Syst. Biol. 66(1):e30–e46, 2017

DOI:10.1093/sysbio/syw056
Advance Access publication August 2, 2016

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*

Received 4 December 2015; reviews returned 31 May 2016; accepted 6 June 2016
Associate Editor: David Bryant

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

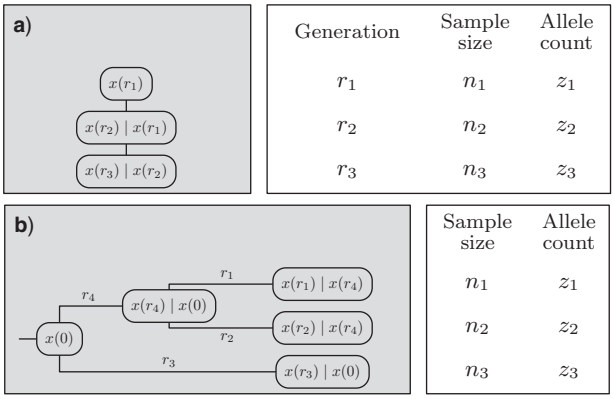

FIGURE 1. Data types. The gray boxes represent the unobserved history of the populations, together with the corresponding population allele frequency $x(r)$, whereas the white boxes indicate the observed data: the generation r when the data are sampled, the size n of the sample, and the allele count z, that is, how many alleles of a given type have been observed among the genotyped individuals. Given the population frequency $x(r_i)$, z_i follows a binomial distribution with size n_i and probability $x(r_i)$. In order to calculate the likelihood of the data, the DAF of $x(r_i)$ is needed. a) Time series data where, typically, one population is sampled at different (known) generations. b) Single time-point data, where multiple populations are sampled just once, typically in the present. The history of the populations is given as a tree. The leaves and internal nodes represent the sampled and ancestral populations, respectively. The branch lengths reflect the amount of time populations have diverged since the split from the ancestral population.

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 to calculate the DAF under the general 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.

A variety of methods that are grounded in the Wright–Fisher model use a range of tests and/or summary statistics to detect population differentiation (Balding and Nichols 1995; 1997; Nicholson et al. 2002; Gaggiotti and Foll 2010), or carry out genome-wide scans for selection (Foll and Gaggiotti 2008; Coop et al. 2010; Gautier et al. 2010; Gautier 2015). Several of these methods use some of the approaches for calculating the DAF discussed here. However, they do not directly use or estimate the effect of the different evolutionary forces on the DAF. Therefore, we do not review such methods and refer the reader instead to Haasl and Payseur (2015) for details.

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; Schiffels 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

TABLE 1. Overview of recent inference methods for the bi-allelic Wright–Fisher model

Reference	Data	Mut	Mig	Sel	Approach	Availability
Markov chain theory						
Mathieson and McVean (2013)[a]	T		×	×	Normal	-
Gompert (2015)[a]	T			×	Beta	spatpg
Diffusion approximation						
Bollback et al. (2008)	T			×	Finite-difference	-
Gutenkunst et al. (2009)	S	×	×		Finite-difference	∂a∂i
Lukić and Hey (2012)	S	×	×		Spectral decomposition	MultiPop
Malaspinas et al. (2012)	T			×	Numerical approximation	upon request
Gautier and Vitalis (2013)	S				Spectral decomposition	KimTree
Steinrücken et al. (2014)	T	×		×	Spectral decomposition	spectralHMM
Vitalis et al. (2014)	S		×	×	Stationary DAF	SelEstim
Živković et al. (2015)	S	×		×	Spectral decomposition	upon request
Ferrer-Admetlla et al. (2016)	T	×		×	Numerical approximation	ApproxWF
Moment-based approximations						
Sirén et al. (2011)	S				Beta	-
Pickrell and Pritchard (2012)	S		×		Normal	TreeMix
Lacerda and Seoighe (2014)	T			×	Normal	upon request
Hui and Burt (2015)	T				Beta	NB
Tataru et al. (2015)	S				Beta with spikes	SpikeyTree
Terhorst et al. (2015)	T			×	Normal	EandR-timeseries

Notes: The table indicates what type of data the method uses (*Data*): time series data from one population (T) or single time-point data from multiple populations (S); if the method models new mutations (*Mut*), migration (*Mig*) or selection (*Sel*); which type of approach is used for calculating the DAF (*Approach*); and whether the method is publicly available (*Availability*). All methods model genetic drift.
[a]analyze jointly time series data from multiple populations. The table covers only the more recent inference methods.

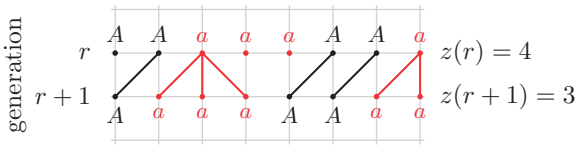

FIGURE 2. Dynamics in the pure drift bi-allelic Wright–Fisher model. The child inherits the parental allele.

corresponding allele frequency. The random mating of the population leads to a count of A alleles in generation $r+1$ that is binomially distributed (Fisher 1930; Wright 1931; Crow and Kimura 1970; Ewens 2004)

$$z(r+1)\,|\,z(r) \sim \mathrm{Bin}\big(2N, x(r)\big). \tag{1}$$

Here, $\mathrm{Bin}(n,p)$ is the binomial distribution with sample size n and probability p. The genetic variation present in the population is due to ancestral polymorphism, and because no new variation is added, the A allele is eventually fixed or lost (Fig. 3a).

The goal is to determine the DAF: the distribution $f(x;r)$ of $x(r)$, after evolving for r generations from an initial frequency $x(0)$ (Fig. 3b). We first calculate the first two moments of the DAF. From the binomial sampling, the mean and variance over one generation are given by

$$\mathrm{E}[x(r+1)\,|\,x(r)] = x(r),$$
$$\mathrm{Var}\big(x(r+1)\,|\,x(r)\big) = \frac{1}{2N}x(r)(1-x(r)).$$

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

$$\mathrm{E}[x(r)\,|\,x(0)] = x(0), \tag{2}$$
$$\mathrm{Var}\big(x(r)\,|\,x(0)\big) = x(0)(1-x(0))\left(1-\left(1-\frac{1}{2N}\right)^r\right). \tag{3}$$

For large N, we can approximate the variance by

$$\mathrm{Var}\big(x(t)\,|\,x(0)\big) \approx x(0)(1-x(0))\left(1-e^{-t}\right), \tag{4}$$

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 et al. 1951; Wakeley 2005)

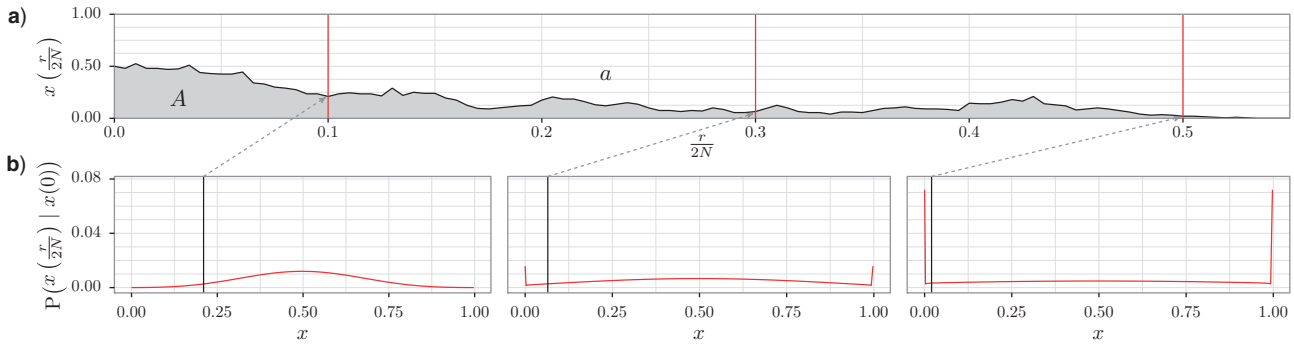

FIGURE 3. a) Simulation under the pure drift model (equation (1)) with $2N=200$ and $x(0)=0.5$. The vertical bars indicate three sampled time-points. The x-axis denotes the time measured in scaled number of generations. b) DAF at the three sampled time-points. The vertical bars indicate the simulated allele frequencies.

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 $\delta t=1/(2N)$ such that one time unit corresponds to $2N$ generations. Then, we have

$$2Nx(t+\delta t)\,|\,x(t)\sim \mathrm{Bin}(2N, x(t)),$$

from which we can approximate

$$x(t+\delta t)\,|\,x(t)\approx \mathrm{N}(x(t), x(t)(1-x(t))\delta t). \tag{5}$$

Here, $\mathrm{N}(\mu,\sigma^2)$ is the normal distribution with mean μ and variance σ^2. Equation (5) corresponds to the time-homogeneous stochastic differential equation

$$dx(t)=a(x)dt+\sqrt{b(x)}dw(t), \tag{6}$$

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)

$$\frac{\partial f(x;t)}{\partial t}=-\frac{\partial}{\partial x}\Big\{a(x)f(x;t)\Big\}+\frac{1}{2}\frac{\partial^2}{\partial x^2}\Big\{x(1-x)f(x;t)\Big\}, \tag{7}$$

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 (3, 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

distribution is a transient distribution (equation (5)) which is appropriate for very short evolutionary times, whereas the stationary DAF under linear evolutionary pressure is given by a beta distribution (see Box 1, equation (B.9)).

Several authors used the normal distribution (Nicholson et al. 2002; Coop et al. 2010; Gautier et al. 2010; Pickrell and Pritchard 2012; Lacerda and Seoighe 2014; Terhorst et al. 2015), which takes the form

$$x(r)\,|\,x(0) \sim \mathrm{N}\Big(\mathrm{E}[x(r)\,|\,x(0)], \mathrm{Var}(x(r)\,|\,x(0))\Big). \qquad (8)$$

Equations (5) and (8) are equivalent under pure drift when the number of generations r is small relative to the population size $2N$. Then, by using the approximation

$$1-\left(1-\frac{1}{2N}\right)^r \approx 1-e^{-\frac{r}{2N}} \approx \frac{r}{2N},$$

in the variance equation (3), we recover equation (5) from equation (8) with $\delta t = r/(2N)$.

Balding and Nichols (1995; 1997) first proposed the use of the Dirichlet distribution, the multivariate generalization of the beta distribution, for the multi-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,

$$x(r)\,|\,x(0) \sim \mathrm{Beta}\Big(\mathrm{E}[x(r)\,|\,x(0)], \mathrm{Var}(x(r)\,|\,x(0))\Big),$$

where $\mathrm{Beta}(m,v)$ is the beta distribution parameterized by mean m and variance v. We note here that a beta distribution always verifies the condition $v < m(1-m)$. For the alternative parameterization with shapes α and β, we have the relation

$$\begin{aligned}\alpha &= \left(\frac{m(1-m)}{v}-1\right)m,\\ \beta &= \left(\frac{m(1-m)}{v}-1\right)(1-m).\end{aligned} \qquad (9)$$

Although both the normal and beta distributions have been used for inference, they differ in accuracy. One major difference comes from the support of the distributions. The allele frequency $x(r)$ always lies between 0 and 1, and, under the Wright–Fisher model, there can be a positive probability for $x(r)$ being either 0 or 1 (the allele is lost or fixed, respectively). The normal distribution is defined over the whole real line, and a positive probability can exist outside $[0,1]$. If $x(0)$ is intermediate and r is small, the probability that $x(r)$ falls outside of $[0,1]$ is small and therefore can be ignored (Pickrell and Pritchard 2012; Lacerda and Seoighe 2014; Terhorst et al. 2015). If $x(0)$ is close to the boundaries, the normal distribution from equation (8) can be truncated to $[0,1]$. The probabilities in the intervals $(-\infty,0]$ and $[1,\infty)$ are added as two atoms at 0 and 1 and serve as the loss and fixation probabilities, respectively (Nicholson et al. 2002; Coop et al. 2010; Gautier et al. 2010). Gautier and Vitalis (2013) noted that the truncated normal distribution no longer has the true variance of the DAF.

Unlike the normal distribution, the beta distribution has support in $[0,1]$. However, due to its continuous nature, the beta distribution cannot account for the discrete events that $x(r)$ can be 0 or 1. Tataru et al. (2015) addressed this issue and introduced a new approximation, the beta with spikes, a beta distribution for the polymorphic frequencies ($0 < x(r) < 1$), supplemented by two spikes at 0 and 1 accounting for the loss and fixation probabilities. Then the distribution of $x(r)$ is

$$x(r)\,|\,x(0) \sim \mathrm{Beta}^{\star}\Big(\mathrm{E}[x(r)\,|\,x(0)], \mathrm{Var}\big(x(r)\,|\,x(0)\big), p_0(r), p_1(r)\Big),$$

where $\mathrm{Beta}^{\star}(m,v,p_0,p_1)$ is the beta with spikes distribution parameterized by mean m, variance v, and probabilities p_0 and p_1 found at 0 and 1, respectively. This is given by

$$\mathrm{Beta}^{\star}(x;m,v,p_0,p_1) = p_0\delta(x) + p_1\delta(1-x) + \frac{1-p_0-p_1}{\mathrm{B}(\alpha^{\star},\beta^{\star})}\mathrm{Beta}(x;m^{\star},v^{\star}).$$

Here, $\delta(x)$ is the Dirac delta function, introduced to account for the non-zero probabilities at the boundaries, and $m^{\star}$ and $v^{\star}$ are the mean and variance of the beta distribution for the polymorphic frequencies, given by (Tataru et al. 2015)

$$m^{\star} = \frac{m-p_1}{1-p_0-p_1}, \qquad v^{\star} = \frac{v+m^2-p_1}{1-p_0-p_1} - \left(m^{\star}\right)^2.$$

The beta function $\mathrm{B}(\alpha^{\star},\beta^{\star})$ acts as a normalization factor, where $\alpha^{\star}$ and $\beta^{\star}$ are the shape parameters of $\mathrm{Beta}(m^{\star},v^{\star})$ (equation (9)).

Using the equations (2) and (3) for the mean and variance, the normal and beta approximations of the DAF can be written in closed form. However, the loss and fixation probabilities are not known in closed form, and therefore, the beta with spikes relies on a recursive approach to calculate these probabilities (see Tataru et al. (2015) for details).

The moment-based approximations have been used in a series of inference methods (Table 1). Hui and Burt (2015) used the beta distribution to infer the effective size of one population undergoing pure drift from time series data. Sirén et al. (2011) and Tataru et al. (2015) used single time-point data to infer divergence times between populations evolving under pure drift. Sirén et al. (2011) used the beta distribution, and therefore could not accurately model the alleles that are close to being lost or fixed. Tataru et al. (2015) used the beta with spikes approximation and demonstrated that the addition of spikes leads to a more accurate inference compared with merely using the beta distribution.

Quality of approximations.—We evaluated the accuracy of the approximations to the true DAF obtained

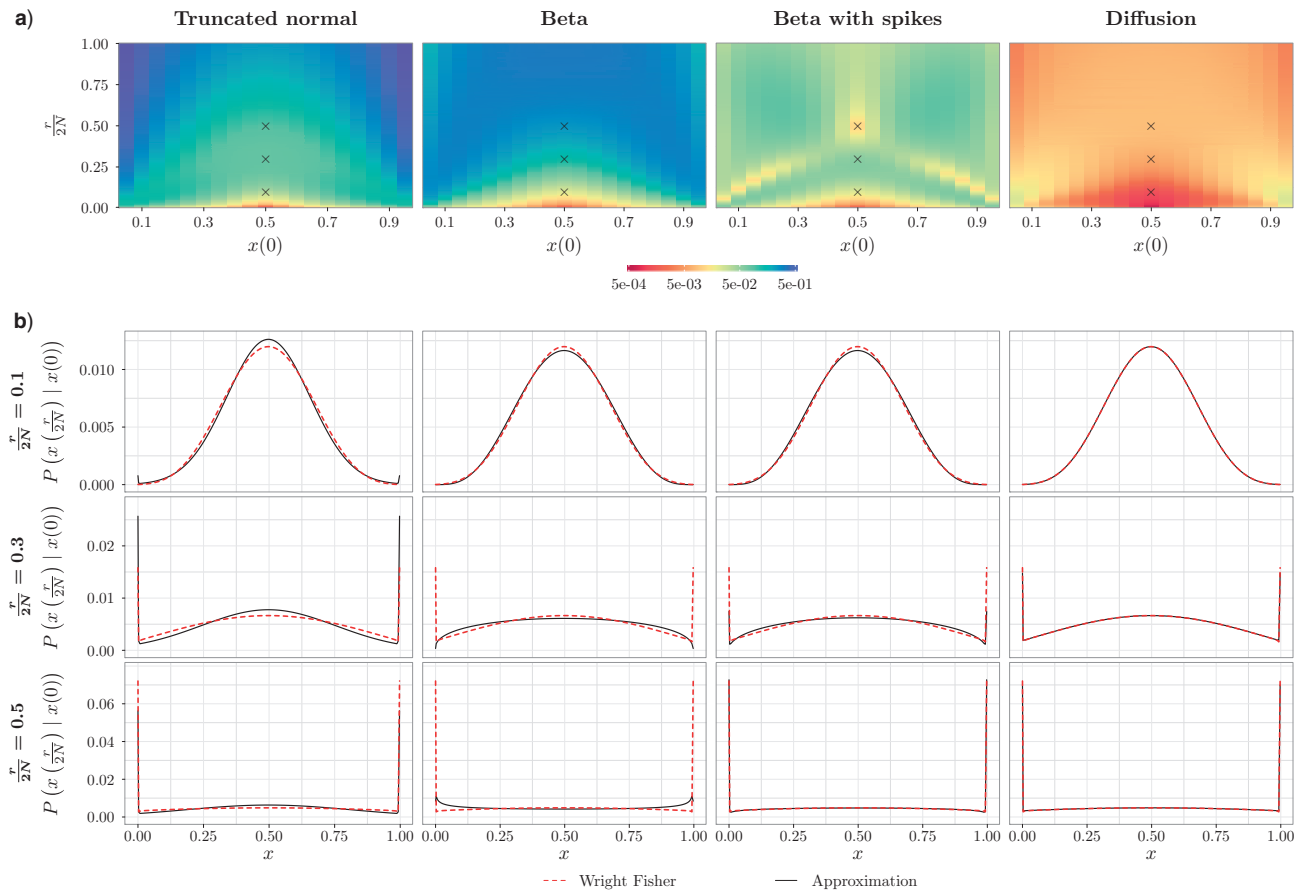

FIGURE 4. Fit of various approximations to the pure drift true DAF, calculated using the Markov chain property for $2N=200$ and a range of $x(0)$ and $r/(2N)$. Each column shows a different type of approximation, indicated at the top of the figure. a) Hellinger distance on log scale between the approximated 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 $\tau=0.01$ 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.

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

$$z(r+1)\,|\,z(r)\sim \mathrm{Bin}\big(2N,(1-u-v)x(r)+v\big). \qquad (10)$$

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 (Sirén 2012). Tataru et al. (2015) provided the

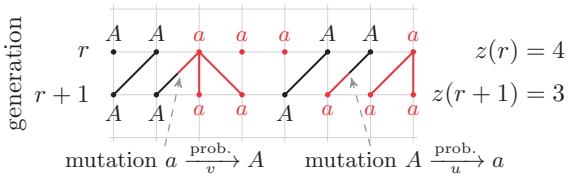

FIGURE 5. Dynamics in the bi-allelic Wright–Fisher model with mutations. If the parental allele is A, the child has the same allele with probability $1-u$, and a mutation occurs with probability u. If the parental allele is a, the child allele is a with probability $1-v$, and becomes A with probability v.

formulas:

$$\mathrm{E}[x(t)\,|\,x(0)] = \frac{\nu}{\mu+\nu} + \left(x(0) - \frac{\nu}{\mu+\nu}\right) e^{-(\mu+\nu)t}, \tag{11}$$

$$\begin{aligned}\mathrm{Var}\big(x(t)\,|\,x(0)\big) &= \frac{\mu\nu}{(\mu+\nu)^2(2(\mu+\nu)+1)}\left(1-e^{-\left(2(\mu+\nu)+1\right)t}\right) \\ &\quad - \left(x(0) - \frac{\nu}{\mu+\nu}\right)^2 e^{-2(\mu+\nu)t}\left(1-e^{-t}\right) \\ &\quad + \frac{\mu-\nu}{(\mu+\nu)(\mu+\nu+1)}\left(x(0) - \frac{\nu}{\mu+\nu}\right) \\ &\quad \cdot e^{-(\mu+\nu)t}\left(1-e^{-(\mu+\nu+1)t}\right),\end{aligned} \tag{12}$$

where $t = r/(2N)$, $\mu = 2Nu$, and $\nu = 2Nv$.

Diffusion approximation.—The diffusion approximation of the Wright–Fisher with neutral mutations is obtained in a similar way as for pure drift. Let $\mu = 2Nu$ and $\nu = 2Nv$ 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

$$a(x) = -\mu x + \nu(1-x). \tag{13}$$

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 2ν and 2μ (Crow and Kimura 1970; Ewens 2004). 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 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 $\mu = \nu = 0.05$, 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)

$$z(r+1)\,|\,z(r) \sim \mathrm{Bin}(2N, (1-m)x(r) + mx_c). \tag{14}$$

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

$$z(r+1)\,|\,z(r) \sim \mathrm{Bin}(2N, g(x(r))). \tag{15}$$

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

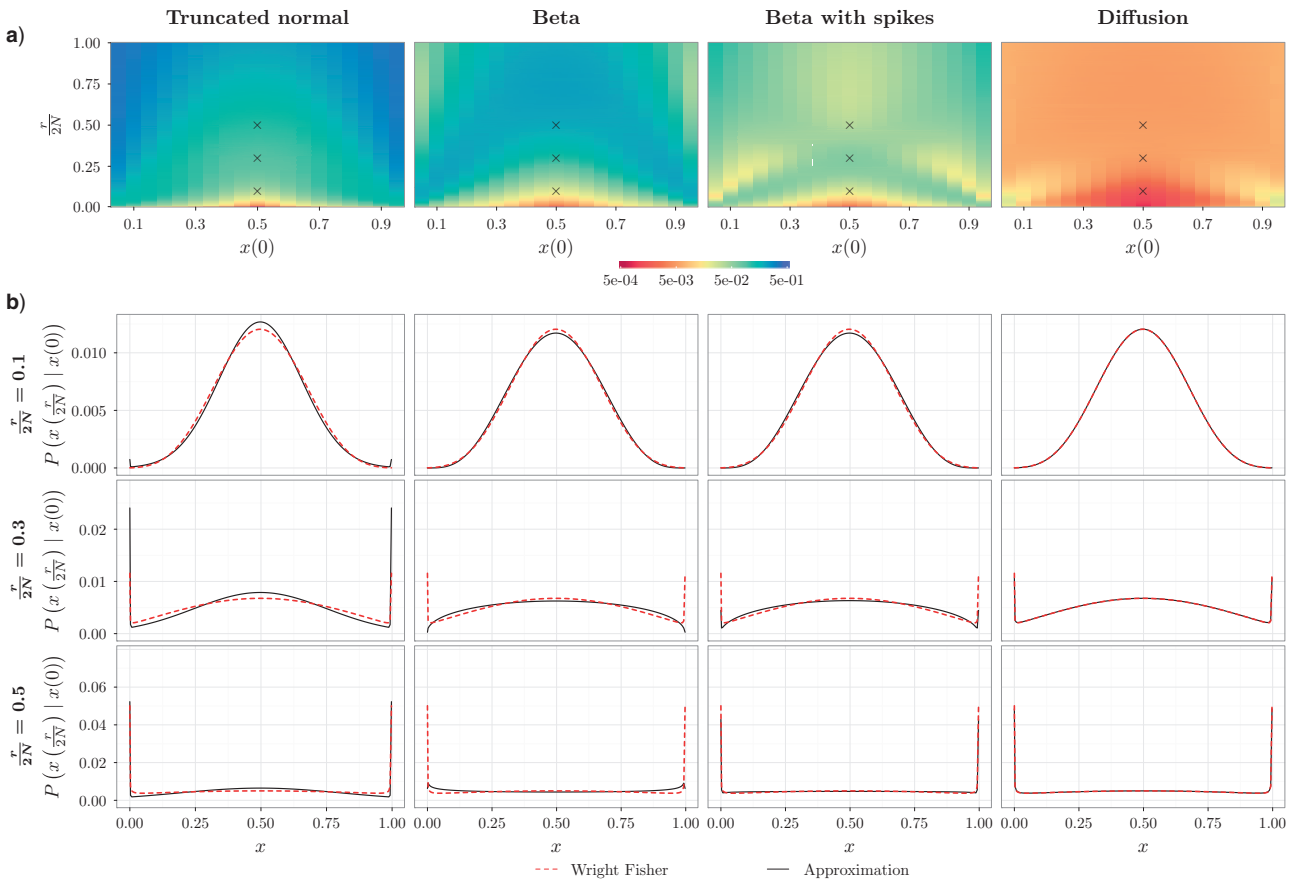

FIGURE 6. Fit of various approximations to the true DAF with neutral mutations, calculated using the Markov chain property for $2N=200$, $\mu=\nu=0.05$ and a range of $x(0)$ and $r/(2N)$. Each column shows a different type of approximation, indicated at the top of the figure. a) Hellinger distance on 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 $r/(2N)$. Calculations are performed as for Figure 4. For comparison purposes, the a) heatmap and b) y-axis scales are the same as in Figure 4.

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 $2Nm=\mu+\nu=0.01$ 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

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

Consider the general bi-allelic Wright–Fisher process, where $g:[0,1]\to[0,1]$ 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:

$$g(x)=(1-a)x+b, \quad \text{for } 0\le b\le a<1, \tag{B.2}$$

where a and b are given by

$$\begin{array}{lll} \text{Pure drift:} & a=0, & b=0, \\ \text{Mutation:} & a=u+v, & b=v, \\ \text{Migration:} & a=m, & b=mx_c. \end{array} \tag{B.3}$$

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)

$$\text{E}[x(t)\,|\,x(0)]=\frac{B}{A}+e^{-At}\left(x_0-\frac{B}{A}\right), \tag{B.4}$$

$$\begin{aligned} &\text{Var}\big(x(t)\,|\,x(0)\big) \\ &\quad=\frac{B}{A}\left(1-\frac{B}{A}\right)\frac{1-e^{-(2A+1)t}}{2A+1} \\ &\qquad-\left(x_0-\frac{B}{A}\right)^2 e^{-2At}\left(1-e^{-t}\right) \\ &\qquad+\left(1-\frac{2B}{A}\right)\left(x_0-\frac{B}{A}\right)e^{-At}\frac{1-e^{-(A+1)t}}{A+1}. \end{aligned} \tag{B.5}$$

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):

$$g(x)=\frac{(1+s)x^2+(1+sh)x(1-x)}{(1+s)x^2+2(1+sh)x(1-x)+(1-x)^2} \tag{B.6}$$

$$\approx x+sx(1-x)(h+(1-2h)x), \tag{B.7}$$

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,

$$g(x)=(1-a)\big\{x+sx(1-x)(h+(1-2h)x)\big\}+b. \tag{B.8}$$

Stationary distribution: When $A,B\neq 0$, variation is constantly introduced in the population and the DAF has a stationary distribution given by (up to a normalization constant),

$$f(x)\propto x^{2B-1}(1-x)^{2(A-B)-1}e^{Sx(2h+(1-2h)x)}, \tag{B.9}$$

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 $t\to\infty$.

has frequency x and selection is parameterized by coefficient s and dominance parameter h, the three possible genotypes have the following frequencies (assuming Hardy–Weinberg equilibrium) and fitness (Crow and Kimura 1970)

Genotype	AA	Aa	aa
Frequency	x^2	$2x(1-x)$	$(1-x)^2$
Fitness	$1+s$	$1+sh$	1

The allele count $z(r+1)$ still follows the process given in equation (15), with the evolutionary pressure function from equation (B.7).

Dynamics and moments.—The first two moments of the DAF for the general linear evolutionary pressure (equations (B.4) and (B.5)) can be obtained using the law of total expectation and variance, respectively. These take

the form

$$\mathrm{E}[x(r+1)\,|\,x(0)] = \mathrm{E}\big[g(x(r))\,|\,x(0)\big], \tag{16}$$

$$\begin{aligned}\mathrm{Var}\big(x(r+1)\,|\,x(0)\big) &= \frac{1}{2N}\mathrm{E}[x(r+1)\,|\,x(0)] - \mathrm{E}[x(r+1)\,|\,x(0)]^2 \\ &\quad + \left(1-\frac{1}{2N}\right)\mathrm{E}\Big[g(x(r))^2\,|\,x(0)\Big].\end{aligned} \tag{17}$$

The evaluation of $\mathrm{E}\big[g(x(r))\,|\,x(0)\big]$ and $\mathrm{E}\Big[g(x(r))^2\,|\,x(0)\Big]$ 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 $\mathrm{E}\Big[g(x(r))^i\,|\,x(0)\Big]$ by only using the first two moments by relying on a Taylor series. This will yield a recursion for calculating the mean and variance of the DAF. The Taylor series can be evaluated around the deterministic trajectory of $x(r)$ (Barton and Otto 2005; Terhorst et al. 2015), or around the pre-calculated mean of $x(r)$ (Lacerda and Seoighe 2014).

To obtain the Taylor series about the deterministic trajectory, we decompose $x(r)$ as $x(r)=\bar{x}(r)+\tilde{x}(r)$, where $\bar{x}(r)=g(\bar{x}(r-1))$ represents the deterministic trajectory followed by the allele frequency in the infinite-population limit, and $\tilde{x}(r)$ is the random disturbance away from $\bar{x}(r)$. Then,

$$\mathrm{E}[x(r)\,|\,x(0)] = \bar{x}(r) + \mathrm{E}[\tilde{x}(r)\,|\,x(0)], \tag{18}$$

$$\mathrm{Var}\big(x(r)\,|\,x(0)\big) = \mathrm{Var}\big(\tilde{x}(r)\,|\,x(0)\big). \tag{19}$$

From equations (16) and (18) we obtain, using the Taylor series for $\mathrm{E}\big[g(x(r))\,|\,x(0)\big]$ about $\bar{x}(r)$,

$$\begin{aligned}\mathrm{E}[\tilde{x}(r+1)\,|\,x(0)] \approx \mathrm{E}[\tilde{x}(r)\,|\,x(0)]\cdot\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{\bar{x}(r)} \\ + \frac{1}{2}\mathrm{E}\Big[\tilde{x}(r)^2\,|\,x(0)\Big]\cdot\left.\frac{\mathrm{d}^2 g}{\mathrm{d}^2 x}\right|_{\bar{x}(r)}.\end{aligned}$$

Similarly, from the Taylor series of $\mathrm{E}\Big[g(x(r))^2\,|\,x(0)\Big]$ about $\bar{x}(r)$, and using equations (17), (18), and (19) we obtain the recursion for $\mathrm{E}\Big[\tilde{x}(r+1)^2\,|\,x(0)\Big]$,

$$\begin{aligned}&\mathrm{E}\Big[\tilde{x}(r+1)^2\,|\,x(0)\Big] \\ &\approx \frac{1}{2N}\bar{x}(r+1)\big(1-\bar{x}(r+1)\big) \\ &\quad + \frac{1}{2N}\big(1-2\bar{x}(r+1)\big)\mathrm{E}[\tilde{x}(r+1)\,|\,x(0)] \\ &\quad + \left(1-\frac{1}{2N}\right)\mathrm{E}\Big[\tilde{x}(r)^2\,|\,x(0)\Big]\left(\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{\bar{x}(r)}\right)^2.\end{aligned}$$

By iterating the recursions above and calculating numerically the first two moments of $\tilde{x}(r)$, 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 $S=2Ns$, we obtain the following infinitesimal mean

$$a(x) = -\nu x + \mu(1-x) + Sx(1-x)(h+(1-2h)x).$$

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; Ewens 2004) and is, up to a normalization constant, given by a tilted beta distribution

$$f(x) \propto x^{2\nu-1}(1-x)^{2\mu-1}e^{Sx(2h+(1-2h)x)}. \tag{20}$$

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. (2015) 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 ($h=0.5$) and used a Taylor series about the mean of $x(r)$, 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 $r/(2N)=0.5$ 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-allelic. 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 $z(r+1)$ is

$$z(r+1)|z(r)\sim\text{Mult}(2N,x(r)). \tag{21}$$

Here, $\text{Mult}(n,p)$ 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,

$$\text{E}[x(t)\,|\,x(0)]=x(0), \tag{22}$$

$$\begin{aligned}&\text{Var}\big(x(t)\,|\,x(0)\big)\\&\quad=\left(1-e^{-t}\right)\big(\text{diag}\{x(0)\}-x(0)'x(0)\big),\end{aligned} \tag{23}$$

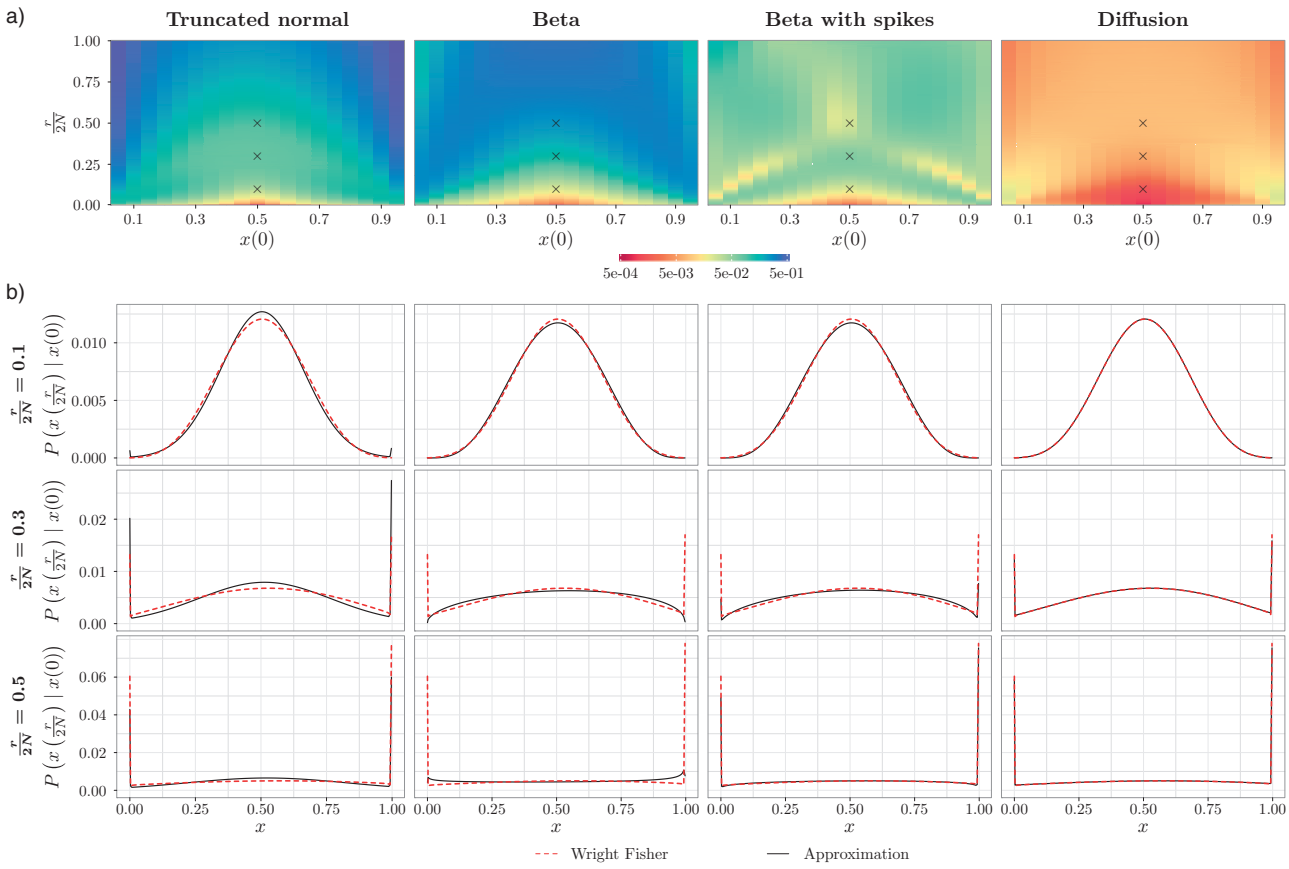

FIGURE 7. Fit of various approximations to the true DAF with selection, calculated using the Markov chain property for $2N=200$, $S=h=1$, $\mu=\nu=0$ and a range of $x(0)$ and $r/(2N)$. Each column shows a different type of approximation, indicated at the top of the figure. a) Hellinger distance on log 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 $r/(2N)$. Calculations are performed as for Figure 4. For comparison purposes, the a) heatmap and b) y-axis scales are the same as in Figure 4.

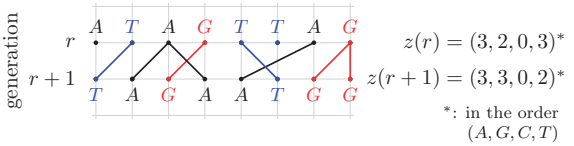

FIGURE 8. Dynamics in the pure drift $K=4$ multi-allelic Wright–Fisher model for (A,G,C,T). The child inherits the parental allele.

where $'$ 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 multi-dimensional 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

$$x(t) \mid x(0) \sim \text{Dirichlet}(\alpha),$$

where $\alpha=(\alpha_1,\ldots,\alpha_K)>0$. This implies that allele $i=1,\ldots,K$ has marginal distribution

$$x_i(t) \sim \text{Beta}(\alpha_i, \alpha_0-\alpha_i), \qquad \text{with } \alpha_0=\sum_{i=1}^{K}\alpha_i.$$

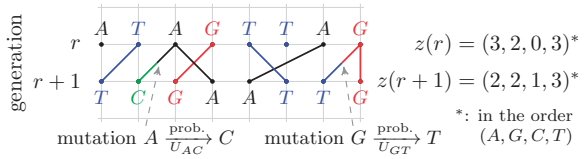

FIGURE 9. Dynamics in the $K=4$ multi-allelic Wright–Fisher model with mutations for (A,G,C,T). If the parental allele is i, the child receives the same allele with probability U_{ii} and another allele j with probability U_{ij}, for $i,j\in\{A,G,C,T\}$.

Under the Dirichlet distribution, the mean and covariance of the DAF are

$$\mathrm{E}[x(t)\,|\,x(0)]=\frac{\alpha}{\alpha_0}, \tag{24}$$

$$\mathrm{Var}\big(x(t)\,|\,x(0)\big) =\frac{1}{\alpha_0+1}\left(\mathrm{diag}\left\{\frac{\alpha}{\alpha_0}\right\}-\left(\frac{\alpha}{\alpha_0}\right)'\frac{\alpha}{\alpha_0}\right). \tag{25}$$

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

$$x(0)=\frac{\alpha}{\alpha_0}, \qquad \text{and} \qquad 1-e^{-t}=\frac{1}{\alpha_0+1}.$$

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_{ij} is the probability of an i allele to mutate to a j allele, the multinomial distribution of $z(r+1)$ becomes

$$z(r+1)\,|\,z(r)\sim\mathrm{Mult}(2N,x(r)U),$$

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 Sirén 2016),

$$\mathrm{E}[x(t)\,|\,x(0)]=x(0)e^{Qt}, \tag{26}$$

$$\begin{aligned}&\mathrm{Var}\big(x(t)\,|\,x(0)\big)\\ &=\int_0^t e^{-s}\left(e^{Qs}\right)'\mathrm{diag}\left\{x(0)e^{Q(t-s)}\right\}\left(e^{Qs}\right)ds\\ &\quad-\left(e^{Qt}\right)'x(0)'x(0)e^{Qt}\left(1-e^{-t}\right).\end{aligned} \tag{27}$$

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 Sirén 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

$$Q_{ij}=2N(U_{ij}-I_{ij})=\begin{cases}\dfrac{q}{K-1} & \text{if } i\neq j\\ -q & \text{if } i=j\end{cases}$$

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

$$e^{Qt}=e^{-\epsilon\frac{t}{2}}\left(I-\frac{E}{K}\right)+\frac{E}{K},$$

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

$$\mathrm{E}[x(t)\,|\,x(0)]=e^{-\epsilon\frac{t}{2}}\left(x(0)-\frac{\mathbf{e}}{K}\right)+\frac{\mathbf{e}}{K}, \tag{28}$$

$$\begin{aligned}&\mathrm{Var}\big(x(t)\,|\,x(0)\big)\\ &=\frac{1}{K}\left(I-\frac{E}{K}\right)\frac{1}{1+\epsilon}\left(1-e^{-(1+\epsilon)t}\right)\\ &\quad-\left(x(0)-\frac{\mathbf{e}}{K}\right)'\left(x(0)-\frac{\mathbf{e}}{K}\right)e^{-\epsilon t}\left(1-e^{-t}\right)\\ &\quad+\left(\mathrm{diag}\left\{x(0)-\frac{\mathbf{e}}{K}\right\}-\left(x(0)-\frac{\mathbf{e}}{K}\right)'\frac{\mathbf{e}}{K}\right.\\ &\qquad\left.-\frac{\mathbf{e}'}{K}\left(x(0)-\frac{\mathbf{e}}{K}\right)\right)\\ &\qquad\times e^{-\epsilon\frac{t}{2}}\frac{1}{1+\frac{\epsilon}{2}}\left(1-e^{-(1+\frac{\epsilon}{2})t}\right),\end{aligned} \tag{29}$$

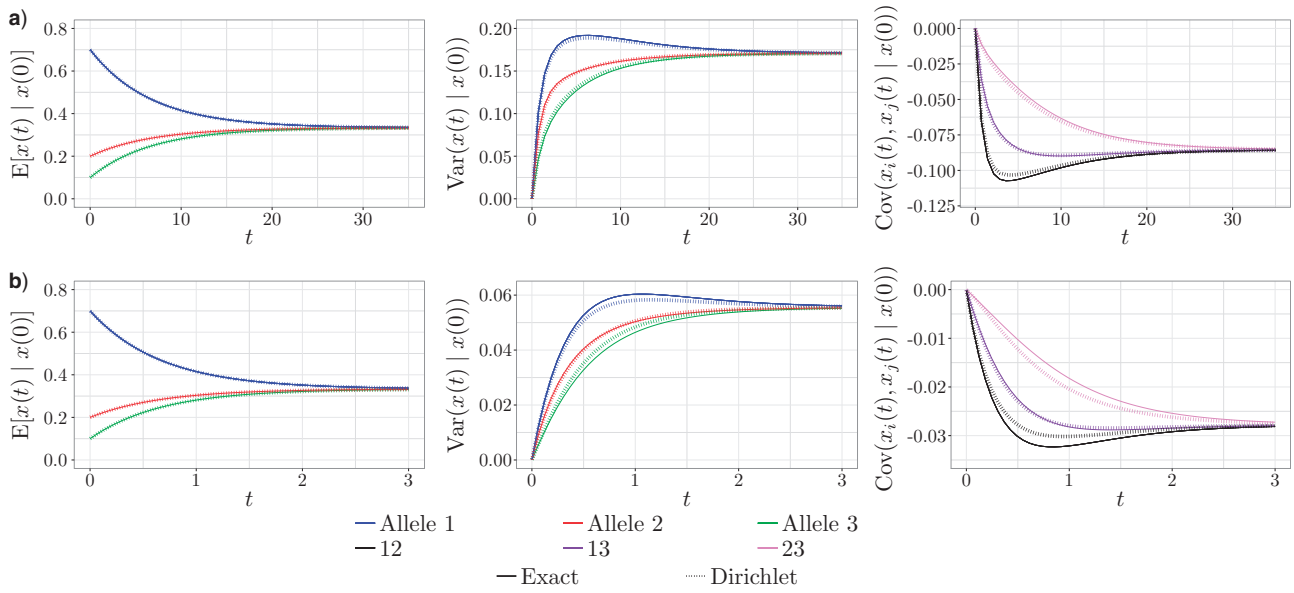

FIGURE 10. Fit of the Dirichlet distribution (dotted lines) to the true mean and covariance of the multi-allelic JC Wright–Fisher model (solid lines) with a) $q=0.1$ ($\epsilon=0.3$ small), and b) $q=1$ ($\epsilon=3$ large). All six plots are calculated for $K=3$, $x(0)=(0.7,0.2,0.1)$, $2N=100$ and different values of t.

where $\mathbf{e}$ is the $1\times K$ vector with 1 in every entry.

For $t\to\infty$, these reduce to

$$\mathrm{E}[x(t)\,|\,x(0)]=\frac{\mathbf{e}}{K},$$
$$\mathrm{Var}\big(x(t)\,|\,x(0)\big)=\frac{1}{K}\left(I-\frac{E}{K}\right)\frac{1}{1+\epsilon}.$$

We note that these moments are the same as for a Dirichlet distribution with $\alpha=\mathbf{e}/(K\epsilon)$, and indeed the Dirichlet distribution is the stationary DAF of the multi-allelic JC 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

$$\begin{aligned}&\mathrm{Var}\big(x(t)\,|\,x(0)\big)\\&\quad\propto \mathrm{diag}\{\mathrm{E[x(t)]}\}-\mathrm{E[x(t)]'E[x(t)]}\\&\quad=\frac{1}{K}\left(I-\frac{E}{K}\right)-\left(x(0)-\frac{\mathbf{e}}{K}\right)'\left(x(0)-\frac{\mathbf{e}}{K}\right)e^{-\epsilon t}\\&\qquad+\left(\mathrm{diag}\left\{x(0)-\frac{\mathbf{e}}{K}\right\}-\left(x(0)-\frac{\mathbf{e}}{K}\right)'\frac{\mathbf{e}}{K}\right.\\&\qquad\left.-\frac{\mathbf{e}'}{K}\left(x(0)-\frac{\mathbf{e}}{K}\right)\right)e^{-\epsilon\frac{t}{2}},\end{aligned}\tag{30}$$

where we used the expression for the mean in equation (28). By comparing equations (29) and (30), we observe that the expressions are approximately proportional with proportionality constant $1-e^{-t}$ when ϵ is small, which corresponds to the pure drift case. Regardless of the parameter ϵ, 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 ϵ (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 Hobolth 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

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 in the limit of a large population size (Lessard and 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 very much a field of its own with active method development in the choice of relevant summary statistics, quasi-sufficiency (Joyce and Marjoram 2008), and various algorithms to obtain computationally efficient approximations of the likelihood (Excoffier et al. 2013). Simulation-based methods can be a source of inspiration for developing methods seeking to specify the DAF under a wider range of scenarios and less restrictive assumptions. Such an example is the inference of selection from time series data from experimental evolution, where the population size undergoes 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.

Funding

This work has been supported, in part, by the European Research Council under the European Unions Seventh Framework Program (FP7/20072013, ERC grant number 311341) and the Danish Research Council (grant number DFF4002-00382).

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. Genetica 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. 2015. 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):2025–2035.

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 allele 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. 6(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 polysomics. Amer. Nat. 88(843):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 polymorphism-aware phylogenetic models. Mol. Biol. Evol. 30(10):2249–2262.
De Maio N., Schrempf 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. 349–358.
Durrett R. 2008. Probability models for DNA sequence evolution. New York: Springer.
Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phyl. Evol. 94:447–462.
Elyashiv E., Sattath S., Hu T.T., Strustovsky A., McVicker G., Andolfatto P., Coop G., Sella G. 2014. A genomic map of the effects of linked selection in Drosophila. arXiv preprint arXiv:1408.5461 [q-bio.PE].
Ewens W.J. 1972. The sampling theory of selectively neutral alleles. Theor. Pop. Biol. 3(1):87–112.
Ewens W.J. 2004. Mathematical Population Genetics 1: I. Theoretical Introduction, Vol 27. Springer Science & Business Media.
Excoffier L., Dupanloup I., Huerta-Sánchez E., Sousa V.C., Foll M. 2013. Robust demographic inference from genomic and snp data. PLOS Gen. 9(10):e1003905.
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.
Ferrer-Admetlla A., Leuenberger C., Jensen J.D., Wegmann D. 2016. An approximate Markova model for the Wright–Fisher diffusion and its application to time series data. Genetics 203(2):831–846.
Fisher R.A. 1930. The genetical theory of natural selection. Oxford: Clarendon.
Foll M. and Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180(2):977–993.
Foll M., Poh Y.-P., Renzette N., Ferrer-Admetlla A., Bank C., Shim H., Malaspinas A.-S., Ewing G., Liu P., Wegmann D., Caffrey D.R., Zeldovich K.B., Bolon D.N., Wang J.P., Kowalik T.F., Schiffer C.A., Finberg R.W., Jensen J.D. 2014. Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Gen. 10(2):e1004185.
Freeman W.H., Wang J., Whitlock M.C. 2003. Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163(1):429–446.
Fu Y.-X., Li W.-H. 1999. Coalescing into the 21st century: an overview and prospects of coalescent theory. Theor. Pop. Biol. 56(1):1–10.
Gaggiotti O.E., Foll M. 2010. Quantifying population structure using the F-model. Mol. Ecol. Res. 10(5):821–830.
Gautier M., Vitalis R. 2013. Inferring population histories using genome-wide allele frequency data. Mol. Biol. Evol. 30(3):654–668.
Gautier M., Hocking T.D., Foulley J.-L. 2010. A Bayesian outlier criterion to detect SNPs under selection in large data sets. PLoS one 5(8):e11913.
Gautier M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics 201(4):1555–1579.
Gompert Z. 2015. Bayesian inference of selection in a heterogeneous environment from genetic timeseries data. Mol. Ecol. 25(1): 121–134
Gutenkunst R.N., Hernandez R.D., Williamson S.H., Bustamante C.D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Gen 5(10):e1000695.
Haasl R.J., Payseur B.A. 2015. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. Mol. Ecol. 25(1):5–23.
Hobolth A., Sirén 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. Schierup M.H. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. PLoS Gen. 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 200(1):285–293.
Jenkins P.A., Fearnhead P., Song Y.S. 2014. Tractable diffusion 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. 41(3):144.
Kimura M. 1955b. Stochastic processes and distribution of gene frequencies under natural selection. In: Cold Spring Harbor Symposia 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. App. Prob. 1(2):177–232.
Kingman J.F.C. 1982a. Exchangeability and the evolution of large populations. In: Koch G., Spizzichino 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. App. Prob. 19A:27–43.
Kingman J.F.C. 1982c. The coalescent. Sto. Proc. Appl. 13(3):235–248.
Kingman J.F.C. 2000. Origins of the coalescent: 1974-1982. Genetics 156(4):1461–1463.
Kolmogorov A. 1931. ber die analytischen methoden in der wahrscheinlichkeitsrechnung. Mathematische Annalen 104:415–458.
Kuhner M.K. 2009. Coalescent genealogy samplers: windows into population history. Tren. Ecol. Evol. 24(2):86–93.
Lacerda M., Seoighe C. 2014. Population genetics inference for longitudinally-sampled mutantsunder strong selection. Genetics 198(3):1237–1250.
Lessard S., Kermany A.R. 2012. Fixation probability in a two-locus model by the ancestral recombination–selection graph. Genetics 190(2):691–707.
Le Cam L.M., Lo Yang G. 2000. Asymptotics in statistics: some basic concepts. 2nd ed. New York: Springer.
Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S.V. 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phyl. Evol. 53(1):320–328.
Liu X., Fu Y.-X. 2015. Exploring population size changes using SNP frequency spectra. Nat. Gen. 47:555–559.
Li H., Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature 475(7357):493–496.
Lukić S., Hey J. 2012. Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-africa expansion. Genetics 192(2):619–639.
Mailund T., Halager A.E., Westergaard M., Dutheil J.Y., Munch K., Andersen L.N., Lunter G., Prüfer K., Scally A., Hobolth A., Schierup M.H. 2012. A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. PLoS Gen. 8(12):e1003125.
Malaspinas A.-S., Malaspinas O., Evans S.N., Slatkin M. 2012. 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):1167–1180.

Mathieson 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.

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

Muirhead 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.

Nielsen R., Slatkin M. 2013. An introduction to population genetics: theory and applications. Sunderland, Massachusetts: Sinauer Associates, Inc.

Paul J.S., Steinrücken M., Song Y.S. 2011. An accurate sequentially Markov conditional sampling distribution for the coalescent with recombination. Genetics 187(4):1115–1128.

Pickrell J.K., Pritchard J.K. 2012. Inference of population splits and mixtures from genomewide allele frequency data. PLoS Gen. 8(11):e1002967.

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

Schiffels S., Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nat. Gen. 46:919–925.

Schraiber J.G. 2014. A path integral formulation of the Wright–Fisher process with genic selection. Theor. Pop. Biol. 92:30–35.

Sheehan S., Harris K., Song Y.S. 2013. Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. Genetics 194(3): 647–662.

Sirén J., Marttinen P., Corander J. 2011. Reconstructing population histories from single nucleotide polymorphism data. Mol. Biol. Evol. 28(1):673–683.

Sirén J. Hanage W.P., Corander J. 2013. Inference on population histories by approximating infinite alleles diffusion. Mol. Biol. Evol. 30(2):457–468.

Sirén J. 2012. Statistical models for inferring the structure and history of populations from genetic data [PhD Thesis] University of Helsinki, Faculty of Science, Department of Mathematics and Statistics Helsinki.

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

Steinrücken M., Bhaskar A., Song Y.S 2014. A novel spectral method for inferring general diploid selection from time series genetic data. Ann. App. Stat. 8(4):2203.

Steinrücken M., Jewett E.M., Song Y.S. 2016. SpectralTDF: transition densities of diffusion processes with time-varying selection parameters, mutation rates, and effective population sizes Bioinformatics 32(5):795–797.

Tataru P., Bataillon T., Hobolth A. 2015. Inference under a Wright–Fisher model using an accurate beta approximation. Genetics 201:1133–1141.

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

Vitalis R., Gautier M., Dawson K.J., Beaumont M.A. 2014. Detecting and measuring selection from gene frequency data. Genetics 196: 799–817.

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 E.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: 97–159.

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):253–259.

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

Wright S. 1945. The differential equation of the distribution of gene frequencies. Proc. Nat. Acad. Sci. USA 31(12):382.

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

Živković D., Steinrücken M., Song Y.S., Stephan W. 2015. Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. Genetics 200:601–617.