Statistical Inference in the Wright–Fisher Model Using Allele Frequency Data

PAULA TATARU†, MARIA SIMONSEN†, THOMAS BATAILLON, AND ASGER HOBOLTH*

Bioinformatics Research Centre, Aarhus University, Aarhus, 8000, Denmark.
*Correspondence to be sent to: Bioinformatics Research Centre, Aarhus University, C.F. Møllers Allé 8, Aarhus C 8000, Denmark. E-mail: asger@birc.au.dk
†These authors contributed equally to this work

Abstract.—The Wright–Fisher model provides an elegant mathematical framework for understanding allele frequency data. In particular, the model can be used to infer the demographic history of species and identify loci under selection. A crucial quantity for inference under the Wright–Fisher model is the distribution of allele frequencies (DAF). Despite the apparent simplicity of the model, the calculation of the DAF is challenging. We review and discuss strategies for approximating the DAF, and how these are used in methods that perform inference from allele frequency data. Various evolutionary forces can be incorporated in the Wright–Fisher model, and we consider these in turn. We begin our review with the basic bi-allelic Wright–Fisher model where random genetic drift is the only evolutionary force. We then consider mutation, migration, and selection. In particular, we compare diffusion-based and moment-based methods in terms of accuracy, computational efficiency, and analytical tractability. We conclude with a brief overview of the multi-allelic process with a general mutation model. [Allele frequency, diffusion, inference, moments, selection, Wright–Fisher.]

A central goal of population genetics is to infer the past history of populations and describe the evolutionary forces that have shaped their genetic variation. The Wright–Fisher model (Fisher 1930; Wright 1931) explicitly accounts for the effects of various evolutionary forces—random genetic drift, mutation, selection—on allele frequencies over time. This model can also accommodate the effect of demographic forces such as variation in population size through time and/or migration connecting populations. Information about these evolutionary and demographic forces can, in principle, be retrieved from allele frequency data. The questions that researchers can answer and the types of inference they can make depend on the type of genetic data available, which can be broadly divided into two categories.

One type of data is a time series of allele frequencies from a single population (Fig. 1a). Here, the task is often to quantify the amount of drift that has influenced the changes in allele frequencies over time. This is done by estimating the size of the ideal Wright–Fisher population that best accounts for the patterns of genetic drift observed in the data, or, in other words, to estimate the effective population size. Furthermore, an important goal could be to identify those loci that have been under positive selection over the time interval considered.

The second type of data consists of allele frequencies from multiple populations, typically collected in the present (Fig. 1b). In this situation, the task is often to infer divergence times, population sizes, mutation rates, and, if applicable, migration rates between populations. Additionally, there is also considerable interest in evaluating the role of selection in shaping the observed data. Typical questions are: Do allele frequencies in regions of interest harbor footprints of selection? What is the overall importance of purifying selection on a specific set of sites (e.g., non-coding regions of functional interest or non-synonymous positions in gene coding regions)? We emphasize that this second type of data is very similar to the type of data analyzed in phylogenetics. In both instances, information is gained as new mutations arise at the nucleotide level and the fate of these mutations is influenced by the different evolutionary and demographic forces of interest. The difference between phylogenetics and population genetics essentially resides in the time scales that are modeled. Phylogenetics is often concerned with long time scales, and the data contain one sample per species. Differences among the sequences are most often substitutions. Population genetics typically considers data where several samples are available within a species, and many differences are detected due to mutations that are still segregating (polymorphic). Interestingly, these two time scales tend to merge when considering data sets containing sequences of individuals that comprise recently diverged species, as both types of differences—mutations that are still polymorphic and mutations that have been fixed as substitutions—have to be modeled jointly.

To infer the evolutionary history of a population, model-based approaches in population genetics have to rely on an explicit model for the evolution of populations. The Wright–Fisher model (Fisher 1930; Wright 1931) occupies a central position in this endeavour. It provides an elegant mathematical framework for modeling allele frequency data. The dynamics of the model are well understood (Kimura 1955a, 1955b, 1964; Crow and Kimura 1956; Crow and Kimura 1970; Ewens 1972; Crow 1987; Ewens 2004) but inference under the Wright–Fisher model is complicated due to the lack of a simple closed-form analytical expression for the distribution of allele frequencies (DAF). Common to all inference methods is the need to determine the DAF, either at equilibrium or over specified time intervals.

Here, we focus on how the DAF is influenced by demographic and evolutionary forces and concentrate on both classical and more recent attempts to calculate...
the DAF that enable accurate yet tractable population genetics inference. We begin our review with the basic bi-allelic Wright–Fisher model by considering, in turn, the forces of pure genetic drift, mutation, migration, and selection. For each of these forces, we provide expressions for the mean and variance of the DAF, and discuss and compare the approaches used to obtain the DAF. We also review implementations of the inference methods (Table 1).

Although the bi-allelic Wright–Fisher model captures a major part of data types, in particular single-nucleotide polymorphisms (SNPs), some loci are intrinsically multi-allelic. We therefore also briefly discuss recent progress on the multi-allelic Wright–Fisher model. We investigate if one of the widely used approximations for the multi-allelic DAF can capture adequately the first two moments of the DAF, and point to limitations of the approximation.

Next to the Wright–Fisher model, the coalescent (Kingman 2000, 1982a, 1982b, 1982c) and Moran (Moran 1958) models occupy an important role in the field. The coalescent process is dual to the Wright–Fisher model: although the Wright–Fisher model describes the evolution of a population forward in time in discrete non-overlapping generations, the coalescent process is built backwards in time, and arises as an approximation to the Wright–Fisher model when the population size is large. Unlike the coalescent, the Moran model is a forward-in-time process, and it is often regarded as an equivalent to the Wright–Fisher model (but see Bhaskar and Song 2009). Both the coalescent and Moran models have been analyzed extensively and their dynamics are in several cases more amenable to mathematical analysis (Donnelly 1984; Ewens 2004; Hobolth et al. 2007; Muirhead and Wakeley 2009; Li and Durbin 2011; Paul et al. 2011; Vogl and Clemente 2012). However, the Moran model is hardly ever used for inference (but see, e.g., De Maio et al. 2013, 2015), whereas the coalescent is typically restricted to a handful of individuals (Hobolth et al. 2007; Li and Durbin 2011; Paul et al. 2011; Mailund et al. 2012; Sheehan et al. 2013; Schifflers and Durbin 2014; Rasmussen et al. 2014) and does not use allele frequency data (but see, e.g., Liu and Fu 2015). Therefore, we do not include the coalescent and Moran models in this review, and refer the reader instead to Fu and Li (1999); Durrett (2008); Kuhner (2009); Liu et al. (2009); Wakeley (2009); Nielsen and Slatkin (2013); Edwards et al. (2016).

**Bi-allelic Wright–Fisher model**

The Wright–Fisher model assumes a randomly mating population of finite size reproducing in discrete non-overlapping generations, by allowing the individuals in generation \( r + 1 \) to choose parents at random from the previous generation \( r \). The model describes the stochastic behavior through time of the frequency of an allele at a locus. This frequency is influenced by a series of evolutionary forces that, as discussed below, change the probability of choosing a parent. Here, we consider a diploid population of size \( 2N \) which contains only two alleles, denoted \( A \) and \( a \). Below we review methods used to obtain the DAF of allele \( A \) after a certain amount of generations.

**Pure Drift**

The Wright–Fisher model, in its simplest form, only considers random genetic drift (Fig. 2), where the stochastic fluctuations in the allele frequency are purely determined by the random mating of the population. This assumption is appropriate for the analysis of loci that have small mutation rates and the analysis of recently diverged populations, leaving little time for mutation to create new alleles, and where we expect an overall negligible effect of selection.

**Dynamics and moments.**—Let \( z(r) \) be the number of \( A \) alleles in generation \( r \) and \( x(r) = z(r)/(2N) \) be the
The mean and variance after \( r \) generations can be obtained by iterating the two expressions above or from alternative derivations (Wright 1942; Crow 1954; Crow and Kimura 1970). The result is

\[
E[x(r) | x(0)] = x(0),
\]

\[
\text{Var}(x(r) | x(0)) = x(0)[1 - x(0)] \left(1 - \left(1 - \frac{1}{2N}\right)^r\right).
\]

For large \( N \), we can approximate the variance by

\[
\text{Var}(x(t) | x(0)) \approx x(0)[1 - x(0)] \left(1 - e^{-\frac{t}{2N}}\right).
\]

where \( t = r/(2N) \). Note that this implies that \( N \) can be estimated by equation (4) only if \( r \) is known, otherwise only the ratio \( t = r/(2N) \) can be estimated.

**Markov chain theory.**—Because the allele frequency at generation \( r + 1 \) only depends on generation \( r \), the Wright–Fisher model is a discrete-time finite-space Markov chain. Using this property, the DAF can be obtained from classical Markov chain theory (Karlin and Taylor 1975), where the transition probabilities are given by equation (1) (Williamson and Slatkin 1999). However, this procedure quickly becomes computationally infeasible, as the transition probability matrix has a size of \( (2N+1) \times (2N+1) \). By recognizing that most of the probability mass from equation (1) is centered around \( z(r) \), the computational demand can be reduced by evaluating, storing and using only the transition probabilities that are large enough to contribute significantly to the DAF (Wang 2001; Freeman et al. 2003). Under the assumption of large \( N \), diffusion theory (see below) shows that the population size acts as a scaling factor (Feller 1951; Wakeley 2005).
and therefore one could calculate the DAF using a smaller $N$. This approach was used by De Maio et al. (2013; 2015), though they relied on the Moran model rather than the Wright–Fisher. Alternatively, if $N$ is large enough such that the allele frequencies can be treated as continuous, the Markov chain can be built over discretized allele frequencies, and thus the computational burden is controlled by the number of bins. The original discrete binomial sampling probability from equation (1) is then replaced by the continuous normal or beta distributions (Mathieson and McVean 2013; Gompert 2015).

**Diffusion approximation.**—One way to calculate the DAF is to take advantage of the diffusion approximation to the Wright–Fisher model, which is appropriate when the population size $N$ is large, such that both allele frequencies and time can be treated as continuous. Diffusion theory uses two fundamental equations, the Kolmogorov forward and backward equations (Kolmogorov 1931). The forward equation was first used by Wright (1945) to calculate the rate of decay and stationary DAF, whereas Kimura (1957) used the backward equation first to study the problem of fixation. Let us define a new time scale by $b! = 1/(2N)$ such that one time unit corresponds to $2N$ generations. Then, we have

$$2Nz(t + b!t) | x(t) \sim \text{Bin}(2N, x(t)),$$

from which we can approximate

$$z(t + b!) | x(t) \approx N(x(t), x(t)(1 - x(t))b!).$$

Equation (5) corresponds to the time-homogeneous stochastic differential equation

$$\frac{dx(t)}{dt} = a(x)dt + \sqrt{b(x)}dW(t),$$

where $(W(t); t \geq 0)$ is a standard Brownian motion, and $a(x)$ and $b(x)$ are the infinitesimal mean and variance, respectively. For the Wright–Fisher model, $b(x) = x(1 - x)$, whereas $a(x)$ has different forms depending on the evolutionary forces. Under pure drift, $a(x) = 0$, as is evident from equation (5).

The DAF $f(x; t)$ at time $t$ is now determined by the forward Kolmogorov (or Fokker–Planck or diffusion) equation (Kolmogorov 1931; Crow and Kimura 1970; Ewens 2004)

$$\frac{df(x; t)}{dt} = -\frac{d}{dx}\left[f(x; t) a(x)\right] + \frac{1}{2} \frac{d^2}{dx^2}\{f(x; t) [1 - x]f(x; t)\},$$

with boundary condition $x = x(0)$ for $t = 0$. This equation can be solved using different approaches (Table 1). Kimura first described how the DAF can be calculated under pure drift (Kimura 1955a) using the spectral decomposition of equation (7), which results in an infinite sum of scaled Gegenbauer polynomials. In practice, the infinite sum needs to be truncated and the optimal truncation level depends on the convergence properties. This controls the accuracy, but also the computational performance. The diffusion equation can also be solved using purely numerical methods. Chang and Cooper (1970) developed a finite-difference scheme to numerically solve any diffusion equation, whereas Zhao et al. (2013) proposed a finite-volume scheme to solve the Wright–Fisher diffusion equation. Gautier and Vitalis (2013) relied on the solution proposed by Kimura (1955a) to estimate divergence times between populations that have been evolving under pure drift, from single time-point data.

**Moment-based approximations.**—The use of the diffusion approximation is limited in practice due to the high computational burden. Cavalli-Sforza and Edwards (1967) approximated pure drift as a Brownian motion process, and current moment-based approximations are reminiscent of that approach, in that they are based on mathematically convenient instrumental distributions. By relying on the equations for the mean (2) and variance (4), we can fit to the true DAF distributions that can be parameterized solely through the first two moments, such as the normal and beta distributions. These two distributions arise as special cases of the DAF approximated from the diffusion theory: the normal

![Figure 3](https://academic.oup.com/sysbio/article-abstract/66/1/e30/2670014/3070014)
For the alternative parameterization with shapes 

\[ \frac{.0011}{.0012} \]

when the number of generations 

Equations (5) and (8) are equivalent under pure drift

where Beta( 

\( m \) by mean 

distribution always verifies the condition 

\( v \)

\( x \)

can be truncated to 

boundaries, the normal distribution from equation (8)

between 0 and 1, and, under the Wright–Fisher model, 

and a positive probability can exist outside 

0 or 1 (the allele is lost or fixed, respectively). The 

Gautier et al. (2010). Gautier and Vitalis (2013) noted that 

respectively (Nicholson et al. 2002; Coop et al. 2010; 

and 1 and serve as the loss and fixation probabilities, 

for the bi-allelic Wright–Fisher model, the DAF 

allelic Wright–Fisher (see the multi-allelic section

below). For the bi-allelic Wright–Fisher model, the DAF 
can be approximated with a beta distribution as follows,

\[ x(r) | x(0) \sim \text{Beta} \left( E[x(r) | x(0)], \text{Var}(x(r) | x(0)) \right) \]

where Beta( 

\( m \) and variance \( v \). We note here that a beta 
distribution parameterized by mean \( m \) and variance \( v \).

\( r \)

\( m \)

\( v \)

\( x \)

\( m \)

\( v \)

\( v \)

\( v \)

\( m \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)

\( v \)
are discretized as in Tataru et al. (2015). The diffusion DAF is calculated as in Zhao et al. (2013), with population is by allowing the alleles to mutate (Fig. 5).

The most common way to introduce variation in a population is by allowing the alleles to mutate (Fig. 5).

**Neutral Mutations**

The most common way to introduce variation in a population is by allowing the alleles to mutate (Fig. 5).

**Dynamics and moments.**—If \( u \) is the probability of a mutation from \( A \) to \( a \), and \( v \) is the probability for the reverse event, the sampling probability from equation (1) is changed by allowing each individual to undergo a mutation after choosing its parent. Therefore, the individual is carrying an \( A \) allele if the parent had an \( A \) allele (probability \( x(r) \)) and there was no mutation (probability \( 1 - u \)), or the parent had an \( a \) allele (probability \( 1 - x(r) \)) and it mutated (probability \( v \)), leading to a sampling probability

\[
x(r)(1-u)+(1-x(r))v=(1-u-v)x(r)+v.
\]

Then, the binomial distribution of \( z(r+1) \) becomes

\[
z(r+1) \sim \text{Bin}(2N, (1-u-v)x(r)+v).
\]

For large \( N \), Crow and Kimura (1956) derived general formulas for all moments of \( x(r) \). The mean and variance after \( r \) generations of evolution can also be obtained by repeated use of the laws of total expectation and variance (Sire\n
2017

TATARU ET AL.—STATISTICAL INFERENCE IN THE WRIGHT–FISHER MODEL

e35

FIGURE 4. Fit of various approximations to the pure drift true DAF, calculated using the Markov chain property for 2N = 200 and a range of \( x(0) \) and \( r/2N \). Each column shows a different type of approximation, indicated at the top of the figure. a) Hellinger distance on log scale between the approximated and true DAF. The three ‘×’s in each of the heatmaps indicate the combinations of \( x(0) \) and \( r/2N \) used in b). b) True (dashed lines) and approximated (solid lines) DAF for \( x(0)=0.5 \) and different values of \( r/2N \). The truncated normal, beta and beta with spikes are discretized as in Tataru et al. (2015). The diffusion DAF is calculated as in Zhao et al. (2013), with \( r=0.01 \) and \( K=2N \). We used 2N = 200 for computational reasons, but we see similar patterns for larger \( N \). From the Markov chain property, using the Hellinger distance (Le Cam and Yang 2000), which lies between 0 and 1, with 0 indicating a perfect match of the two distributions. The diffusion approximation is the most accurate, whereas the truncated normal and beta distributions are the least accurate (Fig. 4). They approximate the true DAF well when the probability mass is away from the boundaries: \( x(0) \) is close to 0.5 and the generation \( r \) is not too large. As \( r \) increases, the frequency drifts away from \( x(0) \) and more and more probability accumulates at the boundaries. The beta distribution fails to capture this, whereas the atoms and spikes in the truncated normal and beta with spikes distributions, respectively, approximate these probabilities with various degrees of accuracy. Overall, the beta with spikes distribution is more accurate than both the truncated normal and beta distributions.
formulas:

\[ E[x(t) | x(0)] = \frac{v}{\mu + v} \left( x(0) - \frac{v}{\mu + v} \right) e^{-(\mu + v)t}, \]

\[ \text{Var}(x(t) | x(0)) = \frac{\mu v}{(\mu + v)^2(2(\mu + v) + 1)} \left( 1 - e^{-2(\mu + v)t} \right). \]  

Diffusion approximation.—The diffusion approximation of the Wright–Fisher model with neutral mutations is obtained in a similar way as for pure drift. Let \( \mu = 2Nu \) and \( v = 2Nv \) be the scaled mutation rates, and we again scale the time in units of \( 2N \) generations. Recall that the infinitesimal variance is independent of the evolutionary forces. For neutral mutations, the infinitesimal mean is given by

\[ d(x) = -\mu x + (1-x). \]  

When new variation is constantly introduced in the population, after enough time, the allele frequency will reach a stationary distribution. This was first obtained by Wright (1931) by noting that at stationarity, the mean and variance are unchanged between successive generations. Later on, the stationary DAF was re-derived using alternative methods, including diffusion (Wright 1945; 1938). The stationary DAF for neutral mutations is given by a beta distribution with shape parameters \( 2v \) and \( 2\mu \) (Crow and Kimura 1970; 1977; Song and Steinrücken 2012). Note that this result is in agreement with the mean (equation (11)) and variance (equation (12)) in the limit \( t \to \infty \).

The spectral decomposition method developed by Kimura (1955a) to calculate the DAF under pure drift was extended to calculate the DAF with recurrent mutation (Crow and Kimura 1956; 1970; Song and Steinrücken 2012), and to incorporate mutation rates and population sizes that vary in time in a piecewise constant manner (Steinrücken et al. 2016).

Moment-based approximations.—Using the moments of the DAF for the bi-allelic Wright–Fisher model with neutral mutations (equations (11) and (12)), the moment-based approximations are obtained just as for pure drift.

Quality of approximations.—The non-zero mutation probabilities introduce variation in the population, and reduce the loss and fixation probabilities relative to pure drift (Figs. 4 and 6). For example, under pure drift, the probability that the mutation is lost (fixed) at \( r/(2N) = 0.5 \) is 0.072, while when alleles mutate with \( \mu = v = 0.05 \), the probability is reduced to 0.05. As more of the probability mass is now found away from the 0 and 1 boundaries, all approximations have an overall improved fit to the true DAF (Fig. 6).

Migration

In its simplest form, the migration model describes the evolution of the allele frequency in one population that sends migrants, with probability \( m \), to an infinitely large population with constant allele frequency \( x_c \), and receives immigrants such that the population size stays constant over time. Then the allele count at generation \( r+1 \) is given by (Crow and Kimura 1970)

\[ z(r+1) \sim \text{Bin}(2N, (1-m)x(r) + m x_c). \]  

Under pure drift, the sampling among the alleles in generation \( r \) is done uniformly (equation (1)). However, as different evolutionary pressures act on the allele, the sampling probability is changed, as observed for neutral mutations and migration in (14). We can capture all the evolutionary pressures acting on the allele in a function \( g: [0,1] \to [0,1] \) which alters the sampling probability of the binomial distribution from equation (1). We then obtain the more general process

\[ z(r+1) | z(r) \sim \text{Bin}(2N, g(x(r))) \]  

The evolutionary pressures for pure drift, mutation, and migration are linear in \( x \) (see Box 1) and are therefore collectively called linear pressure (Crow and Kimura 1970). It is this linearity that allows the calculation of the first two moments of the DAF in closed form. One can formulate a general linear evolutionary pressure model, where pure drift, mutation and migration are special cases (see Box 1).

The migration model from equation (14) is a good approximation if the immigrants represent a random sample of the entire species (Crow and Kimura 1970). This is often not the case, and migrants are typically exchanged by at least two populations that have non-constant allele frequencies. This leads to an evolutionary pressure \( g \) that is dependent on the generation, and the DAFs of both populations need to be modeled jointly.

Markov chain theory.—Mathieson and McVean (2013) inferred effective population sizes and migration rates from time series data (Table 1) while modeling multiple...
populations distributed on a lattice, where neighboring populations exchange migrants every generation.

**Diffusion approximation.**—Gutenkunst et al. (2009) built a diffusion equation to model jointly the allele frequencies in multiple populations. They solved this equation using the finite-difference scheme to infer divergence time between populations, mutation, and migration rates. From the joint DAF, Gutenkunst et al. (2009) calculated the expected multi-population allele frequency spectrum (AFS), which summarizes allele frequency data. Because the dimension of the AFS depends on the number of populations, the time needed to compute the AFS grows exponentially with the number of populations. This limited their analysis to only three populations. Lukić and Hey (2012) also calculated the expected AFS, but they extended the spectral decomposition method to calculate the joint DAF of multiple populations that exchange migrants, while accounting for de novo mutations. The implementation of Lukić and Hey (2012) was optimized to use little memory, and can therefore tackle more than three populations. However, compared with Gutenkunst et al. (2009), it has a lower computational speed on two and three populations.

**Moment-based approximations.**—Pickrell and Pritchard (2012) used the normal distribution to infer divergence times between populations that have been evolving under pure drift and have exchanged migrants. Due to their use of the normal distribution, the method is not accurate for alleles with frequencies close to 0 or 1.

**Quality of approximations.**—As both the neutral mutation (equation (10)) and migration (equation (14)) models are special cases of the general linear evolutionary pressure model (Box 1), the quality of the approximations is similar. The approximation quality shown in Figure 6, where \( \mu = \nu = 0.05 \), also applies for \( 2N\mu = \nu = 0.01 \) and \( x_c = \nu / (\mu + \nu) = 0.5 \).

**Selection**

When selection is present, the different genotypes are transmitted to the next generation with different probabilities, determined by their fitness. If the \( A \) allele
Consider the general bi-allelic Wright–Fisher process, where \( g: [0, 1] \rightarrow [0, 1] \) captures the evolutionary pressures acting on the allele,
\[
z(r+1) | z(r) \sim \text{Bin}(2N, g(x(r))). \tag{B.1}
\]

The function \( g \) can take different forms.

**General linear evolutionary pressure:**
\[
g(x) = (1-a)x + b, \quad \text{for } 0 \leq b \leq a < 1,
\]
where \( a \) and \( b \) are given by

- Pure drift: \( a = 0, \quad b = 0 \),
- Mutation: \( a = u + v, \quad b = v \),
- Migration: \( a = m, \quad b = mx \).

Let \( A = 2Na, B = 2Nb \) and \( t = r/(2N) \). For large \( N \), the mean and variance for the DAF are given by (Tataru et al. 2015)
\[
E[x(t)|x(0)] = \frac{B}{A} e^{-At} \left[ x_0 - \frac{B}{A} \right], \tag{B.4}
\]
\[
\text{Var}(x(t)|x(0)) = \frac{B}{A} \left( 1 - \frac{B}{A} \right) \frac{1 - e^{-2At}}{2A+1} - \left( x_0 - \frac{B}{A} \right)^2 e^{-2At} \left( 1 - e^{-t} \right) + \left( 1 - \frac{2B}{A} \right) \left( x_0 - \frac{B}{A} \right) e^{-At} \frac{1 - e^{-(A+1)t}}{A+1} \tag{B.5}
\]

For pure drift, \( A = B = 0 \) and we set \( 0/0 = 1 \). Note that equations (2), (4), (11), and (12) can be obtained as special cases of the above.

**Selection (non-linear evolutionary pressure):**
\[
g(x) = \frac{(1+s)x^2 + (1+sh)x(1-x)}{(1+s)x^2 + 2(1+sh)x(1-x) + (1-x)^2}
\approx x + sx(1-x)(h+(1-2h)x), \tag{B.6}
\]

where the approximation relies on the selection coefficients \( s \) and \( sh \) being small (Crow and Kimura 1970).

**Selection with linear evolutionary pressure:** Alleles can undergo linear evolutionary pressure and selection jointly. Then,
\[
g(x) = (1-a) \left\{ x + sx(1-x)(h+(1-2h)x) \right\} + b. \tag{B.8}
\]

**Stationary distribution:** When \( A, B \not= 0 \), variation is constantly introduced in the population and the DAF has a stationary distribution given by (up to a normalization constant),
\[
f(x) \propto x^{2s-1} (1-x)^{2(A-B)-1} e^{sx(2h+(1-2h)x)}, \tag{B.9}
\]
where \( S = 2Ns \) is the scaled selection coefficient. When \( s = 0 \), we obtain a beta distribution with shape parameters \( 2B \) and \( 2(A-B) \), which is in agreement with the expressions for mean and variance in the limit \( t \to \infty \).

---

**Box 1  Evolutionary models for the bi-allelic Wright–Fisher**

The allele count \( z(r+1) \) still follows the process given in equation (15), with the evolutionary pressure function from equation \((B.7)\).
the form
\[
E[x(t+1) | x(0)] = E[g(x(t)) | x(0)].
\]  
(16)

Var(x(t+1) | x(0))
\[
= \frac{1}{2N} E[x(t+1) | x(0)] - E[x(t+1) | x(0)]^2
\]  
+ \left(1-\frac{1}{2N}\right) E[g(x(t))^2 | x(0)].
\]  
(17)

The evaluation of \(E[x(t(x(r))) | x(0)]\) and \(E[g(x(t))^2 | x(0)]\) typically requires all moments of \(x(r)\). However, these can be written as functions of only the first two moments when \(g(x)\) is a linear function in \(x\), allowing the above recursions to be solved in closed form (Tataru et al. 2015). When the allele is under selection and \(g(x)\) is no longer linear, we can approximate \(E[g(x(t))^2 | x(0)]\) by using only the first two moments by relying on a Taylor series for \(E[g(x(t))]\) about \(\bar{x}x\). By iterating the recursions above and calculating numerically the first two moments of \(x(t)\), we can recover the mean and variance of the DAF after \(r\) generations.

Markov chain theory.—Mathieson and McVean (2013) and Gompert (2015) inferred selection from time series data by discretizing continuous allele frequencies and building a Markov chain with normal and beta transition probabilities, respectively (Table 1). Gompert (2015) additionally allowed for variability in time of selection coefficients and population sizes.

Diffusion approximation.—For a Wright–Fisher model with drift, mutation and selection, specified by equations (B.1), (B.2), (B.3), and (B.8), and letting \(S = 2N\mu\), we obtain the following infinitesimal mean
\[
a(x) = -ux + \mu (1-x) + sx(1-x)(k(1-2h)x).
\]

The diffusion equation when selection is present is the most difficult to solve. However, the stationary distribution is known in closed form (Wright 1937; Crow and Kimura 1970; Eyre 2004) and is, up to a normalization constant, given by a tilted beta distribution
\[
f(x) \propto x^{2v_1-1} \left(1-x\right)^{2v_2-1} \left(1-x^2(1-x)^k(1-2h)x\right).
\]  
(20)

We note here that the diffusion limit to the Wright–Fisher model requires that the parameters involved in the evolutionary pressure, \(u, v, m, s\), and \(sh\), are all in the order of 1/(2N), such that the resulting scaled parameters, 2Nu, 2Nv, 2Nm, 2Ns, and 2Nsh, are in the order of 1. This is the source of the approximation of equation (B.6) with equation (B.7), and of the common practice of simplifying expressions by removing “small” terms (Feller et al. 1951; Wakeley 2005). It also indicates that in the diffusion limit, the population size \(N\) acts as a scaling factor, and a rescaling of the parameters and time by a constant factor will not affect the DAF. This result is responsible for the notion that it is impossible to estimate, for example, the mutation rate and effective population size separately. However, although it may be true that there is low power in doing so, this is simply a consequence of the assumptions of the diffusion approximation. These might be expected to break down in cases in which the diffusion is not appropriate (Wakeley 2005). In this respect, the moment-based approximations are free of the small parameters assumption, especially because the mean and variance of the general linear evolutionary pressure can be calculated without making the approximation of large \(N\) (Tataru et al. 2015). Therefore, moment-based approximations might be more appropriate when the evolutionary pressure is strong (Lacerda and Seoighe 2014).

Using the spectral decomposition of the diffusion equation, Kimura (1955b; 1957) found the DAF when selection is present. This approach was extended by Song and Steinrücken (2012) to improve the convergence properties for stronger selection, whereas
Steinrücken et al. (2016) developed it further to model selection coefficients that vary over time in a piecewise constant manner. The DAF was also calculated using a finite-difference scheme (Bollback et al. 2008), finite-volume scheme (Zhao et al. 2013), a path integral formalism (Schraiber 2014) and other numerical approaches (Malaspinas et al. 2012; Ferrer-Admetlla et al. 2016).

Bollback et al. (2008); Steinrücken et al. (2014); Malaspinas et al. (2012) and Ferrer-Admetlla et al. (2016) estimated jointly selection coefficients and effective population sizes from time series data from one population. Ferrer-Admetlla et al. (2016) could additionally infer mutation rates. Živković et al. (2013) used the spectral decomposition of Song and Steinrücken (2012) to infer mutation, selection and variable population size from present data from one population. Vitalis et al. (2014) used the stationary distribution of the DAF when multiple populations exchange migrants and experience selection. As they used the stationary DAF, they could not recover any information about the divergence of the populations. We would like to note here that although the method of Gutenkunst et al. (2009) can in principle incorporate selection, the inference software does not estimate selection coefficients.

Moment-based approximations.—Using the numerically approximated moments of the DAF, the truncated normal and beta distributions are obtained as previously. The beta with spikes approximation has not been extended to include selection. However, the approximation developed by Tataru et al. (2015) for the loss and fixation probabilities should still be reasonable if the selection pressure is small and the loss and fixation probabilities are mainly dominated by genetic drift. Moment-based approximations have had limited use for inference of selection due to the difficulties in calculating the first two moments of the DAF. Both Lacerda and Seoighe (2014) and Terhorst et al. (2015) estimated effective population sizes and selection coefficients from time series data, using the normal distribution and the Taylor expansion approach. One critical difference between the two is that Lacerda and Seoighe (2014) assumed additive selection ($r=0.5$) and used a Taylor series about the mean of $x_t$, whereas Terhorst et al. (2015) made no assumptions about dominance and used a Taylor series about the deterministic trajectory. Additionally, Terhorst et al. (2015) were the first to incorporate linkage, but in practice their model is limited to jointly analyze only a small number of loci (typically 3).

Quality of approximations.—Relative to pure drift, positive selection acts by increasing the expected frequency and probability of fixation of the $A$ allele, and decreasing the probability of loss (Figs. 4 and 7). For example, under pure drift and with a beginning frequency of $x(0)=0.5$, the probability that the mutation is lost (fixed) at $t/(2N)=0.5$ is 0.072 (0.072), while when selection is present with $S=1$, the probability is reduced (increased) to 0.06 (0.078). Overall, for $S=1$, all approximations have a fit to the true DAF (Fig. 7) that is very similar to that for pure drift (Fig. 4). We note here that $S=1$ is a very small selection coefficient. For larger values of $S$, the Taylor series approach leads to estimated values for the mean $m$ and variance $v$ for which $v=m(1-m)$, and these cannot be fitted by a beta distribution.

**Multi-Allelic Wright–Fisher Model**

The bi-allelic Wright–Fisher model is typically a very good approximation for SNP data (because the per-nucleotide mutation rate is typically small), but due to highly mutable sites, ancestral polymorphism, very large sample size or large evolutionary distance, a number of SNPs may contain 3 or 4 alleles. Furthermore, highly variable loci (e.g., short tandem repeats) are still widely used, especially in forensics (Balding and Nichols 1997; Balding and Steele 2015), and are typically multi-allele.

In these cases, the data can be analyzed using the multi-allelic Wright–Fisher model, an extension of the bi-allelic model. Instead of following the frequency of one allele, which is sampled from a binomial distribution from one generation to the next, the multi-allelic model describes the joint distribution of the $K$ alleles present in the population, which are now sampled from one generation to the next from a multinomial distribution.

**Pure Drift**

Similar to the bi-allelic model, the simplest form is the pure random genetic drift model, where the stochastic fluctuations in the allele frequencies are purely determined by the random mating of the finite population (Fig. 8).

**Dynamics and moments.**—Let $z_i(r)$ be the number of $i$ alleles in generation $r$, $z(r)=z_1(r),\ldots,z_K(r)$ and $x(r)=z(r)/(2N)$ be the corresponding allele frequency. The distribution of $z(t+1)$ is

$$z(t+1)|z(t)\sim Mult(2N, x(t)).$$

Here, $Mult(n, p)$ is the multinomial distribution with sample size $n$ and probability vector $p$.

To determine the mean and covariance of the DAF, we move from discrete generations to continuous time, where one time unit corresponds to 2N generations, and set $t=r/(2N)$. Then,

$$E[x(t)|x(0)]=x(0),$$

$$Var(x(t)|x(0))=$$

$$= \left(1-e^{-t}\right)\{\text{diag}(x(0))-x(0)x(0)'\},$$
where $\mathbf{t}$ denotes vector transpose. These formulas are natural extensions of equations (2) and (4).

**Diffusion approximation.**—Diffusion theory can be extended from the bi-allelic to the multi-allelic case. We will not cover this here, but refer to Ewens (2004; section 4.8, p. 151) for a general discussion of multidimensional diffusion processes, and Ewens (2004; section 5.10, p. 192) for the $K$-allele pure drift Wright–Fisher model. In particular, Ewens (2004) mentions that a generalization of equation (7) can be formulated and that a generalization of Kimura’s solution in terms of orthogonal polynomials exists.

**Moment-based approximations.**—The beta distribution is a natural choice for approximating the DAF for the bi-allelic Wright–Fisher model, and it provides a good approximation when the allele is not close to being lost or fixed (Figs. 4, 6, and 7). It is therefore natural to approximate the DAF for the multi-allelic Wright–Fisher using the generalization of the beta distribution, the Dirichlet distribution (Balding and Nichols 1995, 1997). Just like for the bi-allelic case, where the beta distribution arises as the stationary DAF under linear evolutionary pressure, the Dirichlet distribution is the stationary DAF for a specific mutation model (Ewens 2004) (see below).

Under the Dirichlet model, also called the Balding–Nichols model (Balding and Steele 2015), the allele frequency vector $x(t)$ follows a Dirichlet distribution

$$x(t) | x(0) \sim \text{Dirichlet}(\mathbf{a})$$

where $\mathbf{a}=(a_1, \ldots, a_K) > 0$. This implies that allele $i=1, \ldots, K$ has marginal distribution

$$x_i(t) \sim \text{Beta}(a_i, a_0-a_i), \quad \text{with } a_0 = \sum_{i=1}^K a_i.$$
Under the Dirichlet distribution, the mean and covariance of the DAF are

\[ \mathbb{E}[x(t)|x(0)] = \frac{\alpha}{\alpha_0}, \]
\[ \text{Var}(x(t)|x(0)) = \frac{1}{\alpha_0 + 1} \left( \frac{\alpha}{\alpha_0} \right) - \left( \frac{\alpha}{\alpha_0} \right) \left( \frac{\alpha}{\alpha_0} \right) - \frac{1}{\alpha_0 + 1}. \]  

The mean and covariance of the DAF (equations (22) and (23)) are equivalent to those under the Dirichlet distribution (equations (24) and (25)) when

\[ x(0) = \frac{\alpha}{\alpha_0} \quad \text{and} \quad 1 - e^{-t} = \frac{1}{\alpha_0 + 1}. \]

Therefore, the Dirichlet distribution can accurately capture the true mean and covariance of the multi-allelic pure drift Wright-Fisher model.

**Neutral Mutations**

Just as is the case for the bi-allelic model (Fig. 3), when the alleles evolve under pure drift, eventually the process will reach a monomorphic state, where only one of the alleles will be present in the population. The variation can be maintained in the population by allowing mutations (Fig. 9).

**Dynamics and moments.—** If \( U_i \) is the probability of an \( i \) allele to mutate to a \( j \) allele, the multinomial distribution of \( z(r+1) \) becomes

\[ z(r+1) \sim \text{Mult}(2N, x(r)U). \]

where the mutation probabilities are stored in a \( K \times K \) matrix \( U \). By specifying the structure of \( U \), different evolutionary mutation models can be formulated, such as the Jukes-Cantor (JC) model, parent independent mutation model, infinite alleles model, Kimura model, and single-step mutation model (Felsenstein 2004).

The mean and covariance of the DAF in continuous time \( t=r/(2N) \) are obtained using the rate matrix \( Q=2N(U-I) \), where \( I \) is the identity matrix, from the diffusion approximation (Hobolth and Sírén 2016),

\[ \mathbb{E}[x(t)|x(0)] = x(0)e^{Qt}, \]
\[ \text{Var}(x(t)|x(0)) = \int_0^t e^{-s} \left( e^{Qs} - I \right) \left( e^{Qs} - I \right) ds \]
\[ = e^{Qt}x(0)e^{Qs} \left( 1 - e^{-s} \right). \]

These general formulas make it possible to numerically calculate the mean and covariance for any mutation model. In practice, the mean can be calculated using one of the many available numerical procedures for matrix exponentials (Moler and Van Loan 2003). Calculating the covariance, which involves integrals of matrix exponentials, is more tedious, but this can be done numerically using the eigenvalue decomposition of the rate matrix (Hobolth and Sírén 2016).

The JC is the most simple mutation model, where all mutation probabilities are equal, \( U_{ij} = u/(K-1) \), for all \( i \neq j \). The entries in the rate matrix for the JC model are given by

\[ Q_{ij} = 2N(U_{ij} - I_{ij}) = \begin{cases} q_{ij} & \text{if } i \neq j \\ -q & \text{if } i = j \end{cases} \]

where \( q = 2Nu \). The rate matrix can be written in matrix form as

\[ Q = \frac{q}{K-1}(E-IK), \]

where \( E \) is the \( K \times K \) matrix with 1 in every entry. We can now obtain a closed-form solution for the matrix exponential \( e^{Qt} \), namely

\[ e^{Qt} = e^{tQ} \left( I - E \right) + E \frac{E}{K}, \]

where \( s = 2aK/(K-1) \). The mean and covariance in the JC model are found from equations (26) and (27) and given by

\[ \mathbb{E}[x(t)|x(0)] = e^{tQ} \left( x(0) - \frac{e}{K} \right) + \frac{e}{K}, \]
\[ \text{Var}(x(t)|x(0)) = \frac{1}{K} \left( I - E \right) \frac{1}{1+s} \left( 1 - e^{-rK} \right) \]
\[ - \left( x(0) - \frac{e}{K} \right) \left( x(0) - \frac{e}{K} \right) e^{rK} \left( 1 - e^{-rK} \right) \]
\[ + \left( \text{diag} \left[ x(0) - \frac{e}{K} \right] \right) - \left( x(0) - \frac{e}{K} \right) \frac{e}{K} \]
\[ - \frac{e}{K} \left( x(0) - \frac{e}{K} \right) \]
\[ \times e^{rK} \frac{1}{1+s} \left( 1 - e^{-rK} \right). \]
we observe that the expressions are approximately
\[ q = e \]
where \( e \) is the 1 \( \times K \) vector with 1 in every entry.

For \( t \to \infty \), these reduce to
\[
E[x(t) | x(0)] = \frac{e}{K}
\]
\[
\text{Var}(x(t) | x(0)) = \frac{1}{K} \left( I - \frac{E}{K} \right) \frac{1}{1+\epsilon}.
\]

We note that these moments are the same as for a Dirichlet distribution with \( \alpha = e/(Ke) \), and indeed the Dirichlet distribution is the stationary DAF of the multi-allelic Wright–Fisher model (Ewens 2004).

Moment-based approximations.—The mean and covariance of the Dirichlet distribution (equations (24) and (25)) are equivalent to those under the JC model if the covariance approximately fulfills the proportionality condition
\[
\text{Var}(x(t) | x(0)) = \frac{1}{K} \left( I - \frac{E}{K} \right) \frac{1}{1+\epsilon}.
\]

Regardless of the parameter \( \epsilon \), the expressions are also approximately proportional, with proportionality constant \( t \), when the evolutionary distance \( t \) is small. Finally, for large \( t \), the proportionality constant is 1, because the Dirichlet distribution is the stationary distribution for the JC model. These analytical considerations are confirmed by Figure 10. The Dirichlet distribution cannot accurately capture the mean and covariance of the JC model for intermediate values of \( t \), and the deviation is very clear for large values of \( \epsilon \) (Fig. 10b). Therefore, care should be taken when using the Dirichlet distribution in practice. Because the JC is the most simple mutation model, with just one parameter, one could expect that the fit of the Dirichlet distribution could be even more problematic for more complex mutation models. An important step in developing more appropriate distributions for the DAF under the multi-allelic Wright–Fisher model is made by Sirén et al. (2013) and Hubollth and Sirén (2016), but in general more research is needed in this direction.

**CONCLUSION AND PERSPECTIVES**

We have provided a broad overview of methods to calculate the DAF under the Wright–Fisher model. These methods have a number of working assumptions in common. Here, we discuss in turn each of these and how current methods tackle these issues or potentially could be improved to do so.

Virtually all methods presented here rely on unlinked loci, with an exception worth mentioning using a moment-based approach (Terhorst et al. 2015). Several

![Figure 10](https://academic.oup.com/sysbio/article-abstract/66/1/e30/2670014/fig10)
inference methods built on the coalescent process analyze pairs of linked neutral loci (Li and Durbin 2011; Paul et al. 2011; Sheehan et al. 2013; Schiffels and Durbin 2014; Rasmussen et al. 2014). Some theoretical results do exist for linked bi-allelic selected loci (Kermany 2012), but these have limited use for inference methods. To our knowledge, equivalent results for the joint DAF of two loci are not available, but see Jenkins et al. (2014) for an approximation for loosely linked loci. These results could be used as the basis for including recombination in inference methods, but so far these types of approximations have not really permeated the field of inference under the Wright–Fisher model.

The methods we have presented here also implicitly assume very simplified demographic scenarios. In some instances, especially if the data contain individuals sampled from populations with complex demography, it might be easier, although more computationally intensive, to rely on simulations under the Wright–Fisher model to perform inference (Excoffier et al. 2013). In particular, much progress has been made in using Approximate Bayesian Computation (ABC) that relies on a series of summary statistics from simulations and either rejection sampling or more sophisticated methods to obtain approximated posterior distributions for the parameters of interest (Beaumont et al. 2002; Blum 2010; Marin et al. 2012). This is becoming permissive and is expected to undergo periodic bottlenecks, followed by population expansion. This is typical of most experimental setups, where the population size is experimentally controlled (Foll et al. 2014).

Ultimately, the directions of future method development are likely to be conditioned by the type of data modeled and the evolutionary or demographic questions of interest that motivate the need for inference. As genome-wide re-sequencing becomes increasingly common and replaces most SNP genotyping and exome sequencing, we can expect that the data will increasingly consist of polymorphism counts among tightly linked sites.

One direction worth exploring is using the Wright–Fisher model to learn about how selection varies along the genome, and thereby shapes genome-wide diversity. Some progress has been made in inferring mutation rates and selection coefficients by expressing expected local levels of nucleotide diversity as a function of the amount of selection affecting neutral sites due to linkage (Elyashiv et al. 2014).

Finally, at present, most software programs that implement inference methods have been developed for bi-allelic data, whereas inference for multi-allelic data is clearly lagging behind. We have discussed recent attempts to understand and formulate approximations for the DAF under the multi-allelic Wright–Fisher model with mutation. These developments are expected to improve modeling of short tandem repeat data that are still widely used in forensics (Balding and Steele 2015). They might also allow the analysis of a broader range of biological situations where the bi-allelic assumption is not always appropriate, for example, when there is extensive heterogeneity in the mutation rate or the product of effective population size and mutation rate is high, as is the case for microbial and viral genomes.

ACKNOWLEDGMENTS

We would like to thank Jukka Sirén, Dominik Schrempf, the associate editor, editor, and an anonymous reviewer for their constructive suggestions and comments that helped improve the manuscript.

REFERENCES

Balding D.J., Nichols R.A. 1995. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetics 96: 3–12.

Balding D.J., Nichols R.A. 1997. Significant genetic correlations among Caucasians at forensic DNA loci. Heredity 78(6):583–589.

Balding D.J., Steele C.D. 2011. Weight-of-evidence for forensic DNA profiles. Chichester: John Wiley and Sons.

Barton N.H., Otto S.P. 2005. Evolution of recombination due to random drift. Genetics 169(4):2353–2370.

Beaumont M.A., Zhang W., Balding D.J. 2002. Approximate Bayesian computation in population genetics. Genetics 162(4):2023–2035.

Bhaskar A., Song Y.S. 2009. Multi-locus match probability in a finite population: a fundamental difference between the Moran and Wright–Fisher models. Bioinformatics 25(12):i187–i195.

Blum M.G.B. 2010. Approximate Bayesian computation: a nonparametric perspective. J. Amer. Stat. Ass. 105(491):1178–1187.

Bollback J.P., York T.L., Nielsen R. 2008. Estimation of 2nes from temporal allelic frequency data. Genetics 179(1):497–502.

Cavalli-Sforza L.L., Edwards A.W.F. 1967. Phylogenetic analysis: models and estimation procedures. Amer. J. Hum. Gen. 19(3 Pt 1):233.

Chang J.S., Cooper G. 1970. A practical difference scheme for Fokker–Planck equations. J. Comp. Phy. 4(1):1–16.

Coop G., Witonsky D., Rienzo A.D., Pritchard J.K. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185(4):1411–1423.

Crow J.F., Kimura M. 1970. An introduction to population genetics theory: New York, Evanston and London: Harper & Row.
Crow J. F. 1954. Random mating with linkage in polynomials. Amer. Nat. 88(543):431-434.
Crow J. F. 1987. Population genetics history: a personal view. Ann. Rev. Gen. 21(1):1-22.
Crow J. F., Kimura M. 1956. Some genetic problems in natural populations. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, volume 4. Berkeley and Los Angeles: University of California Press, p. 1-22.
De Maio N., Schlötterer C., Kosiol C. 2013. Linking great apes genome evolution across time scales using phylogenome-aware phylogenetic models. Mol. Biol. Evol. 30(10):2249-2262.
De Maio N., Schrumpf D., Kosiol C. 2015. Pomo: An allele frequency-based approach for species tree estimation. Syst. Biol. 64(6):1018-1031.
Donnelly P. 1984. The transient behaviour of the Moran model in population genetics. In: Mathematical Proceedings of the Cambridge Philosophical Society, volume 95. Cambridge: Cambridge University Press, p. 349-359.
Durrett R. 2008. Probability models for DNA sequence evolution. New York: Springer.
Fisher R.A. 1930. The genetical theory of natural selection. Oxford: Clarendon.
Felsenstein J. 2004. Inferring phylogenies. Sunderland, Mass: Sinauer Associates.
Ferrard-Admella A., Leuenberger C., Jensen J.D., Wegmann D. 2016. An approximate Markov model for the Wright–Fisher diffusion and its application to time series data. Genetics 203(2):831–846.
Ferrer R.A. 1930. Theoretical genetics of natural selection. Proc. Roy. Soc. B 100:415–458.
Feller W. 1951. Diffusion processes in genetics. In: Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability University of California Press, Berkeley and Los Angeles, volume 227, p. 246.
Felsenstein J. 2004. Inferring phylogenies. Inc. Sunderland, Massachusetts: Sinauer Associates.
Gavrilets S., Hey J. 2012. Demographic inference using spectral methods. 1st ed. San Diego: Academic Press.
Gutenkunst R.N., Hernandez R.D., Williamson S.H., Bustamante C.D. 2009. Inference of the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5(10):e1000568.
Haal H.J., Paysse B.A. 2015. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complexity. Mol. Ecol. 24(1):5–23.
Hobolth A., Srin J. 2016. The multivariate Wright-Fisher process with mutation: Moment-based analysis and inference using a hierarchical beta model. Theor. Pop. Biol. 108:36–50.
Hobolth A., Christensen O.F., Mailund T., Scharup M.H. 2007. Genomic relationships and speciation times of humans, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. PLoS Genet. 3(2):e7.
Hui T.Y.J., Burt A. 2015. Estimating effective population size from temporally spaced samples with a novel, efficient maximum-likelihood algorithm. Genetics 201(1):285–293.
Jenkins P.A., Fearnehead P., Song Y.S. 2014. Trackable divergence and coalescent processes for weakly correlated loci. arXiv preprint arXiv:1405.6863.
Joyce P., Marjoram P. 2008. Approximately sufficient statistics and bayesian computation. Stat. Appl. Gen. Mol. Biol. 7(1):1544-6115.
Karlin S., Taylor H.M. 1975. A first course in stochastic processes. 2nd ed. San Diego: Academic Press.
Kimura M. 1955a. Solution of a process of random genetic drift with a continuous model. Proc. Nat. Acad. Sci. USA. 43(3):144.
Kimura M. 1955b. Stochastic processes and distribution of gene frequencies under natural selection. In: Cold Spring Harbor Symposium on Quantitative Biology, volume 20. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, p. 33–53.
Kimura M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28(4):882-901.
Kimura M. 1964. Diffusion models in population genetics. J. Appl. Proc. 1(2):175–222.
Kingman J.F.C. 1982a. Exchangeability and the evolution of large populations. In: Koch G., Spitzichino F., editors, Exchangeability in probability and statistics. Amsterdam: North-Holland, p. 97–112.
Kingman J.F.C. 1982b. On the genealogy of large populations. J. Appl. Proc. J.A. 27–43.
Kingman J.F.C. 1982c. The coalescent. Stoc. Proc. Appl. 13(2):1555–1579.
Liu X., Fu Y.-X. 2015. Exploring population size changes using SNP frequency spectra. Nat. Genet. 47:535-539.
Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Biol. Evol. 30(3):654–668.
Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S.V. 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phylog. Evol. 51(2):469-477.
Liu L., Wu S., Lemmon A.R., Lemmon E.M., Leaché A.D., Hui T.Y.J., Burt A. 2015. Estimating allele age and selection coefficient from time-serial data. Genetics 192(2):599–607.
Marin J-M., Pudlo P., Robert C.P., Ryder R.J. 2012. Approximate Bayesian computational methods. Statistics and Computing 22(6):1067–1080.

Matheson I., McVean G. 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics 193(3):973–984.

Moler C., Loan C.V. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. 45(1):3–49.

Monahan P.A. 1958. Random processes in genetics. In: Mathematical Proceedings of the Cambridge Philosophical Society, volume 54. Cambridge: Cambridge University Press, p. 60–71.

Murhead C.A., Wakeley J. 2009. Modeling multiallelic selection using a Moran model. Genetics 182(4):1141–1157.

Nicholson G. Smith A.V., Jónsson F., Gústafsson Ó, Stefánsson K., Donnelly P. 2002. Assessing population differentiation and isolation from single-nucleotide polymorphism data. J. Roy. Stat. Soc. Ser. B (Stat. Meth.) 64(4):695–715.

Nielson R., Statkin M. 2013. An accurate sequentially Markov conditional sampling distribution for the coalescent with recombination. Genetics 192(4):1115–1128.

Pickrell K.J., Pritchard J. K. 2012. Inference of population splits and mixtures from genomewide allele frequency data. PLoS Gen. 8(1):e1002967.

Rasmussen M.D., Huibsz M.J., Gronau I., Siepel A. 2014. Genome-wide inference of ancestral recombination graphs. PLoS Gen., 10(5):e1004342.

Schraiber J.G. 2014. A path integral formulation of the Wright–Fisher model using an accurate beta approximation. Genetics 192:1134–1141.

Terhorst J., Schlotterer C., Song Y.S. 2015. Multi-locus analyses of genomewide time series data from experimental evolution. PLoS Gen. 11(4):e1005069.

Vogl C., Clemente F. 2012. The allele-frequency spectrum in a decoupled Moran model with mutation, drift, and directional selection, assuming small mutation rates. Theor. Pop. Biol. 81(3):197–209.

Wakeley J. 2005. The limits of theoretical population genetics. Genetics 169(1):1–7.

Wakeley J. 2009. Coalescent theory: an introduction. Roberts & Company Publishers, Greenwood Village.

Wang J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Gen. Res. 78(03):243–257.

Williamson S.G., Slatkin M. 1999. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics 152(2):755–761.

Wright S. 1931. Evolution in Mendelian populations. Genetics 16:97–129.

Wright S. 1937. The distribution of gene frequencies in populations. Proc. Nat. Acad. Sci. 23(6):307–320.

Wright S. 1938. The distribution of gene frequencies under irreversible mutation. Proc. Nat. Acad. Sci. 24(7):233–239.

Wright S. 1942. Statistical genetics and evolution. Bull. Amer. Math. Soc. 48(4):223–246.

Wright S. 1943. The differential equation of the distribution of gene frequencies. Proc. Nat Acad Sci. USA 31(2):382.

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