,000$ during a bottleneck that starts $4,000$ generations before present. The population stays at this reduced size, until $1,000$ generations before present, when it experiences exponential growth at rate .25% per generation until the present, see Figure 2 for a schematic.

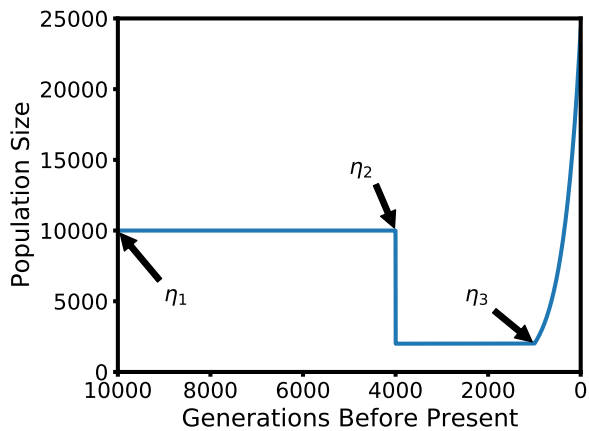


Figure 2: Stylized Out-of-Africa demographic model. The ancestral population size is 10,000, which is reduced to 2,000 during a bottleneck 4,000 generations ago. Exponential growth until the present starts 1,000 generations ago.

By considering different initial times (labeled η_1, η_2, and η_3) for the introduction of the beneficial allele ($10,000$ generations before present, at the beginning of the bottleneck, and at the beginning of the growth phase, respectively), and exhibiting the moments at several points in time afterwards, this model essentially allows us to investigate the dynamics of relaxation to stationarity, during the bottleneck only, or just exponential growth, in a single model, thus covering a wide range of demographic scenarios, providing confidence that our method can accurately account for arbitrary population size histories. In all cases we used moments of order 31 for the numerical solutions. This choice resulted from balancing a high moment order for the accuracy of the moment approximation with computational limitations due to the dimension of the moment vector $\mathbf{M}_n$ increasing cubically with the moment order.

4.1 Fixed Initial Frequencies

In this section, we consider the setting of two loci separated by a recombination distance $\rho = 4N_0 r$. We consider a scenario with two alleles (A and a) at the first locus and two (B and b) at the second locus. At each locus, an allele mutates into the other with rate $\theta = 4N_0 m$ and we consider genic selection on allele A with population scaled selection coefficient $\sigma = 2N_0 s$, where s ranges from 0 to 0.005. In order to validate our results we compare moments obtained from the numerical solution of the ODEs to moments obtained

from simulations using `simuPOP` (Peng and Kimmel, 2005; Peng and Amos, 2008). We consider a variety of different parameter values for the initial haplotype frequencies, selection coefficient, and recombination rate. For the initial haplotype frequencies we vary the initial frequency x_A of the selected allele A and initialize the population such that frequencies of haplotypes AB, Ab, aB, and ab are x_A, 0, $0.5 - x_A$, and 0.5, respectively. The different parameter values used are provided in Table 1 and all combinations of parameters values were tested. We chose these to roughly mirror human genetic parameters, with a per generation per locus mutation probability of $m = 1.25 \cdot 10^{-8}$, and a per generation recombination probability of $r = 10^{-8}$ between two neighboring nucleotide sites. The recombination probability of $r = 10^{-2}$ thus roughly corresponds to a physical distance of $1,000,000$ bp.

Initial Frequency of Beneficial Allele (x_A)	0.01, 0.03, 0.05, 0.1
Selection Coefficient (σ)	0, 0.01, 0.1, 1, 10, 100
Recombination Rate (ρ)	0.0004 ($r = 10^{-8}$), 400 ($r = 10^{-2}$)
Mutation Rate (θ)	0.0005 ($m = 1.25 \cdot 10^{-8}$)

Table 1: Parameters values that are used to compare the numerical method against simulations.

4.1.1 Comparison of Moment Approximation Methodology

We now compare the moments obtained from numerically solving the ODEs when utilizing different moment approximation techniques to the same moments obtained from the simulations described above. As discussed in Section 3, the moment approximation techniques we consider are *logit-linear*, *linear*, *jackknife*, *jackknife-unconstrained*, and *least-squares*. In addition to these methods we test two additional modifications to each, which can be employed in the numerical solution of the ODEs. The first is whether or not to re-normalize the vector of moments after each step of the ODE solver. Specifically, after the solver takes a step the error incurred from the moment approximation may cause the moment vector not to sum to 1, however, we can re-normalize the state vector to sum to 1. The second option is to force the approximated higher-order moment to be "parsimonious" with the lower-order moment. We will defer a detailed description of this modification to Appendix C.2, but it amounts to projecting the original estimate of $\mathbf{M}_{n+1}(t)$ onto the space of order $n+1$ moments which reduce to the original $\mathbf{M}_n(t)$ when the order n moments are computed. Note that by definition the *least-squares* approximation technique already satisfies this requirement so the parsimonious projection does not modify this approximation.

The solutions of the ODEs and simulations were compared as follows. For each combination of parameters in Table 1, we performed 1,000 repetitions using `SimuPOP`. At the end of each repetition, the expression under the integral in equation (2.6) were computed based on the haplotype frequencies in each replicate. These were then averaged over the 1,000 repetitions to get estimates for all moments of order 31. Furthermore, the respective moments were also computed using the numerical approach to solve the ODE with different moment approximation schemes. In order to evaluate the accuracy of the different approximation schemes, we computed (1) the total absolute error and (2) the mean-squared error (MSE) between the simulated moments and those obtained from numerically solving the respective ODEs. In Table 2 and Table 3, for *logit-linear*, *linear*, *jackknife*, *jackknife-unconstrained*, and *least-squares*, we present the combination of renormalization and parsimonious projection which leads to the best results. In all cases, except for *linear*, the combination of options with the lowest absolute error and lowest MSE were the same. For the case of *linear*, including the parsimonious projection resulted in the best absolute error while excluding it resulted in the best MSE. In Table 2 we report the results where we average the errors across all parameter configurations. We note that *logit-linear* outperforms the other methods on both Absolute Error and Mean-Squared Error, by several orders of magnitude in some cases. Similarly, in Table 3 we report the results of the worst performance across all parameter configurations. Specifically, for each combination of approximation type, renormalization, and parsimonious projection we compute the highest Absolute Error and Mean-Squared Error between the different parameter configurations. We then select the approximation type with the minimum (i.e. the mini-max) of these quantities across renormalization, and parsimonious projection. For this metric, the combination of options with the lowest absolute error and lowest MSE were the same with the exception of *Jackknife-Unconstrained*. Again, the scheme *logit-linear* with renormalization outperforms

Approximation Method	Renormalize	Parsimonious	Absolute Error	MSE
Logit-Linear	True	False	**0.000015**	$\mathbf{7.44 \cdot 10^{-8}}$
Linear	True	True	0.000022	$2.94 \cdot 10^{-6}$
Linear	True	False	0.000044	$1.07 \cdot 10^{-6}$
Jackknife	True	True	0.000049	$1.06 \cdot 10^{-5}$
Jackknife-Unconstrained	True	True	0.000043	$8.70 \cdot 10^{-6}$
Least-Squares	True	NA	0.000018	$1.01 \cdot 10^{-6}$

Table 2: Absolute error and mean squared error comparing simulated moments of order 31 with numerical solutions using different moment approximation schemes, averaged over the parameters given in Table 1.

Approximation Method	Renormalize	Parsimonious	Absolute Error	MSE
Logit-Linear	True	False	**0.000024**	$\mathbf{8.44 \cdot 10^{-7}}$
Linear	True	False	0.000120	$1.08 \cdot 10^{-5}$
Jackknife	True	True	0.00025	$1.07 \cdot 10^{-4}$
Jackknife-Unconstrained	True	False	0.00034	$2.84 \cdot 10^{-5}$
Jackknife-Unconstrained	True	True	0.00024	$1.01 \cdot 10^{-4}$
Least-Squares	True	NA	0.000159	$4.30 \cdot 10^{-5}$

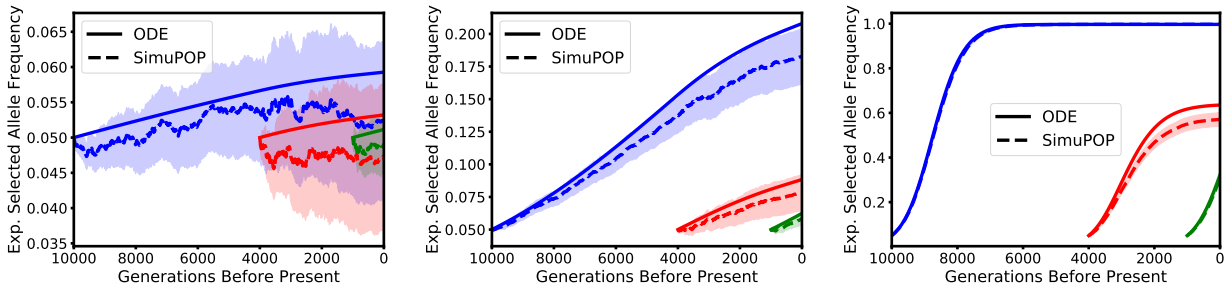
Table 3: Mini-max Absolute error and mean squared error comparing simulated moments of order 31 with numerical solutions using different moment approximation schemes, among the parameters given in Table 1.

all other approximation techniques, in some cases by several orders of magnitude. Based on these simulations we chose *logit-linear* as the moment approximation method used in the remainder of this manuscript.

4.1.2 Comparing Trajectories of Moments

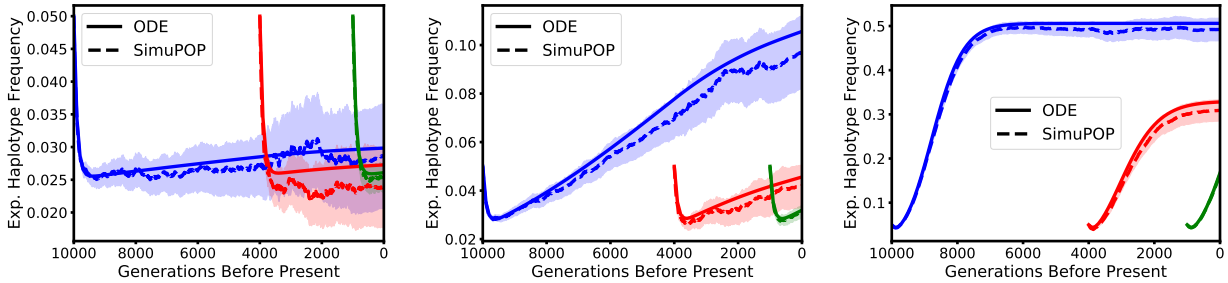
In this section, we will present comparisons of trajectories of the moments between the simulations and the numerical solutions of the ODEs. We begin by noting that moments of lower order can be computed from moments of higher order, and can then subsequently be used to compute quantities of interest, for example expected allele frequencies, expected haplotype frequencies, expected heterozygosity or expected linkage disequilibrium. Thus, we compute the numerical solutions of order 31 first, to ensure sufficient accuracy of the moment approximation, and then compute the moments of lower order that we investigate. In Figures 3, 4, 5, and 6, we compare the dynamics of the solutions to the ODEs with simulations starting the dynamics 10,000 generations before present (η_1), 4,000 generations before present (beginning of the bottleneck, η_2), and 1,000 generations before present (beginning of the exponential growth, η_3). We show the expected values computed from the numerical solution of the ODEs and the mean over the 1,000 simulated replicates, also indicating the region of $\pm 1.96 \times$standard error, corresponding to a 95% confidence interval. Figure 3 displays the expected trajectory of the selected allele A computed using the numerical solutions of the ODEs and from the average of the `SimuPOP` simulations. We show the trajectories for $x_A = 0.05$ and the three largest selection coefficients $\sigma = 1, 10,$ and 100. We can see that the solution to the ODEs accurately tracks the values from the simulations, and thus captures the population dynamics well. As expected, a larger selection coefficient leads to a faster increase of the allele frequency towards fixation at 1. The increase is slower during the bottleneck (red trajectories), since the efficacy of selection is reduced in the smaller population.

Next, in Figure 4, we show the trajectories of the expected frequency of haplotype AB, which carries the beneficial allele A. Here, we show the scenarios with a recombination rate of $\rho = 400$. This choice exhibits a broader spectrum of dynamics, since in the case $\rho = 0.0004$ linkage is very strong, and thus the dynamics are similar to the single locus case in Figure 3 (see Figure D.1). With $\rho = 400$ however, due to the strong recombination, we observe an initial decrease in the frequency of the AB haplotype, since the A allele is only introduced on the B background, and then quickly spreads to the b background as well. Subsequently, the dynamics are then again dominated by selection, and the haplotype AB increases, but only to a frequency of 0.5, since the Ab haplotype will also rise to that frequency in this scenario. Again, the increase is faster for stronger selection. To complement these dynamics, in Figure 5 we also depict the trajectory of the expected



(a) Selection coefficient $\sigma = 1$. (b) Selection coefficient $\sigma = 10$. (c) Selection coefficient $\sigma = 100$.

Figure 3: Expected trajectory of frequency of beneficial allele A. Blue, red, and green correspond to the dynamics starting 10,000 generations, 4,000 generations (beginning of the bottleneck), and 1,000 generations before present (beginning of exponential growth), respectively. The dotted and solid lines correspond to the simulation average and ODE solution, respectively, with the shaded region indicating a 95% confidence interval.



(a) Selection coefficient $\sigma = 1$. (b) Selection coefficient $\sigma = 10$. (c) Selection coefficient $\sigma = 100$.

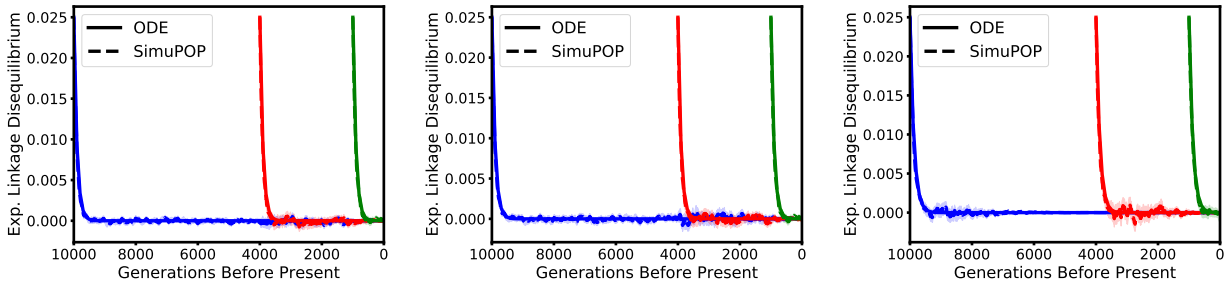
Figure 4: Expected frequency of the AB haplotype, carrying the beneficial allele. Again, blue, red, and green correspond to the dynamics starting 10,000 generations, 4,000 generations (beginning of the bottleneck), and 1,000 generations before present (beginning of exponential growth), respectively. The dotted and solid lines correspond to the simulation average and ODE solution, respectively, with the shaded region indicating the 95% confidence interval.

linkage disequilibrium (LD) between locus A and locus B:

$$\begin{aligned} \mathrm{D}(t) &:= \mathbb{E}\big[X_{(A,B)}(t)X_{(a,b)}(t) - X_{(A,b)}(t)X_{(a,B)}(t)\big] \\ &= \frac{M_{\mathbf{e}_{(A,B)}+\mathbf{e}_{(a,b)}}(t)}{2} - \frac{M_{\mathbf{e}_{(A,b)}+\mathbf{e}_{(a,B)}}(t)}{2}, \end{aligned} \tag{4.1}$$

where $\mathbf{e}_i$ is the vector with 1 in component i, and 0 otherwise. We observe that it rapidly decreases, coinciding with the decrease of the expected frequency of the AB haplotype, but once it reaches zero, the haplotype frequencies starts increasing again.

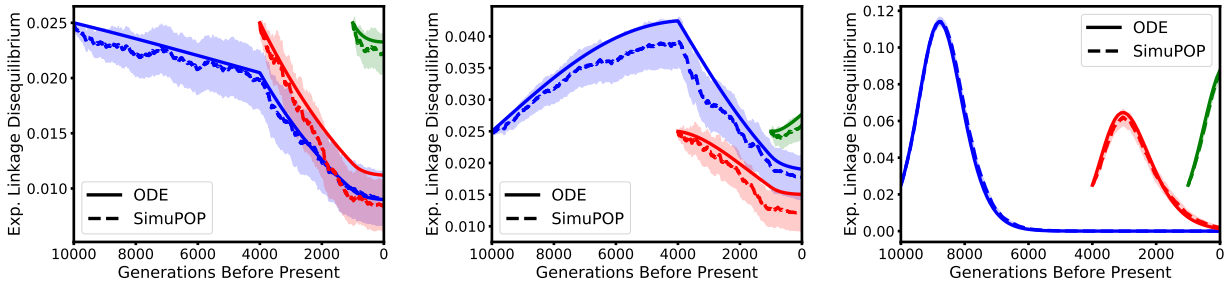
Lastly, we also investigate the dynamics of the expected LD in the scenario with the lower recombination rate $\rho = 0.0004$, displayed in Figure 6. This lower recombination rate leads to much slower decay in LD, allowing us to exhibit the intricate interaction between selection, recombination, and demography. In the case of strong selection, we observe that LD can actually increase, the effect is however dependent on the population size. As can be seen in Figure 6(a) and Figure 6(b), the rate of decrease changes substantially when the bottleneck at 4,000 generations before present and the exponential growth 1,000 generations before present is encountered. In addition to exhibiting the complex dynamics of the genetic variation in the population in these scenarios, the close match between the simulations and the numerical solutions demonstrates that our method accurately captures this dynamics and correctly accounts for the variable population size



(a) Selection coefficient $\sigma = 1$. (b) Selection coefficient $\sigma = 10$. (c) Selection coefficient $\sigma = 100$.

Figure 5: Expected linkage Disequilibrium between locus A and B for $\rho = 400$. Blue, red, and green correspond to the dynamics starting 10,000 generations, 4,000 generations (beginning of the bottleneck), and 1,000 generations before present (beginning of exponential growth), respectively. The dotted and solid lines correspond to the average of the simulations and ODE estimates, respectively, with the shaded region representing a 95% confidence interval.

history.



(a) Selection coefficient $\sigma = 1$. (b) Selection coefficient $\sigma = 10$. (c) Selection coefficient $\sigma = 100$.

Figure 6: Linkage Disequilibrium between locus A and B for $\rho = 0.0004$. Blue, red, and green correspond to the dynamics starting 10,000 generations, 4,000 generations (beginning of the bottleneck), and 1,000 generations before present (beginning of exponential growth), respectively. The dotted and solid lines correspond to the average of the simulations and ODE estimates, respectively, with the shaded region representing a 95% confidence interval.

In summary, we observe that our method captures the population dynamics of the expected frequencies and linkage disequilibrium well. Combined with the results of Section 4.1.1, this provides evidence that our method efficiently and accurately computes the correct dynamics of the genetic variation at two-linked loci across a wide range of parameter settings, including selection at one of the loci.

4.2 Initializing at Stationarity

We are furthermore interested in applying our method to study scenarios in which a novel beneficial allele arises in a population at stationarity, that is, the linked neutral allele is mutation-drift equilibrium. Under the Wright-Fisher diffusion used here, this stationary distribution is given by a Beta(θ, θ) distribution (cf. Durrett, 2008, Sec. 7.5). In this case, the initial haplotype frequencies are not fixed values, but random and follow a certain distribution. Here, we initialize the ODEs for the moments with the moments of this distribution. Characterizing these moments requires additional assumptions about how the new mutation arises in the population. In particular, say the new beneficial allele arises at a low frequency x_A. The new allele will fall either on the B or the b background. The probability of obtaining a sample containing both

haplotype AB and Ab under this distribution is 0, and thus the only non-zero moments are those were either $n_{(A,B)} = 0$ or $n_{(A,b)} = 0$. In addition, the probability of A arising on the background B (or b) is proportional to the frequency of allele B in the population. As a result, the focal allele A tends to arise linked with the major allele more frequently than the minor allele. We give a detailed derivation of the initial moments for this scenario in Appendix C.4.1.

We chose the initial frequency for the beneficial allele as $x_A = 0.05$ and initialize the ODEs using this distribution. This frequency is chosen to be 0.05 because, when initialized with fewer beneficial alleles, the selected allele is purged from the population most of the time, leading to very little difference from the neutral case for the expected quantities considered here . Computing the solution for different recombination rates allows us to explore how the impact of selection on linked neutral variation changes with increasing recombinational distance from the focal site under selection. Again, in the remainder of this section we compare the numerical solutions of the ODEs to the corresponding quantities obtained from simulations using `simuPOP`. We assume the same demographic model as above and, assuming that two adjacent base-pairs are separated by a recombination distance of $\rho = 0.0004$, explore a 100 kbp region around the focal locus under selection. We consider 40 equally spaced neutral loci on either side of the selected side, the first of which is assumed to be perfectly linked (i.e. $\rho = 0$) with the selected locus and the outermost is separated by a recombination distance of $\rho = 50,000 \times 0.0004 = 20$. In the `simuPOP` simulations we simulate a population in which each individual has 41 loci, and for the numerical method, we solve the ODEs for each neutral locus paired with the selected locus, thus we solve it 40 times with different recombination rates. Note that we will mirror the plot in all cases so that the selected locus is in the center and it captures the dynamics of a 100 kbp region with a selected locus in the center. Since each of the 40 loci effectively represents 1,250 nucleotide sites, we also increase the population scaled mutation rate to $\theta = 0.625$ in order to achieve the appropriate levels of neutral diversity. Finally, as in the previous section, we reduce the population size in the simulations by a factor of 10 to allow for tractable simulations and scale the parameters accordingly.

In Figure 7 we examine the expected heterozygosity at the neutral loci for different selection coefficients, defined as

$$\begin{aligned} H(t) &:= \mathbb{E}\big[\big(X_{(a,B)}(t) + X_{(A,B)}(t)\big)\big(X_{(a,b)}(t) + X_{(A,b)}(t)\big)\big] \\ &= \frac{1}{2}M_{\mathbf{e}_{(a,B)}+\mathbf{e}_{(a,b)}}(t) + \frac{1}{2}M_{\mathbf{e}_{(a,B)}+\mathbf{e}_{(A,b)}}(t) + \frac{1}{2}M_{\mathbf{e}_{(A,B)}+\mathbf{e}_{(a,b)}}(t) + \frac{1}{2}M_{\mathbf{e}_{(A,B)}+\mathbf{e}_{(A,b)}}(t), \end{aligned}$$

where A denotes the beneficial allele. The left-most column shows the results when the beneficial allele is introduced 10,000 generations before present (η_1), thus experiencing the full demography, the middle column shows starting from just before the bottleneck (4,000 generations before present, η_2), and the last column shows the results when starting at the beginning of the exponential growth (1,000 generations before present, η_3). In the example starting from just before the bottleneck we initialize the dynamics with the stationary distribution computed based on the larger (pre-bottleneck) population size, while the simulations starting before the exponential growth are assumed to be at stationarity using the reduced size of the bottleneck. Along the rows from top to bottom, the selection coefficients are $\sigma = 1, 50,$ and 100. Note that in all cases, at the time the beneficial allele is introduced, the level of heterozygosity is the same across all neutral loci. As expected, this background level is lower when the beneficial allele is introduced after the bottleneck. Moreover, over time, selection reduces the heterozygosity at the selected locus, and this effect is stronger for larger selection coefficients. The neutral loci most closely linked to the selected locus also experience this reduction in heterozygosity, but the effect diminishes with increasing recombination distance, resulting in the characteristic "V" shape indicating a loss of genetic diversity near the selected locus, the hallmark signature of genetic hitchhiking. At the same time, the overall level of heterozygosity also decreases as the population enters the bottleneck. Moreover, in the case of strong selection, once the beneficial allele has swept to high frequency, the minimum of the "V" starts to increase again, and the signature of the selective sweep starts to erode. For completeness, we provide similar Figures for $x_A = 0.03$ (Figure D.2) and $x_A = 0.01$ (Figure D.3) in the appendix.

To complement these considerations on the dynamics of expected heterozygosity at the neutral sites, we also investigated a measure of LD in the same scenarios. Due to the stationarity assumption at the neutral site, regular LD at the time when the beneficial allele is introduced can be positive or negative with equal probability, and thus the expected LD as given in equation (4.1) is 0 at all later times. For this reason, we

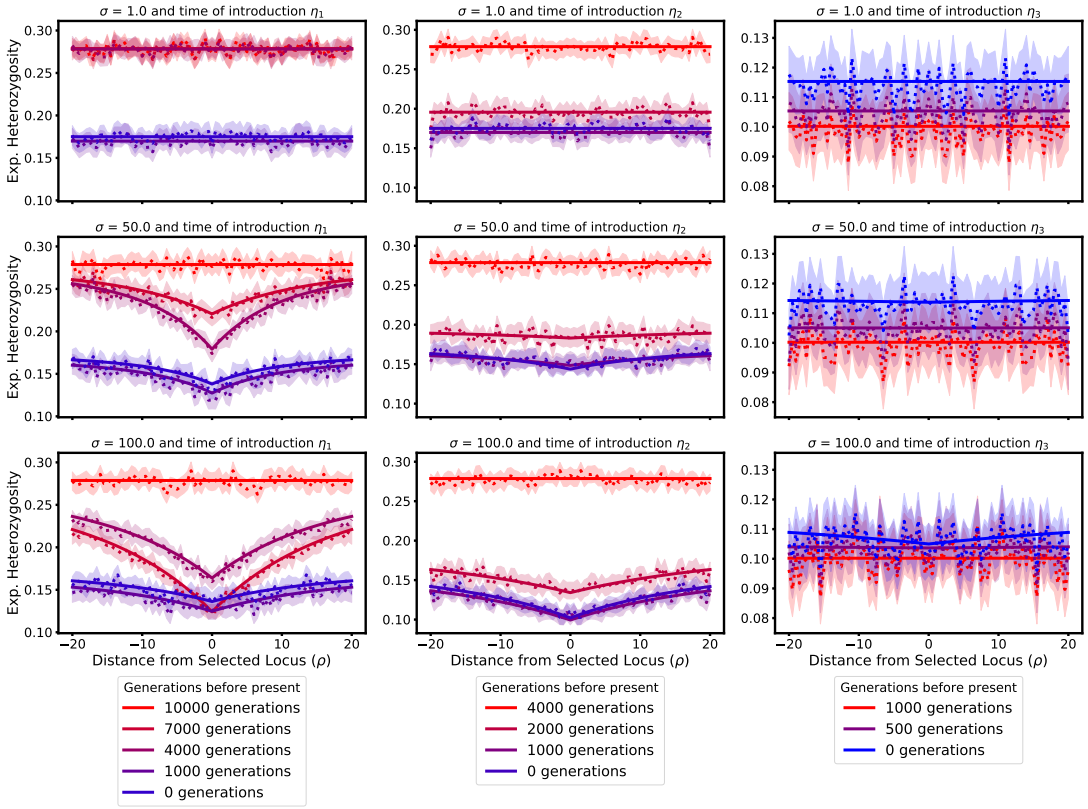


Figure 7: Expected heterozygosity across a 100 kbp region with different selection coefficients and times that the beneficial allele is introduced. The dotted and solid lines represent the simulations and numerical results, respectively, with the shaded region indicating a 95% confidence interval.

consider the expectation of LD squared, that is,

$$D^2(t) := \mathbb{E}\big[\big(X_{(A,B)}(t)X_{(a,b)}(t) - X_{(A,b)}(t)X_{(a,B)}(t)\big)^2\big],$$

which can be expressed in terms of the moments of order 4. Figure 8 shows $D^2(t)$ in the same scenarios and for the same times as Figure 7. Since in the initial distribution, the beneficial allele arises on a given neutral background, that is, it is entirely linked to one neutral allele, at the time of introduction, we see a certain level of D^2. At the loci that are far away from the focal locus, D^2 goes to zero over time. However, closer to the selected locus, the dynamics of the selected allele causes D^2 to increase strongly. Interestingly, for small selection coefficients, in contrast to little impact on the expected heterozygosity, we see a strong impact on D^2. This is likely due to weak selection increasing the allele frequency continuously over long time intervals, and thus constantly building up LD. On the other hand, for strong selection, where the beneficial allele sweeps to fixation quickly, we observe a quick build-up of LD, but this signal also vanishes quickly once the beneficial allele is close to fixation. For the strongest selection coefficient, we do not observe this build-up, since our resolution in time is likely not fine enough. Again, the smaller population size during the bottleneck has a dampening effect. For completeness, we provide similar Figures for $x_A = 0.03$ (Figure D.4) and $x_A = 0.01$ (Figure D.5) in the appendix.

As Figure 7 and Figure 8 show, our method can be used to elucidate the extent of reduction in heterozygosity or inflation of linkage disequilibrium at certain recombination distances from the focal locus, and thus ultimately to characterize how wide the footprint of genetic hitchhiking extends in a variety of scenarios. Moreover, we observe again that the numerical solutions fall within the confidence bounds of the simulations, indicating that our numerical method is highly accurate in capturing the dynamics. In addition, the numerical solutions result in fairly smooth curves for the expected quantities of interest, whereas the curves estimated from the simulations exhibit a large degree of variability, which could only be reduced by

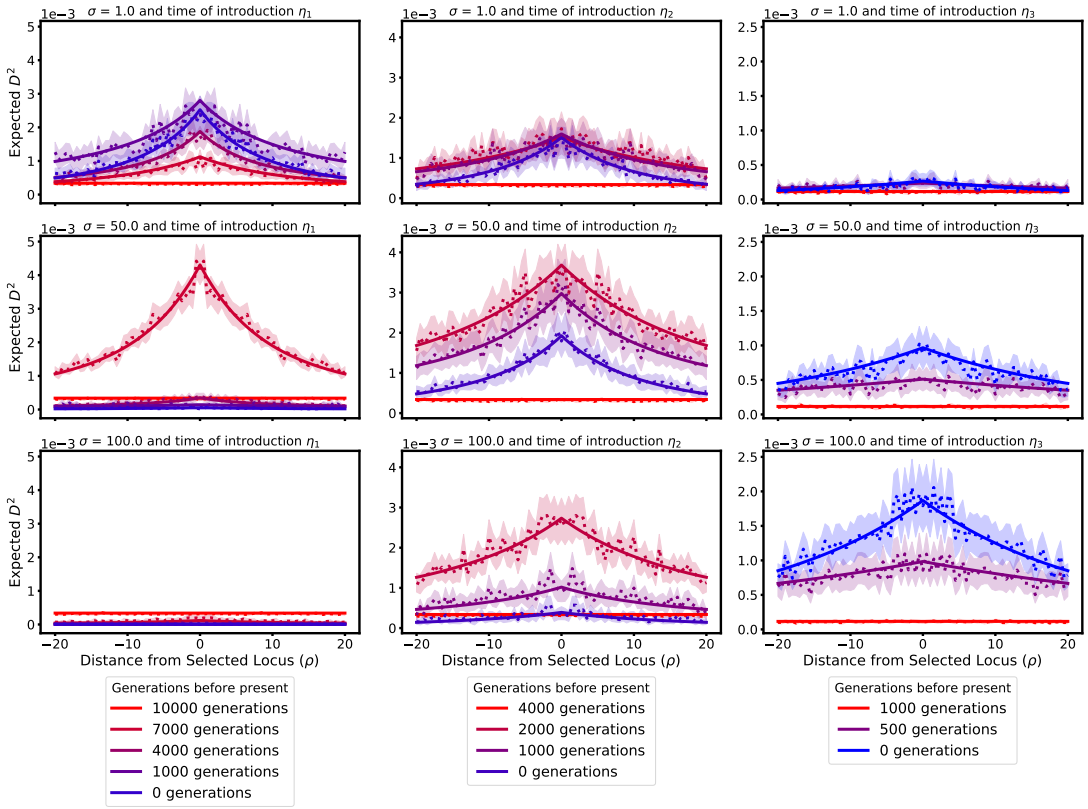


Figure 8: Expected D^2 across a 100 kbp region with different selection coefficients and times that the beneficial allele is introduced. The dotted and solid lines represent the simulation and ODE results, respectively, with the shaded region representing a 95% confidence interval.

performing many additional simulations, and would drastically increase computation time. In cases with much lower mutation rates then considered in our comparisons we would need many replicates of the simulations to even observe a single mutant at the neutral locus, resulting in even more variable estimates of very small quantities. However, the numerical solutions of the ODEs can readily incorporate the respective dynamics and yield accurate results at all scales.

5 Impact of Selection on Local Site-Frequency-Spectrum

In this section, we use our method to investigate the impact of selection at a focal locus on summary statistics of the neutral variation in a surrounding genomic window. Specifically, we will use the numerical solutions of the ODEs to efficiently and accurately compute the expectation of the folded site-frequency-spectrum (SFS) in this genomic window. We will consider only the full demographic model η_1, see Figure 2, that is, the beneficial allele is introduced 10,000 generations before present. We set the per locus mutation rate to $\theta = 0.0005$ and the recombination rate between adjacent sites to $\rho = 0.0004$. In order to compute the expected SFS for the whole window, we compute the two-locus moments using the ODEs over on a grid of 20 different recombination rates where the minimum is $\rho = .0004$, and the maximum is the maximum observed in the region of interest. As we will use window sizes 100 kbp, 500 kbp, and 1 Mbp, with the selected locus in the center, the maximum ρ is $20, 100$, and 200, respectively. and the interior grid-points have logarithmic spacing. For locus separated from the selected locus by a given recombination distance ρ that is not on the grid, we compute its vector of moments by interpolating between the moments computed on the grid. From these two-locus moments, which yield the two locus sampling probabilities, we can compute an individual locus' contribution to the SFS by marginalizing over the selected locus which yields the vector of single-locus

moments (sampling probabilities) at the neutral locus. Specifically, for a given sample size n and number of minor alleles k, we compute the linear combination of two-locus moments

$$S_k^n(t) := \sum_{\substack{i \in E \\ i_B = k}} M_i(t) + \sum_{\substack{i \in E \\ i_b = k}} M_i(t),$$

which yields the contribution to the k-th entry of the expected folded SFS by a neutral locus separated from the focal selected locus by the given recombination distance. Summing these quantities over all loci across the window of interest yields the full folded expected SFS.

We consider three different window sizes (100 kbp, 500 kbp, and 1 Mbp) in which the locus at the center is under genic selection with three different population scaled selection coefficients ($\sigma = 1, 50,$ and 100), and compute the expected folded SFS for a sample of size 31. Again, we note that the choice of this order strikes a balance between accuracy of the moment approximation technique and the size of the moment vector limiting computation. In principle, one could compute the expected SFS for larger samples, by computing the moments of a moderate order, and then using the solutions to approximate the higher orders using the moment approximation techniques outlined in Section 3. The initial frequency of the beneficial allele is $x_A = 0.05$ in each case. To validate the results we compare with simulations using `SLiM` (Haller and Messer, 2019; Haller et al., 2019), generated as follows. First, 1000 `SLiM` replicates are generated, omitting the neutral mutations. The simulations are conditioned on the selected allele reaching the initial frequency of the corresponding ODEs. Specifically, for each `SLiM` replicate, we run the simulation until the selected allele either exceeds x_A or is lost from the population. If it exceeds x_A, we calibrate the demographic model so that this generation is considered the generation when the selected allele is introduced (η_1). If the allele is lost from the simulation, we simply restart the simulation. As the simulations proceed we record the multi-locus genealogies that are simulated by `SLiM` for the entire genomic region and, once the simulation is complete, we compute the branch lengths of the genealogies in order to estimate the expected SFS for that replicate, and average across replicates. Again, for efficiency, we rescale the population sizes by a factor of 10 and adjust the parameters accordingly. In addition to the reduction in population size, there are several approximations and differences to note. The first is that `SLiM` considers an infinite alleles model while the Wright-Fisher diffusion, on which our method is based, uses a recurrent mutation model. However, for low per site mutation rates, this difference has no impact on the *folded* SFS. The second is that conditioning the beneficial allele on reaching frequency 0.05 in the `SLiM` simulation is not exactly equivalent to initializing the ODEs with the selected allele at frequency 0.05 and being completely linked to one neutral allele. This is true for two reasons. The first is that the frequency is conditioned on being *greater than or equal to* 0.05 when the demography is started. The simulation may very well be greater than the threshold, and thus occasionally we may be using a slightly lower initial frequency in the ODE than in the simulations. The second is that `SLiM` allows for recombination to occur before it exceeds the frequency which slightly changes the distribution on the haplotype frequencies at generation zero. However, we argue that these differences are negligible and indeed observe that the numerical solutions to the ODEs and the simulations align very well.

Figure 9 shows the expected folded SFS for the different window sizes and the different selection coefficients. At 10,000 generations before present, when the beneficial allele is introduced, the SFS exhibits the characteristic neutral shape, with an abundance of low frequency minor alleles and fewer alleles at intermediate frequencies. Once the beneficial allele is introduced and subject to selection, for strong selection we observe a reduction in intermediate frequency variants, since these are depleted by selection, as expected. From the start of the bottleneck to the end, we observe a general reduction of genetic diversity, depleting alleles in all frequency classes. Once the exponential growth is encountered, we observe the characteristic inflation of low frequency minor alleles. The effects of selection are most pronounced for a small window size, and barely noticeable for large windows. This is to be expected, since the large windows contain a large number of neutral sites sufficiently removed from the selected locus to not be affected by the dynamics, and thus the neutral variation dominates the SFS. However, note that the total amount of variation in the window (see y-axis in the plots) is proportionately lower in the small windows. Thus, the SFS in the small window is more noisy when computed from real data, indicating that to characterize selection, one needs to strike a balance between large windows, with more statistical power in estimating the SFS, but less impact of selection, and small windows, where the impact of selection is greater, but fewer data points. Lastly, while

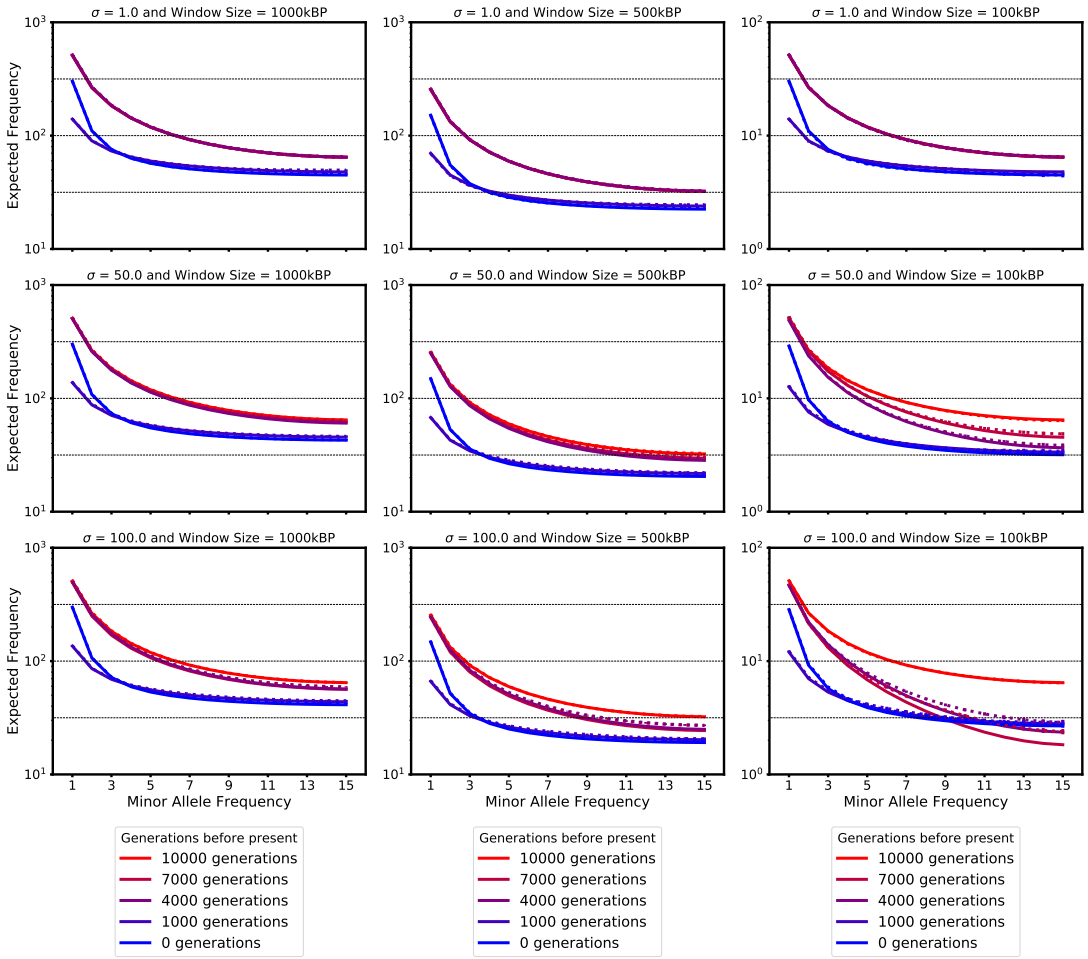


Figure 9: Local folded site-frequency-spectra for windows of size 100 kbp, 500 kbp, and 1 Mbp, for selection coefficients $\sigma = 1$, 50, and 100. The demographic history is given in Figure 2 and the beneficial allele is introduced 10,000 generations before present (η_1). Note that for the third columns the y-axis limits have changed.

the dynamics of the SFS in the given scenarios follows our expectation from population genetic principles, the fact that they are exhibited by our numerical solutions and match closely with the simulations shows again that our novel method indeed captures the relevant population genetic processes accurately.

6 Discussion

We have presented a framework for computing moments of the transition density of the Wright-Fisher diffusion, or equivalently sampling probabilities, at two loci separated by an arbitrary recombination distance when selection is acting on the genetic variation. In addition, the framework accounts for arbitrary population size histories. We have explored our implementation of this framework in the case where genic selection is acting at one of the loci and demonstrated that it accurately captures the dynamics of relevant statistics under genetic hitchhiking in a demographic model that includes both a bottleneck and exponential growth, by comparing it to simulations using `SimuPOP` and `SLiM`. We believe this framework and our implementation enables progress in at least two directions.

On the one hand, it allows exploration of the impact of direct or linked selection on frequently used statistics computed from genomic data. As we have shown, the method can be used to compute a variety of quantities of interest under a range of demographic scenarios and varying strength of selection, more effi-

ciently than using simulations. These quantities include expected allele or haplotype frequency trajectories, heterozygosity, and linkage disequilibrium. As patterns of LD are integral to a variety of existing population genetic tools, our framework provides a means of elucidating the impact of selection and demographic history on LD to allow for a better understanding of the patterns observed in genomic data. In addition, the method can be used to compute the expected local SFS which allows for the exploration of linked and background selection on the SFS. Moreover, the outlined methodology provides a tractable means of calibrating the parameters for existing population genetics tools for detecting selection, for example, by elucidating the size of the genomic footprint under a given mode of selection and population size history.

On the other hand, our numerical framework is a key building block for developing likelihood-based tools to estimate the location, strength, and mode of selection from population genomic data, while accounting for arbitrary demographic scenarios. Including neutral variation around the selected locus into the likelihood model, has increased power over single-locus methods (Pavlidis and Alachiotis, 2017; Hejase et al., 2020). A possible approach is to use the two locus sampling probabilities computed using our method in a composite likelihood framework similar to the one introduced by McVean et al. (2002), that is, multiplying the likelihoods for each locus in a window paired with the candidate selected locus. Such a framework would allow one to compute maximum likelihood estimates of selection coefficients in general models of selection, and apply likelihood-based tests to determine significance.

Our framework and its implementation can be extended into several useful directions. Implementing functionality for general modes of diploid selection requires no additional theoretical developments. However, since general diploid selection requires increasing the order of the approximated moments by two, additional testing needs to be performed in order to ensure its accuracy. Nonetheless, we do expect our novel approximation scheme to generalize well to this case. Moreover, we have presented our framework in the setting of a single panmictic population, accounting for arbitrary population size history. It is natural to consider extending the framework to multiple populations with an arbitrary demographic history. This would enable applying it in situations were a beneficial alleles arises in only one population, and contrasting the genetic variation between populations, to study ancestral selection using the signatures from multiple extant population, or to study selection on introgressed genetic variation. Extending the theoretical framework to multiple populations is possible, since Ragsdale and Gravel (2019) do use multi-population moments of lower order to estimate demographic parameters under neutrality. However, extensions of the framework including selection present additional challenges for the implementation. Obtaining an accurate approximation of the dynamics under selection necessitates solving the ODEs for moments of sufficient high order to ensure the accuracy of the moment approximation when "closing" the ODEs. A straightforward extension wold necessitate computing moments of high order in all populations, and thus the dimensionality equals the number of single population moments exponentiated by the number of populations. Since this number quickly grows to become infeasible, additional approximations will be required. A potential approach could be to compute the single locus dynamic for the selected allele first using higher order moments, and then "attach" the dynamics of lower order moments for the linked neutral loci in a second step. Other approaches could include reducing the ODEs from a system that keeps track of all moments, to a system where only bulk quantities, like sums of moments, are considered. Ultimately, the framework and implementation introduced in this paper present a promising foundation upon which to build these applications and extensions.

Software Availability

A `python` implementation of the numerical procedure as well as scripts to create the figures in this manuscript can be found at `https://github.com/steinrue/two_locus_selection_moments`.

Acknowledgements

We thank Aaron Ragsdale, Simon Gravel, and the population genetics community in Chicago for many inputs and helpful feedback on the methodology and manuscript.

A Derivation of Moments

A.1 Drift

Recall the term of the generator corresponding to genetic drift, $\mathcal{L}_t^{\text{drift}}$, defined in (2.2). Then the contribution of genetic drift to the time derivative of moment $M_{\mathbf{n}}(t)$ is,

$$
\begin{aligned}
&C(\mathbf{n}) \int_{\Delta_K} \mathcal{L}_t^{\text{drift}} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= \frac{C(\mathbf{n})}{2} \int_{\Delta_K} \sum_{i,j \in E} \lambda_t y_i (\delta_{i,j} - y_j) \frac{\partial^2}{\partial y_i \partial y_j} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= \frac{C(\mathbf{n})}{2} \lambda_t \int_{\Delta_K} \sum_{i,j \in E} y_i (\delta_{i,j} - y_j) n_i (n_j - \delta_{i,j}) \boldsymbol{y}^{\mathbf{n} - e_i - e_j} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= \frac{C(\mathbf{n})}{2} \lambda_t \int_{\Delta_K} \left[\sum_{i \in E} (1 - y_i) n_i (n_i - 1) \boldsymbol{y}^{\mathbf{n} - e_i} - \sum_{\substack{i,j \in E \\ i \neq j}} n_i n_j \boldsymbol{y}^{\mathbf{n}} \right] p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= \frac{C(\mathbf{n})}{2} \lambda_t \int_{\Delta_K} \sum_{\substack{i,j \in E \\ i \neq j}} n_i (n_i - 1) \boldsymbol{y}^{\mathbf{n} - e_i + e_j} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} - \frac{1}{2} \lambda_t \sum_{\substack{i,j \in E \\ i \neq j}} n_i n_j M_{\mathbf{n}}(t) \\
&= \frac{1}{2} \lambda_t \sum_{\substack{i,j \in E \\ i \neq j}} \frac{C(\mathbf{n}) n_i (n_i - 1)}{C(\mathbf{n} - e_i + e_j)} M_{\mathbf{n} - e_i + e_j}(t) - \frac{1}{2} \lambda_t \sum_{\substack{i,j \in E \\ i \neq j}} n_i n_j M_{\mathbf{n}}(t) \\
&= \frac{1}{2} \lambda_t \sum_{\substack{i,j \in E \\ i \neq j}} (n_i - 1)(n_j + 1) M_{\mathbf{n} - e_i + e_j}(t) - \frac{1}{2} \lambda_t \sum_{\substack{i,j \in E \\ i \neq j}} n_i n_j M_{\mathbf{n}}(t),
\end{aligned}
$$

where e_i denotes a vector with all components 0, except the i-th entry, which is equal to 1. It follows that we can define $\mathcal{Q}^n_{\text{drift}}$ as follows,

$$
\begin{aligned}
\mathcal{Q}^n_{\text{drift}}(\mathbf{n}, \mathbf{n} - e_i + e_j) &= \tfrac{1}{2} \lambda_t (n_i - 1)(n_j + 1) && \text{for } \mathbf{n} \in \Gamma^n \text{ and } i, j \in E \\
\mathcal{Q}^n_{\text{drift}}(\mathbf{n}, \mathbf{n}) &= \tfrac{1}{2} \lambda_t && \text{for } \mathbf{n} \in \Gamma^n
\end{aligned}
$$

with all other entries being zero.

A.2 Mutation

Recall the term of the generator corresponding to mutation, $\mathcal{L}^{\text{mut}}$, defined in (2.3). Then the contribution of mutation to the time derivative of moment $M_{\mathbf{n}}(t)$ is,

$$
\begin{aligned}
&C(\mathbf{n}) \int_{\Delta_K} \mathcal{L}^{\text{mut}} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= C(\mathbf{n}) \int_{\Delta_K} \frac{1}{2} \sum_{i \in E} \sum_{j \in E} y_j \theta_{j,i} \frac{\partial}{\partial y_i} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= C(\mathbf{n}) \int_{\Delta_K} \frac{1}{2} \sum_{i \in E} \sum_{j \in E} \theta_{j,i} n_i \boldsymbol{y}^{\mathbf{n} - e_i + e_j} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= \frac{1}{2} \sum_{i \in E} \sum_{j \in E} \theta_{j,i} \frac{n_i C(\mathbf{n})}{C(\mathbf{n} - e_i + e_j)} M_{\mathbf{n} - e_i + e_j}(t) \\
&= \frac{1}{2} \sum_{i \in E} \sum_{j \in E} \theta_{j,i} (n_j + 1) M_{\mathbf{n} - e_i + e_j}(t)
\end{aligned}
$$

From this, we can define $\mathcal{Q}^n_{\text{mut}}$ as follows,

$$\begin{aligned}
\mathcal{Q}^n_{\text{mut}}(\mathbf{n}, \mathbf{n} - e_i + e_j) &= \tfrac{1}{2}\theta_{j,i}(n_j + 1) \text{ for } \mathbf{n} \in \Gamma^n \text{ and } i, j \in E \\
\mathcal{Q}^n_{\text{mut}}(\mathbf{n}, \mathbf{n}) &= \tfrac{1}{2}\lambda_t \qquad\qquad \text{for } \mathbf{n} \in \Gamma^n
\end{aligned}$$

with all other entries being zero.

A.3 Selection

Recall the term of the generator corresponding to mutation, $\mathcal{L}^{\text{mut}}$, defined in (2.4). Then the contribution of natural selection to the time derivative of moment $M_{\mathbf{n}}(t)$ is,

$$\begin{aligned}
&C(\mathbf{n}) \int_{\Delta_K} \mathcal{L}^{\text{sel}} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= C(\mathbf{n}) \int_{\Delta_K} \sum_{i \in E} \boldsymbol{y}^{e_i} \left[\sigma_i(\boldsymbol{y}) - \bar{\sigma}(\boldsymbol{y})\right] \frac{\partial}{\partial y_i} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= C(\mathbf{n}) \int_{\Delta_K} \sum_{i \in E} n_i \left\{ \boldsymbol{y}^{\mathbf{n}} \left[\sum_{j \in E} \sigma_{i,j} y_j - \sum_{j,k \in E} \sigma_{j,k} y_j y_k \right] \right\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= C(\mathbf{n}) \int_{\Delta_K} \sum_{i \in E} \left[\sum_{j \in E} n_i \sigma_{i,j} \boldsymbol{y}^{\mathbf{n}+e_j} - \sum_{j,k \in E} n_i \sigma_{j,k} \boldsymbol{y}^{\mathbf{n}+e_j+e_k} \right] p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\
&= C(\mathbf{n}) \sum_{i \in E} \left[\sum_{j \in E} \frac{n_i}{C(\mathbf{n} + e_j)} \sigma_{i,j} M_{\mathbf{n}+e_j}(t) - \sum_{j,k \in E} \frac{n_i}{C(\mathbf{n} + e_j + e_k)} \sigma_{j,k} M_{\mathbf{n}+e_j+e_k}(t) \right] \\
&= \sum_{i \in E} \sum_{j \in E} \left[\frac{n_i(n_j + 1)}{n + 1} \sigma_{i,j} M_{\mathbf{n}+e_j}(t) - \frac{n(n_i + 1 + \delta_{i,j})(n_j + 1)}{(n + 2)(n + 1)} \sigma_{i,j} M_{\mathbf{n}+e_i+e_j}(t) \right].
\end{aligned} \tag{A.1}$$

It then follows that $\mathcal{Q}^{n+1}_{\text{sel}}$ is defined as follows,

$$\begin{aligned}
\mathcal{Q}^{n+1}_{\text{sel, in}}(\mathbf{n}, \mathbf{n} + e_j) &= \tfrac{n}{n+1}(n_j + 1)\sigma_j(\boldsymbol{y}) \qquad \text{for } \mathbf{n} \in \Gamma^n \text{ and } j \in E \\
\mathcal{Q}^{n+2}_{\text{sel, out}}(\mathbf{n}, \mathbf{n} + e_i + e_j) &= -\tfrac{n(n_i+1+\delta_{i,j})(n_j+1)}{(n+2)(n+1)} \sigma_{i,j} \text{ for } \mathbf{n} \in \Gamma^n \text{ and } i, j \in E
\end{aligned}$$

with all other entries being zero. Since moments of order $n+1$ can be computed from the moments of order $n+2$, $\mathcal{Q}^{n+1}_{\text{sel}}$ can be re-written so that it maps moments of order $n+2$ to those of n. Namely, denoting the new matrix as $\mathcal{Q}^{n+2}_{\text{sel}}$, it can be expressed as $\mathcal{Q}^{n+2}_{\text{sel}} = \mathcal{Q}^{n+1}_{\text{sel, in}} \mathcal{D}_{n+2,n+1} + \mathcal{Q}^{n+2}_{\text{sel, out}}$, where $\mathcal{D}$ is the downsampling matrix, see Appendix C.2 for more details.

A.3.1 Genic Selection

Consider the case of genic selection at locus A. Namely, for all $i, j \in E$, define $\sigma_{i,j}$ as,

$$\sigma_{i,j} = \frac{\sigma i_1 + \sigma_{j_1}}{2}.$$

From the fourth line of (A.1) we have,

$$C(\mathbf{n}) \int_{\Delta_K} \mathcal{L}^{\text{sel}} \{\boldsymbol{y}^{\mathbf{n}}\} p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} = C(\mathbf{n}) \int_{\Delta_K} \sum_{i \in E} \left[\sum_{j \in E} n_i \sigma_{i,j} \boldsymbol{y}^{\mathbf{n}+e_j} - \sum_{j,k \in E} n_i \sigma_{j,k} \boldsymbol{y}^{\mathbf{n}+e_j+e_k} \right] p_t(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \tag{A.2}$$

The first term on the RHS of the above equation can be written as,

$$C(\mathbf{n})\int_{\Delta_K}\sum_{i\in E}\sum_{j\in E} n_i\sigma_{i,j}\boldsymbol{y}^{\mathbf{n}+e_j}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y} = C(\mathbf{n})\int_{\Delta_K}\sum_{i\in E}\sum_{j\in E} n_i\frac{\sigma_{i_1}+\sigma_{j_1}}{2}.\boldsymbol{y}^{\mathbf{n}+e_j}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$$
$$= C(\mathbf{n})\int_{\Delta_K}\left[\sum_{i\in E} n_i\frac{\sigma_{i_1}}{2}\boldsymbol{y}^{\mathbf{n}} + n\sum_{j\in E}\frac{\sigma_{j_1}}{2}\boldsymbol{y}^{\mathbf{n}+e_j}\right]p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$$

Then the second term on the RHS of (A.2) can be written as,

$$C(\mathbf{n})\int_{\Delta_K}\sum_{i\in E}\sum_{j,k\in E} n_i\sigma_{j,k}\boldsymbol{y}^{\mathbf{n}+e_j+e_k}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y} = C(\mathbf{n})\sum_{j,k\in E} n\left(\frac{\sigma_{j_1}+\sigma_{j_1}}{2}\right)\boldsymbol{y}^{\mathbf{n}+e_j+e_k}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$$
$$= C(\mathbf{n})\sum_{j,k\in E} n\sigma_{j_1}\boldsymbol{y}^{\mathbf{n}+e_j+e_k}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$$
$$= C(\mathbf{n})\sum_{j,k\in E} n\sigma_{j_1}\boldsymbol{y}^{\mathbf{n}+e_j}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$$

where the first equality follows from the definition of $\sigma_{j,k}$, the second follows from the symmetry of j and k, and the third follows from the fact that $\sum_{i\in E}\boldsymbol{y}^{e_i}=1$. Substituting this back into equation (A.2) yields,

$$C(\mathbf{n})\int_{\Delta_K}\mathcal{L}^{\text{sel}}\{\boldsymbol{y}^{\mathbf{n}}\}\,p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y} = C(\mathbf{n})\int_{\Delta_K}\left[\sum_{i\in E} n_i\frac{\sigma_{i_1}}{2}\boldsymbol{y}^{\mathbf{n}} - n\sum_{j\in E}\frac{\sigma_{j_1}}{2}\boldsymbol{y}^{\mathbf{n}+e_j}\right]p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$$
$$= \sum_{i\in E} n_i\frac{\sigma_{i_1}}{2}M_{\mathbf{n}}(t) - n\sum_{j\in E}\frac{n_i+1}{n+1}\frac{\sigma_{j_1}}{2}M_{\mathbf{n}+e_j}(t).$$

It then follows that $\mathcal{Q}^{n+1}_{\text{genic}}$ is defined as follows,

$$\begin{aligned}
\mathcal{Q}^{n}_{\text{genic, in}}(\mathbf{n},\mathbf{n}) &= \textstyle\sum_{i\in E} n_i\frac{\sigma_{i_1}}{2} && \text{for } \mathbf{n}\in\Gamma^n\\
\mathcal{Q}^{n+1}_{\text{genic, out}}(\mathbf{n},\mathbf{n}+e_j) &= -n\textstyle\sum_{j\in E}\frac{n_i+1}{n+1}\frac{\sigma_{j_1}}{2} && \text{for } \mathbf{n}\in\Gamma^n \text{ and } j\in E
\end{aligned}$$

with all other entries being zero. As with the case of general selection $\mathcal{Q}^{n}_{\text{genic, in}}$ can be written as $\mathcal{Q}^{n+1}_{\text{genic, in}}$ and combined with $\mathcal{Q}^{n+1}_{\text{genic, out}}$ to get $\mathcal{Q}^{n+1}_{\text{genic}}$.

A.4 Recombination

Recall the term of the generator corresponding to recombination, $\mathcal{L}^{\text{rec}}$, defined in (2.5). Then the contribution of recombination to the time derivative of moment $M_{\mathbf{n}}(t)$ is,

$$\begin{aligned}
&C(\mathbf{n})\int_{\Delta_K}\mathcal{L}^{\text{rec}}\{\boldsymbol{y}^{\mathbf{n}}\}\,p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}\\
&= C(\mathbf{n})\int_{\Delta_K}\frac{1}{2}\sum_{i=(i_A,i_B)\in E}\rho(y_{(i_A,\cdot)}y_{(\cdot,i_B)}-y_i)\frac{\partial}{\partial y_i}\{\boldsymbol{y}^{\mathbf{n}}\}\,p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}\\
&= C(\mathbf{n})\int_{\Delta_K}\frac{1}{2}\sum_{i=(i_A,i_B)\in E}\rho(y_{(i_A,\cdot)}y_{(\cdot,i_B)}-y_i)n_i\boldsymbol{y}^{\mathbf{n}-e_i}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}\\
&= C(\mathbf{n})\int_{\Delta_K}\frac{1}{2}\rho\sum_{i=(i_A,i_B)\in E}\sum_{j\in E^B}\sum_{k\in E^A}y_{(i_A,j)}y_{(k,i_B)}\boldsymbol{y}^{\mathbf{n}-e_i}p_t(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y} - \frac{\rho n}{2}M_{\mathbf{n}}(t)\\
&= \frac{1}{2}\rho\sum_{i=(i_A,i_B)\in E}\sum_{j\in E^B}\sum_{k\in E^A}\frac{n_iC(\mathbf{n})}{C(\mathbf{n}-e_i+e_{(i_A,j)}+e_{(k,i_B)})}M_{\mathbf{n}-e_i+e_{(i_A,j)}+e_{(k,i_B)}}(t) - \frac{\rho n}{2}M_{\mathbf{n}}(t)
\end{aligned}$$

The quantity $\mathcal{Q}^n_{\text{rec}}$ can then be defined as follows,

$$\begin{aligned}
\mathcal{Q}^{n+1}_{\text{rec, in}}(\mathbf{n}, \mathbf{n} - e_i + e_{(i_A,j)} + e_{(k,i_B)}) &= \tfrac{1}{2}\rho \tfrac{n_i C(\mathbf{n})}{C(\mathbf{n} - e_i + e_{(i_A,j)} + e_{(k,i_B)})} && \text{for } \mathbf{n} \in \Gamma^n \text{ and } i \in E, kin E^A, j \in E^B \\
\mathcal{Q}^{n+1}_{\text{rec, out}}(\mathbf{n}, \mathbf{n} + e_j) &= -\tfrac{\rho n}{2} && \text{for } \mathbf{n} \in \Gamma^n
\end{aligned}$$

with all other entries being zero. As in the case of selection $\mathcal{Q}^n_{\text{rec, out}}$ can be written as $\mathcal{Q}^{n+1}_{\text{rec, out}}$ and combined with $\mathcal{Q}^{n+1}_{\text{rec, in}}$ to get $\mathcal{Q}^{n+1}_{\text{rec}}$.

B Proof of Lemma 3.1

Proof of Lemma 3.1. Recall the probability measure

$$\mu := f(y)dy$$

on $[0,1]$ with density $f(\cdot)$ with respect to the Lebesgue measure dy, and the sequence of probability measures

$$\mu^n := \bar{f}_n(y)dy$$

with piece-wise constant densities

$$\bar{f}_n(y) := \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) (n+1) \mathbb{1}_{[\frac{i-0.5}{n}, \frac{i+0.5}{n})}(y)$$

with $y \in [0,1]$, where

$$b_i^n(z) = \binom{n}{i} z^i (1-z)^{n-i}$$

are the Bernstein polynomials. To show weak convergence of μ_n to μ, consider a bounded, Lipschitz continuous function $g : [0,1] \to \mathbb{R}$, that is, there exists a $K > 0$ such that

$$\big| g(x) - g(y) \big| \leq K |x - y|$$

holds for all $x, y \in [0,1]$. Moreover, let $g_{\max} \geq |g(\cdot)|$ be a bound on the function. Then, we have to show that

$$\int_0^1 g\mu^n \to \int_0^1 g\mu$$

for all such $g(\cdot)$. To this end, note that

$$\int_0^1 g\mu = \int_0^1 g(y) f(y) dy$$

holds. Furthermore, for the approximating sequence μ_n, we have

$$\begin{aligned}
\int_0^1 g\mu^n &= \int_0^1 g(y) \bar{f}_n(y) dy \\
&= \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) (n+1) \int_0^1 \mathbb{1}_{[i-\frac{1}{2n}, i+\frac{1}{2n})}(y) g(y) dy \\
&\leq \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) (n+1) \Big(\int_{i-\frac{1}{2n}}^{\frac{i}{n}} \Big(g\Big(\frac{i}{n}\Big) + K\Big(\frac{i}{n} - y\Big)\Big) dy + \int_{\frac{i}{n}}^{i+\frac{1}{2n}} \Big[g\big(\frac{i}{n}\big) + K\big(y - \frac{i}{n}\big)\Big] dy \Big) \\
&= \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) 2(n+1) \int_0^{\frac{1}{2n}} \Big(g\Big(\frac{i}{n}\Big) + Ky \Big) dy \\
&= \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) \frac{n+1}{n} \Big(g\Big(\frac{i}{n}\Big) + K \frac{1}{4n^2} \Big) \\
&= \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) g\Big(\frac{i}{n}\Big) \frac{n+1}{n} + K \frac{n+1}{4n^2},
\end{aligned}$$

where the inequality follows from the Lipschitz continuity of $g(\cdot)$ and setting $x = \frac{i}{n}$, and the last equality holds, because

$$\sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) = 1.$$

We can proceed along the same lines to obtain a lower bound which ultimately yields

$$\begin{aligned} \int_0^1 g\mu^n &= \frac{n+1}{n} \sum_{i=0}^{n} \Big(\int_0^1 b_i^n(z) f(z) dz \Big) g\Big(\frac{i}{n}\Big) + O\big(n^{-1}\big) \\ &= \frac{n+1}{n} \int_0^1 \sum_{i=0}^{n} b_i^n(z) g\Big(\frac{i}{n}\Big) f(z) dz + O\big(n^{-1}\big) \end{aligned} \tag{B.1}$$

The Weierstrass approximation theorem (e.g. Klenke, 2008, Ex. 5.15) applied to the Bernstein polynomials $b_i^n(z)$ can be used to show that

$$\sum_{i=0}^{n} b_i^n(z) g\Big(\frac{i}{n}\Big) \to g(z)$$

uniformly on $[0, 1]$ as $n \to \infty$. Furthermore, this expression can be bounded by

$$\sum_{i=0}^{n} b_i^n(z) g\Big(\frac{i}{n}\Big) \leq g_{\max} \sum_{i=0}^{n} b_i^n(z) = g_{\max}$$

for all x, and thus

$$\int_0^1 g\mu^n \to \int_0^1 g(y) f(y) dy \tag{B.2}$$

follows from the dominated convergence theorem (Klenke, 2008, Cor. 6.26) applied to the integral on the right hand side of equation (B.1). Convergence of the integrals in (B.2) for all bounded Lipschitz continuous functions $g(\cdot)$ is equivalent to the weak convergence of μ_n to μ due to the Portemanteau theorem (Klenke, 2008, Thm. 13.16).

□

C Implementation

Our method relies on numerically solving the ODE (2.7). This is ultimate accomplished using a modified Runge-Kutta-Fehlberg 4(5) method. The two main numerical challenges faced were to (1) quickly and accurately estimate moments of order $n + 1$ and $n + 2$ from those of order n, and then ensuring that the numerical solution to (2.7) remains stable while using the approximation. In this section we will detail how the method is numerically implemented.

C.1 ODE Solver Implementation

We solve the ODE (2.7) numerically using a modified Runge-Kutta-Fehlberg 4(5) method (e.g. Kincaid et al., 2009, Ch. 8.3). Since the state-space is a simplex, the implementation needs to take these constraints into account. In particular, at each step, all negative values are set to zero and values greater than one and set to one. Then, the state-vector is renormalized so that all the components sum to one.

C.2 Parimonious Moments

We now introduce a definition which is useful for the implementation of the ODE. First, consider the following identity:

$$\begin{aligned}\mathbf{M_n} &= C(\mathbf{n}) \int_{\Delta_K} \boldsymbol{y}^{\mathbf{n}} dp(\boldsymbol{y}) = C(\mathbf{n}) \int_{\Delta_K} \boldsymbol{y}^{\mathbf{n}} \left(\sum_{i=1}^{K} y_i \right) dp(\boldsymbol{y}) = C(\mathbf{n}) \int_{\Delta_K} \left(\sum_{i=1}^{K} \boldsymbol{y}^{\mathbf{n}+e_i} \right) dp(\boldsymbol{y}) \\ &= C(\mathbf{n}) \int_{\Delta_K} \left(\sum_{i=1}^{K} \boldsymbol{y}^{\mathbf{n}+e_i} \right) dp(\boldsymbol{y}) = \sum_{i=1}^{K} \frac{C(\mathbf{n})}{C(\mathbf{n}+e_i)} \mathbf{M}_{\mathbf{n}+e_i}.\end{aligned} \tag{C.1}$$

This shows that every moments of order n can be explicity computed as a linear combination of moments of order $n+1$. One can apply this identity multiple time to represent $\mathbf{M}_n$ in terms of $\mathbf{M}_{n+m}$ for any $m > 0$. If the two vectors M_n and M_{n+m} satisfy these identities for all components of M_n we say that they are "parsimonious". Furthermore, one can use (C.1) to define a sparse matrix $\mathcal{D}_{n+m,n}$ which maps $\mathbf{M}_{n+m}$ to $\mathbf{M}_n$.

C.3 Evaluation of Derivative

The time-derivative of the state variable $\mathbf{M}_{n+1}(t)$ is composed of four components (genetic drift, mutation, recombination, selection). For convenience, we split recombination and selection into *in* (positive term) and *out* (negative term). Each of the six terms can then be computed by multiplying the moment vector with a sparse matrix:

1. Net flux due to genetic drift.
2. Net flux due to mutation.
3. Flux in due to selection.
4. Flux out due to recombination.
5. Flux out due to selection.
6. Flux in due to recombination.

Terms 1-4 can be expressed as linear combinations of the order n moments while 5 and 6 need moments of order $n+1$ (or $n+2$ in the case of general diploid selection) and rely on the moment approximation. However, we noticed that in most practical scenarios the ODE is more accurate and stable when using the moments of order n which are parsimonious with those moments of order $n+1$. When evaluating the derivative, we thus employ the vector of moments which are parsimonious with the estimated M_{n+1} rather than the original M_n for 3 and 4. Furthermore since $\mathcal{D}_{n+1,n}$ is sparse this can be implemented efficiently.

C.4 Initial Distribution

In this section we outline how the initial moments for the ODEs are computed. We consider three main types of initial conditions. In the first, the population starts from a fixed haplotypes frequency $x = [x_{AB}, x_{Ab}, x_{aB}, x_{ab}]$. This case is straightforward as the moments can easily be obtained using a multinomial distribution. In the second case, we assume that the neutral site, B, starts from a stationary (beta) distribution while the selected allele A starts from a fixed frequency. Finally, we can also allow for both sites to start from stationarity, modeling selection that starts from standing variation.

C.4.1 One-Locus Stationary Distribution

Suppose we model the case in which a selected allele is introduced at frequency x_A at time 0 while the neutral site is at stationarity. Under this assumption, the distribution of alleles at the neutral locus follows a Beta(θ, θ) distribution. Furthermore, the distribution of the frequency of the allele A is a point mass at frequency x_A. A natural way to implement this seems to just consider the probability measure $\delta_{x_A} \times p_B$ where p_B is the probability measure corresponding to the Beta(θ, θ) distribution. However, this measure is not correct, especially if one is interested in modeling the impact of selection on nearby neutral sites. The problem is that it is modeling a situation in which the new allele A can land on both the B and b background simultaneously. To see this note that if we were to take $\delta_{x_A} \times p_B(x)dx$ with $x_A = 1/2N$ (corresponding to a single selected allele) and computed the sampling probabilities for a sample of size 2, the probability of sampling one AB haplotype and one Ab haplotype would be non-zero. Instead the initial distribution should be

$$\begin{aligned} &I(1 - x_{AB} \geq x_B \geq x_{AB})(x_B \delta_{(x_A,0)}(x_{AB}, x_{Ab}) + (1-x)\delta_{(0,x_A)}(x_{AB}, x_{Ab})) \times p_B(x_B)dx_B \\ &\quad + I(1 - x_A < x_B)\delta_{(x_A,0)}(x_{AB}, x_{Ab} \times p_B(x_B)dx_B \\ &\quad + I(x_B < x_A)\delta_{(0,x_A)}(x_{AB}, x_{Ab}) \times p_B(x_B)dx_B, \end{aligned} \tag{C.2}$$

where $I(\cdot)$ is the indicator function that is 1 if the condition is true, and zero otherwise. This reflects that if one were to randomly introduce a mutation A, it has probability x of landing on the same chromosome as a B and $1 - x$ of arising on the b background. Furthermore the indicator functions reflect the fact that the frequency of B cannot be smaller than the frequency of AB and similarly for b and Ab. In particular, the first line represents the case in which the frequency of B and b is above that of A and thus both AB and Ab are possible. The second (resp. third) line correspond to the case in which b (resp. B) is too small for Ab (resp. Ab) to exist. Note that this measure is discontinuous as x_B increases across the threshold x_A. This is due to the fact that the beta distribution is assuming a population of infinite size, whereas introducing an allele in a finite number of individuals is assuming a finite population. Indeed the probability measure (C.2) is an approximation of both, however it approximates the scenario sufficiently well for small x_A. Given a moment configuration, the moment can be computed by integrating the appropriate multinomial probability-mass-function against (C.2). This can be split of up into three integrals the first corresponding to the case of $x < x_A$.

$$\begin{aligned} M_n(t) = &\int_0^{x_A} 0^{n_{AB}} (x_A)^{n_{Ab}} x^{n_{aB}} (1 - x - x_A)^{n_{ab}} (1-x)p(x)dx \\ &+ \frac{1}{2}\int_{x_A}^{1-x_A} \Big[(x_A)^{n_{AB}} 0^{n_{Ab}} (x - x_A)^{n_{aB}} (1-x)^{n_{ab}} x \\ &\qquad + 0^{n_{ab}} (x_A)^{n_{Ab}} x^{n_{aB}} (1 - x - x_A)^{n_{ab}} (1-x) \Big] p(x)dx \\ &+ \int_{1-x_A}^{1} (x_A)^{n_{AB}} 0^{n_{Ab}} (x - x_A)^{n_{aB}} (1-x)^{n_{ab}} x p(x)dx \end{aligned}$$

Considering the last integral and assuming $n_{Ab} = 0$

$$\begin{aligned} &\int_{1-x_A}^{1} (x_A)^{n_{AB}} (x - x_A)^{n_{aB}} (1-x)^{n_{ab}} x p(x)dx \\ &= \int_{1-x_A}^{1} (x_A)^{n_{AB}} (x - x_A)^{n_{aB}} (1-x)^{n_{ab}} x \frac{x^{\beta_1 - 1}(1-x)^{\beta_2 - 1}}{B(\beta_1, \beta_2)} dx \\ &= \int_{1-x_A}^{1} (x_A)^{n_{AB}} \sum_{i=1}^{n_{aB}} \left[x^i \binom{n_{aB}}{i} (-x_A)^{n_{aB}-i} \right] \frac{x^{\beta_1 + i}(1-x)^{\beta_2 + n_{ab} - 1}}{B(\beta_1, \beta_2)} dx \\ &= \sum_{i=1}^{n_{aB}} \binom{n_{aB}}{i} (-x_A)^{n_{aB}-i} (x_A)^{n_{AB}} \frac{B(\beta_1 + i + 1, \beta_2 + n_{ab}) - B(1 - x_A; \beta_1 + i + 1, \beta_2 + n_{ab})}{B(\beta_1, \beta_2)} \end{aligned}$$

where $B(\cdot,\cdot)$ and $B(\cdot;\cdot,\cdot)$ are the Beta function and incomplete Beta function, respectively. By a similar argument the first term equals:

$$\int_0^{x_A} (x_A)^{n_{Ab}} x^{n_{aB}} (1-x-x_A)^{n_{ab}} (1-x)p(x)dx$$
$$= \sum_{i=1}^{n_{ab}} \binom{n_{ab}}{i} (-x_A)^{n_{ab}-i} (x_A)^{n_{Ab}} \frac{B(x_A;\beta_1+n_{aB},\beta_2+i+1)}{B(\beta_1,\beta_2)}$$

Finally the middle term equals:

$$\frac{1}{2}\int_{x_A}^{1-x_A} \left[(x_A)^{n_{AB}} 0^{n_{Ab}} (x-x_A)^{n_{aB}} (1-x)^{n_{ab}} x + 0^{n_{ab}} (x_A)^{n_{Ab}} x^{n_{aB}} (1-x-x_A)^{n_{ab}} (1-x)\right] p(x)dx$$
$$= \frac{0^{n_{Ab}}}{2} \sum_{i=1}^{n_{aB}} \binom{n_{aB}}{i} (-x_A)^{n_{aB}-i} (x_A)^{n_{AB}} \frac{B(1-x_A;\beta_1+i+1,\beta_2+n_{ab}) - B(x_A;\beta_1+i+1,\beta_2+n_{ab})}{B(\beta_1,\beta_2)}$$
$$+ \frac{0^{n_{AB}}}{2} \sum_{i=1}^{n_{ab}} \binom{n_{ab}}{i} (-x_A)^{n_{ab}-i} (x_A)^{n_{Ab}} \frac{B(1-x_A;\beta_1+n_{aB},\beta_2+i+1) - B(x_A;\beta_1+n_{aB},\beta_2+i+1)}{B(\beta_1,\beta_2)}$$

C.4.2 Two-Locus Stationary Distribution

In addition to modeling the setting in which a selected allele arises on a background which is at stationarity, one may be interested in modeling a situation in which an existing neutral allele gains a selective advantage/disadvantage and selection starts from standing variation. In such a setting, the ODEs need to be initialized from the two-locus stationary distribution. To do this, one can take a similar approach as in `LDpop` (Kamm et al., 2016). Namely, one can formulate a two-locus Moran model which accounts for genetic drift, mutation, and recombination and compute its transition matrix. This transition matrix can then be used to compute the stationary distribution of the moments by either solving the appropriate set of linear equations or via power iteration, which can be more efficient for sparse matrices. For an exact result, one would use the augmented Moran model of Kamm et al. (2016) which includes partially specified haplotypes, while for larger sample sizes ($n \geq 20$) the normal two-locus Moran model of Ethier and Kurtz (1993) serves as a good approximation.

D Additional Simulation Results

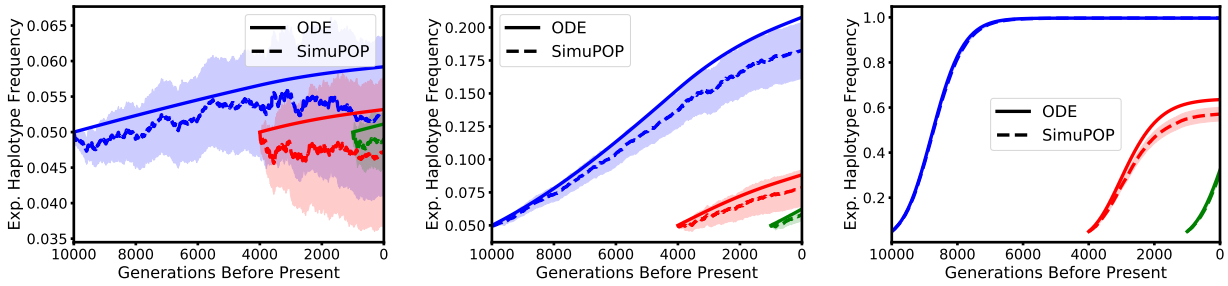
In this section we display some additional results for some parameter configurations not shown in the main text.

D.1 Fixed Initial Frequencies

In Figure D.1, we present the expected frequency of haplotype AB with the same parameters as Figure 4 but $\rho = 0.0004$. We can see that since the two loci are tightly linked the trajectories closely resemble expected trajectory of the A allele in 3.

D.2 Neutral locus at stationarity

In this Section we present analogous results to those displayed in Section 4.2 with the selected allele A is initialized at frequency 0.03 and 0.01 instead of 0.05. Figures D.2 and D.3 display the heterozygosity over time (analogous to Figure 7) with A initialized from 0.03 and 0.05, respectively. Similarly, Figures D.4 and D.5 display D^2 over time (analogous to Figure 8) where A is initialized at frequency 0.01 and 0.03, respectively.



(a) Selection coefficient $\sigma = 1$.

(b) Selection coefficient $\sigma = 10$.

(c) Selection coefficient $\sigma = 100$.

Figure D.1: Expected frequency of the AB haplotype, but for strong linkage ($\rho = 0.0004$). Again, blue, red, and green correspond to the dynamics starting 10,000 generations, 4,000 generations (beginning of the bottleneck), and 1,000 generations before present (beginning of exponential growth), respectively. The dotted and solid lines correspond to the simulation average and ODE solution, respectively, with the shaded region indicating the 95% confidence interval.

References

Bergman, J., Schrempf, D., Kosiol, C., and Vogl, C. (2018). Inference in population genetics using forward and backward, discrete and continuous time processes. *Journal of Theoretical Biology*, **439,** 166 – 180.

Berisa, T. and Pickrell, J. K. (2015). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, **32,**(2) 283–285.

Boyd, S. and Vandenberghe, L. (2018). *Introduction to Applied Linear Algebra: Vectors, Mtrices, and Least Squares*. Cambridge university press.

Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., Adams, M. D., Schmidt, S., Sninsky, J. J., Sunyaev, S. R., White, T. J., Nielsen, R., Clark, A. G., and Bustamante, C. D. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genetics*, **4,**(5) 1–13.

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., of the Psychiatric Genomics Consortium, S. W. G., Patterson, N., Daly, M. J., Price, A. L., and Neale, B. M. (2015). Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, **47,**(3) 291–295.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2,**(4) e64.

Durrett, R. (2008). *Probability Models for DNA Sequence Evolution*. Springer.

Ethier, S. N. and Kurtz, T. G. (1993). Fleming–viot processes in population genetics. *SIAM Journal on Control and Optimization*, **31,**(2) 345–386.

Ethier, S. N. and Kurtz, T. G. (2009). *Markov Processes: Characterization and Convergence*. John Wiley & Sons.

Evans, S. N., Shvets, Y., and Slatkin, M. (2007). Non-equilibrium theory of the allele frequency spectrum. *Theoretical Population Biology*, **71,**(1) 109–119.

Ewens, W. J. (2010). *Mathematical Population Genetics 1: Theoretical Introduction (Interdisciplinary Applied Mathematics)*. Springer.

Fay, J. C. and Wu, C. I. (2000). Hitchhiking under positive darwinian selection. *Genetics*, **155,**(3) 1405–1413.

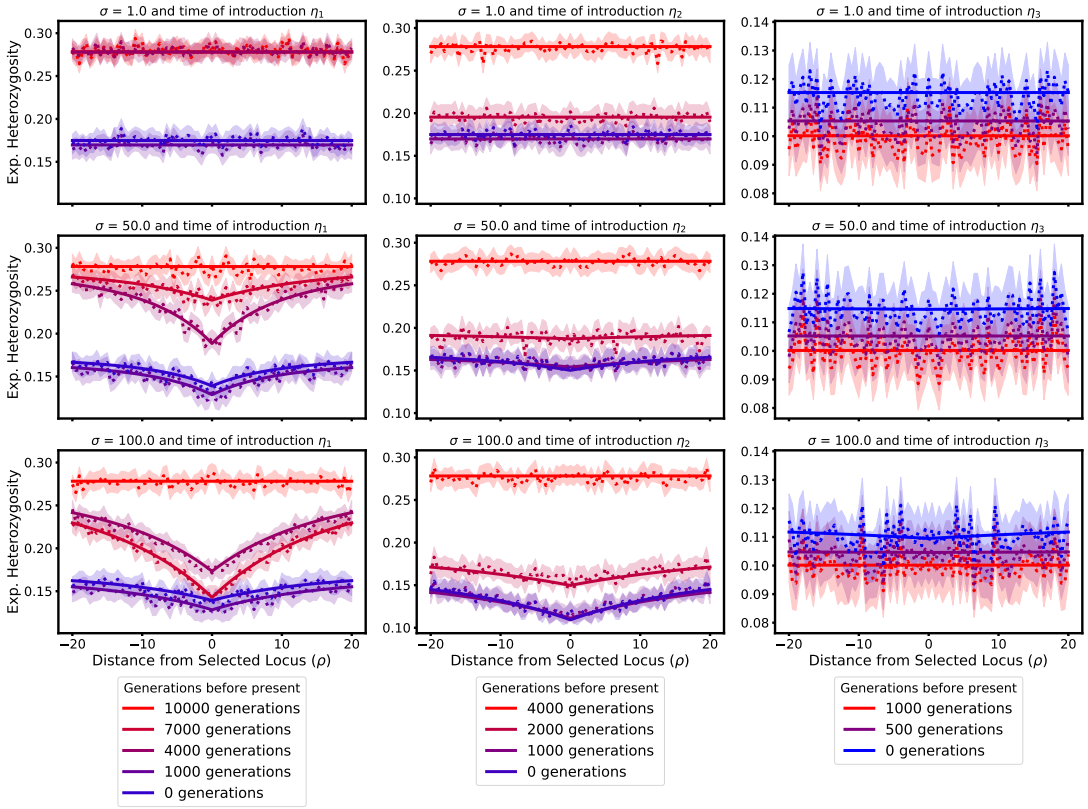


Figure D.2: Heterozygosity across a 100 kbp region with different selection coefficients and starting times. The beneficial allele is introduced at frequency $x_A = 0.03$. The dotted and solid lines represent the simulation and ODE results, respectively, with the shaded region representing a 95% confidence interval.

Gravel, S. and National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project. (2014). Predicting discovery rates of genomic features. *Genetics*, **197,**(2) 601–610.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5,** e1000695.

Haller, B. C. and Messer, P. W. (2019). Slim 3: Forward genetic simulations beyond the wright–fisher model. *Molecular Biology and Evolution*, **36,**(3) 632–637.

Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., and Ralph, P. L. (2019). Tree-sequence recording in slim opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, **19,** (2) 552–566.

He, Z., Beaumont, M. A., and Yu, F. (2020). Numerical simulation of the two-locus wright-fisher stochastic differential equation with application to approximating transition probability densities. *bioRxiv*. `https://doi.org/10.1101/2020.07.21.213769`.

He, Z., Dai, X., Beaumont, M., and Yu, F. (2020). Detecting and quantifying natural selection at two linked loci from time series data of allele frequencies with forward-in-time simulations. *Genetics*, **216,**(2) 521–541.

Hejase, H. A., Dukler, N., and Siepel, A. (2020). From summary statistics to gene trees: Methods for inferring positive selection. *Trends in Genetics*, **36,**(4) 243 – 258.

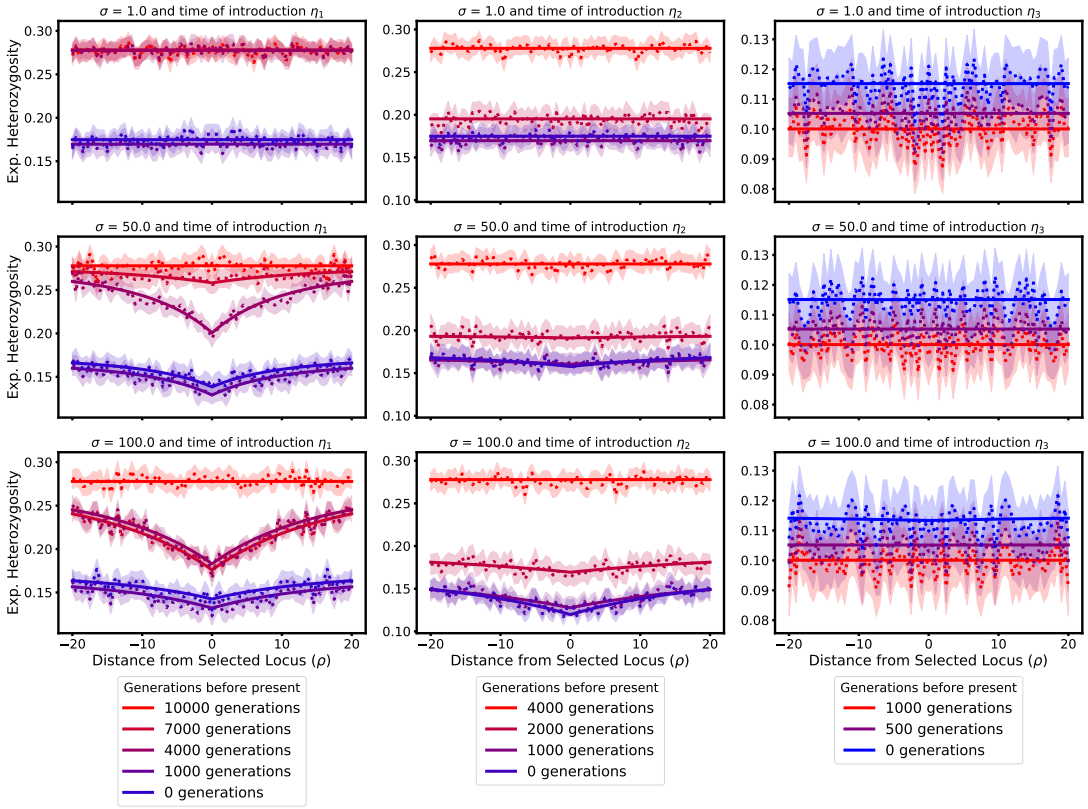


Figure D.3: Heterozygosity across a 100 kbp region with different selection coefficients and starting times. The beneficial allele is introduced at frequency $x_A = 0.01$. The dotted and solid lines represent the simulation and ODE results, respectively, with the shaded region representing a 95% confidence interval.

Hill, W. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, **38,**(6) 226–231.

Jenkins, P. A. and Song, Y. S. (2012). Padé approximants and exact two-locus sampling distributions. *Annals of Applied Probability*, **22,**(2) 576–607.

Jouganous, J., Long, W., Ragsdale, A. P., and Gravel, S. (2017). Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, **206,**(3) 1549–1567.

Kamm, J., Terhorst, J., Durbin, R., and Song, Y. S. (2020). Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, **115,**(531) 1472–1487.

Kamm, J. A., Spence, J. P., Chan, J., and Song, Y. S. (2016). Two-locus likelihoods under variable population size and fine-scale recombination rate estimation. *Genetics*, **203,**(3) 1381–1399.

Kaplan, N. L., Hudson, R. R., and Langley, C. H. (1989). The "hitchhiking effect" revisited. *Genetics*, **123,** (4) 887–899.

Karlin, S. and Taylor, H. E. (1981). *A Second Course in Stochastic Processes*. Elsevier.

Kim, Y. and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160,**(2) 765–777.

Kimura, M. (1955). Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology*, **20,** 33–53.

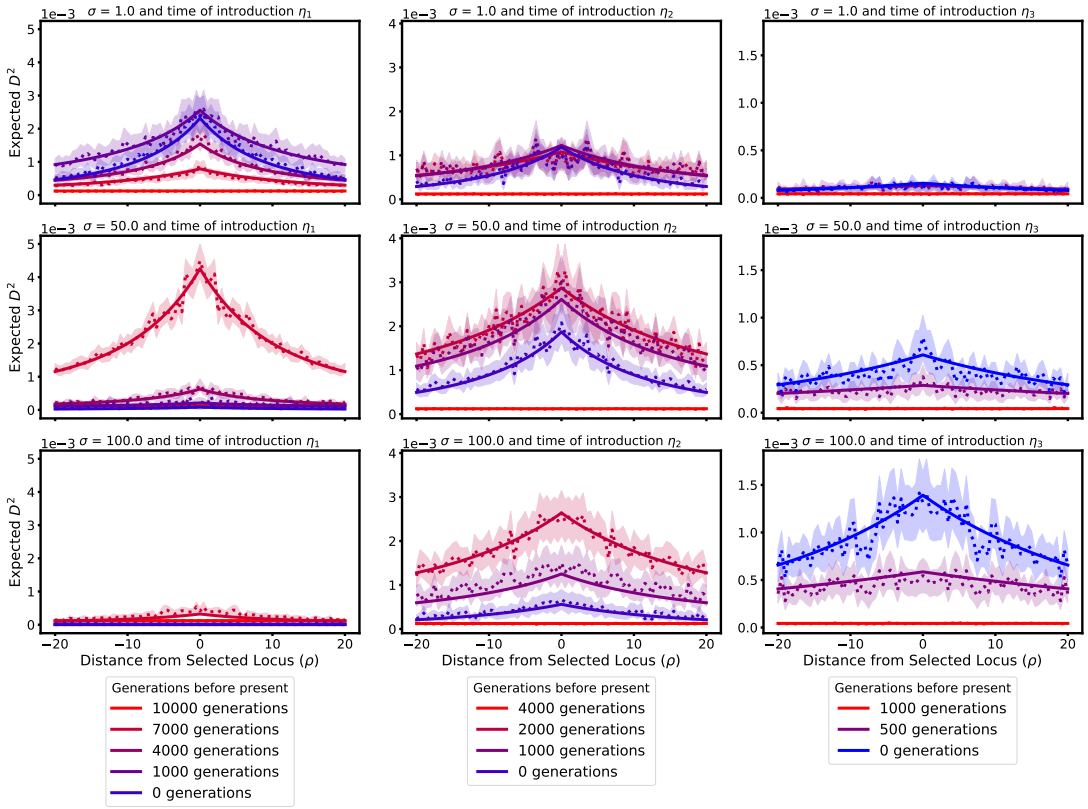


Figure D.4: D_2 across a 100 kbp region with different selection coefficients and starting times. The beneficial allele is introduced at frequency $x_A = 0.03$. The dotted and solid lines represent the simulation and ODE results, respectively, with the shaded region representing a 95% confidence interval.

Kimura, M. (1955). Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences*, **41,**(3) 144–150.

Kincaid, D., Kincaid, D. R., and Cheney, E. W. (2009). *Numerical Analysis: Mathematics of Scientific Computing*, volume 2. American Mathematical Society.

Klenke, A. (2008). *Probability Theory : A Comprehensive Course*. Springer, London.

Maynard Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, **23,**(1) 23–35.

McVean, G., Awadalla, P., and Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160,**(3) 1231–1241.

Øksendal, B. (2003). Stochastic differential equations. In *Stochastic differential equations* 65–84. Springer.

Pavlidis, P. and Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki*, **24,**(1) 1–17.

Peng, B. and Amos, C. I. (2008). Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics*, **24,**(11) 1408–1409.

Peng, B. and Kimmel, M. (2005). simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21,**(18) 3686–3687.

Pritchard, J. K., Pickrell, J. K., and Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20,**(4) R208–R215.

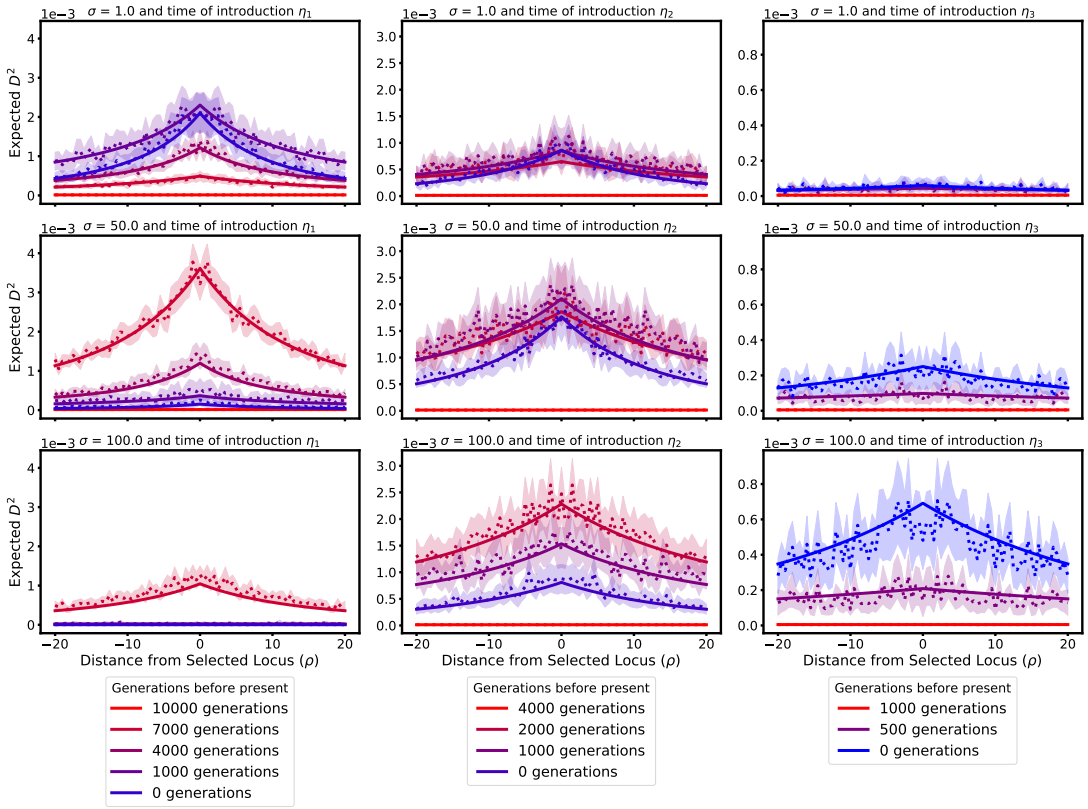


Figure D.5: D_2 across a 100 kbp region with different selection coefficients and starting times. The beneficial allele is introduced at frequency $x_A = 0.01$. The dotted and solid lines represent the simulation and ODE results, respectively, with the shaded region representing a 95% confidence interval.

Ragsdale, A. P. and Gravel, S. (2019). Models of archaic admixture and recent history from two-locus statistics. *PLoS Genetics*, **15,**(6) e1008204.

Ragsdale, A. P. and Gutenkunst, R. N. (2017). Inferring demographic history using two-locus statistics. *Genetics*, **206,**(2) 1037–1048.

Schraiber, J. G., Evans, S. N., and Slatkin, M. (2016). Bayesian inference of natural selection from allele frequency time series. *Genetics*, **203,**(1) 493–511.

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,** 1117–1129.

Spence, J. P. and Song, Y. S. (2019). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances*, **5,**(10).

Steinrücken, M., Wang, Y. X. R., and Song, Y. S. (2013). An explicit transition density expansion for a multi-allelic wright-fisher diffusion with general diploid selection. *Theoretical Population Biology*, **83,** 1–14.

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.

Williamson, S. H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R., and Bustamante, C. D. (2005). Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences*, **102,**(22) 7882–7887.

Zabad, S., Ragsdale, A. P., Sun, R., Li, Y., and Gravel, S. (2020). Assumptions about frequency-dependent architectures of complex traits bias measures of functional enrichment. *bioRxiv*. `https://doi.org/10.1101/2020.10.23.352427`.

Zhao, L., Yue, X., and Waxman, D. (2013). Complete numerical solution of the diffusion equation of random genetic drift. *Genetics*, **194,**(4) 973–985.

Živković, D. and Stephan, W. (2011). Analytical results on the neutral non-equilibrium allele frequency spectrum based on diffusion theory. *Theoretical Population Biology*, **79,**(4) 184–191.