Available online at www.sciencedirect.com
ScienceDirect
Theoretical Population Biology 71 (2007) 109–119
Theoretical Population Biology
www.elsevier.com/locate/tpb

Non-equilibrium theory of the allele frequency spectrum

Steven N. Evans[a,1], Yelena Shvets[a,2], Montgomery Slatkin[b,*,3]

[a]Department of Statistics #3860, University of California at Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, USA
[b]Department of Integrative Biology, University of California at Berkeley, Berkeley, CA 94720-3140, USA

Received 13 April 2006
Available online 23 June 2006

Abstract

A forward diffusion equation describing the evolution of the allele frequency spectrum is presented. The influx of mutations is accounted for by imposing a suitable boundary condition. For a Wright–Fisher diffusion with or without selection and varying population size, the boundary condition is $\lim_{x\downarrow 0} xf(x,t) = \theta\rho(t)$, where $f(\cdot,t)$ is the frequency spectrum of derived alleles at independent loci at time t and $\rho(t)$ is the relative population size at time t. When population size and selection intensity are independent of time, the forward equation is equivalent to the backwards diffusion usually used to derive the frequency spectrum, but this approach allows computation of the time dependence of the spectrum both before an equilibrium is attained and when population size and selection intensity vary with time. From the diffusion equation, a set of ordinary differential equations for the moments of $f(\cdot,t)$ is derived and the expected spectrum of a finite sample is expressed in terms of those moments. The use of the forward equation is illustrated by considering neutral and selected alleles in a highly simplified model of human history. For example, it is shown that approximately 30% of the expected total heterozygosity of neutral loci is attributable to mutations that arose since the onset of population growth in roughly the last 150,000 years.

Keywords: Population genetics; Diffusion theory; Entrance law; Kolmogorov forward equation

1. Introduction

The allele frequency spectrum is the distribution of allele frequencies at a large number of equivalent loci. The term "site-frequency spectrum" (Braverman et al., 1995) is equivalent but emphasizes the application to individual nucleotides rather than alleles at different genetic loci. Here, we will use the frequency spectrum for both terms.

Although models assuming reversible mutation predict an equilibrium distribution of allele frequencies (Wright, 1931), all recent studies of frequency spectra assume irreversible mutation. Under that assumption, an equilibrium is not reached at any locus, but the distribution across polymorphic loci reaches an equilibrium if both population size and selection intensities are constant. The theory predicting the frequency spectrum under irreversible mutation was developed by Fisher (1930), Wright (1938), and Kimura (1964). Kimura (1969) noted that this theory was applicable to nucleotide positions and introduced the "infinitely many sites model." Sawyer and Hartl (1992) incorporated the theory of Fisher, Wright and Kimura into a Poisson random field model for the purpose of estimating the selection intensity from the observed frequency spectrum in a finite sample of chromosomes. Their method has been tested and refined by Bustamante et al. (2001) and Williamson et al. (2004).

Past population growth affects the frequency spectrum. Nei et al. (1975) showed that rapid growth resulted in more low-frequency alleles than expected under neutrality. Tajima (1989) confirmed that conclusion and examined the effect of past population growth on other aspects of the frequency spectrum. Griffiths and Tavaré (1998) developed the coalescent theory for the frequency spectrum of neutral alleles in a population that has experienced arbitrary

*Corresponding author. Fax: +1 510 643 6264.
E-mail addresses: evans@stat.Berkeley.edu (S.N. Evans), yelenashvets@yahoo.com (Y. Shvets), slatkin@Berkeley.edu (M. Slatkin).
URLs: http://www.stat.berkeley.edu/users/evans/, http://ib.berkeley.edu/labs/slatkin/.
[1]Supported in part by NSF Grant DMS-0405778.
[2]Supported in part by NSF Grant DMS-0405778.
[3]Supported in part by NIH Grant R01-GM40282.

0040-5809/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2006.06.005

changes in population size. Nielsen (2000) implemented the Griffiths and Tavaré (1998) simulation method and applied it to human SNP data for the purpose of estimating the growth rate of human populations. Although Nielsen (2000) was not able to reject the hypothesis of no growth, he noted that his analysis was of a small data set. Wakeley et al. (2001) considered the same problem and developed a method that allowed for both ascertainment bias and population sub-division. Wakeley et al. (2001) analyzed a larger SNP data set and found evidence of recent growth. Wooding and Rogers (2002) and Polanski and Kimmel (2003) also modeled the coalescent process underlying the spectrum of neutral alleles and developed analytic theory that allows for exact calculation of the spectrum for large sample sizes.

Griffiths (2003) reviewed and extended the theory of the frequency spectrum, making clear the role of time-reversal. He generalized that theory in two ways. He showed that the spectrum in a finite sample could be obtained from the solution to a backwards equation by assuming sampling with replacement, and he showed that the frequency spectrum in a population of variable size could be derived from the spectrum for a population of constant size when a transformation of the time scale reduces the backwards equation to one for a population of constant size. The transformation of time scales is always possible for neutral alleles, in which case the frequency spectrum in a finite population is the same as that derived by Griffiths and Tavaré (1998). For selected alleles, the frequency spectrum cannot be obtained by the method of Griffiths (2003) except in the special case in which the selection intensity is inversely proportional to the population size at all times in the past.

The frequency spectrum of alleles closely linked to selected loci is also of interest. Braverman et al. (1995), Fay et al. (2002), Kim and Stephan (2002), and others have simulated the effects of selected sites on the frequency spectrum of closely linked neutral sites, with the goal of finding evidence of background selection against deleterious mutations and genetic hitchhiking caused by positive selection of advantageous mutations.

Williamson et al. (2005) recently considered the combined effects of population growth and selection on the frequency spectrum. Their model was of a population that was of a constant size until τ generations in the past, at which time it grew instantaneously by a factor ν and remained at the new size until the present. Williamson et al. (2005) developed a likelihood method for estimating both τ and ν from the spectrum of sites assumed to be neutral and for estimating τ, ν and a scaled selection intensity γ for non-neutral sites. Their method was based on numerical solutions for the frequency spectrum using both Kimura (1955) series solution for neutral alleles and numerical solutions to the backwards equation for selected alleles. Williamson et al. (2005) applied their method to a previously published data set of 301 genes in the human genome and found evidence both of population growth and of purifying selection at non-synonymous sites. In a related study, Bustamante et al. (2005) found evidence of differences in selection intensity among different classes of genes in a data set for more than 6000 loci in humans.

In this paper, we will explore in more detail the allele frequency spectrum in a population of variable size. Our goal is similar to that of Williamson et al. (2005) in modeling the combined effects of selection and population growth. We derive the frequency spectrum from the forward equation, first for a Markov chain and then for a diffusion approximation to that Markov chain. The forward equation provides a natural way to compute the spectrum as it approaches an equilibrium from an arbitrary initial condition and to model the time-dependence of the spectrum resulting from the time-dependence of population size and selection intensity. Furthermore, the forward equation provides a way to incorporate the effects of immigration. We show that our approach recovers known results for an equilibrium population and present some numerical results for an idealized model of recent human populations.

2. Markov chain

The model is of a monoecious randomly mating diploid population containing $N(t)$ individuals at time t, which in this section takes integer values representing discrete non-overlapping generations. We assume a large number of identical and independent loci. At each locus there are only two alleles **A**, the derived allele, and **a**, the ancestral allele. In generation t, the set of loci is described by the row vector with jth element $f_j(t)$ that is the expected numbers of loci at which **A** is found on j chromosomes, $1 \leqslant j \leqslant 2N(t)$. Thus, $f_{2N(t)}(t)$ is the expected number of loci fixed for **A** in generation t. The model assumes that the pool of loci fixed for **a** is so large that it can be assumed to be not reduced by the creation of polymorphic loci by mutation—the infinitely many sites model of Kimura (1969).

The change in $f_j(t)$ because of genetic drift and mutation is described by the set of difference equations

$$f_j(t+1) = \sum_{i=1}^{2N(t)} f_i(t) p_{ij}(t) + M_j(t), \quad 1 \leqslant j \leqslant 2N(t+1). \qquad (1)$$

The first term on the right-hand side represents the combined effect of genetic drift and natural selection on loci that are already polymorphic: in the notation of Ewens (2004), $p_{ij}(t)$ is the probability that a locus with i copies of **A** in generation t will have j copies in generation $t+1$. The $p_{ij}(t)$ are easily derived for the Wright–Fisher and other models (Ewens, 2004).

The second term on the right-hand side represents the creation of new polymorphic loci by mutation and immigration. The influx of mutations is modeled by assuming that each of the $2N(t)$ copies of **a** at a monomorphic locus mutates to an **A** with probability μ per generation.

Therefore

$$M_j(t) = 2N(t)\mu\delta_{1,j} \tag{2}$$

for mutation alone, where $\delta_{1,j} = 1$ if $j = 1$ and 0 otherwise. Under the infinitely many sites model, the mutation rate is assumed to be so low that the effect of mutation on loci already polymorphic can be ignored. However, mutation can be incorporated into $p_{ij}(t)$ if necessary.

Immigration from another population can be accounted for both by modifying p_{ij} to allow for the effect of immigration on loci that are already polymorphic and $M_j(t)$ to allow for the creation of new polymorphic loci. Immigration, unlike mutation, can create polymorphic loci for which $j > 1$ immediately. In this paper, we will restrict our analysis to the case with mutation only.

Given an initial frequency spectrum, $f_j(0)$, Eq. (1) can be iterated to obtain the frequency spectrum at any time in the future. Even if N is constant, there is no equilibrium solution to Eq. (1) because the number of loci fixed for **A** will increase each generation. However, an equilibrium solution for f_j, $1 \leqslant i \leqslant 2N-1$, the spectrum for polymorphic loci, does exist and is found by solving the matrix equation

$$\hat{f} = \hat{f}P + \frac{\theta}{2}e, \tag{3}$$

where P is the $(2N-1)\times(2N-1)$ matrix with elements p_{ij} for $1 \leqslant i,\ j < 2N$, e is a row vector with the first element 1 and the remaining elements 0, and $\theta = 4N\mu$. The solution is

$$\hat{f} = \frac{\theta}{2}e(I-P)^{-1}, \tag{4}$$

where I is the $(2N-1)\times(2N-1)$ identity matrix.

This result is equivalent to that obtained using the backwards equation for the Markov chain. The frequency spectrum of polymorphic loci is proportional to the sojourn times ($\tilde{t}_{1,j}$, in the notation of Ewens, 2004). Eq. (4) is obtained from the solution to Eq. (2.143) of Ewens (2004) by multiplying with $\theta/2$, which is the rate of influx of mutations each generation.

The advantage of the formulation presented here is that it also describes the approach to the equilibrium. The rate of approach depends on the second largest eigenvalue of P, which for a neutral allele is $1 - 1/2N$ (Ewens, 2004). The mean number of loci fixed for **A** increases by

$$\sum_{j=1}^{2N-1} f_j(t)p_{j,2N} \tag{5}$$

per generation, which reduces to $2N\mu P_1$ at equilibrium, where P_1 is the probability of fixation of each mutant.

3. Diffusion approximation and new boundary condition

Let us start by considering the time-homogeneous case with no mutation from the ancestral type, but where we can start at time 0 with some derived alleles already present. Because we want to eventually allow varying population sizes, assume that the population is described by a time-homogeneous Markov chain with state-space $\{0, 1, \ldots, 2N\rho\}$. Suppose that if we shrink space by a factor of $2N\rho$ and speed time up by a factor of $2N$, then this chain converges to a diffusion process on $[0,1]$ with generator $\mathscr{G} = a(x)d/dx + \frac{1}{2}b(x)d^2/dx^2$ for appropriate coefficients (this is the scaling regime that is appropriate for models such as the Wright–Fisher chain with or without selection).

Suppose at time 0 that there are countably many loci at which derived alleles are present, with respective (non-random) frequencies $x_1, x_2, \ldots$. Once we have passed to the diffusion limit, the frequency spectrum at time t is just the intensity measure (that is, the expectation measure) of the point process that comes from starting independent copies of the diffusion process at each of the x_i and letting them run to time t. In other words, the intensity measure is obtained by taking the sum of point masses and moving it forwards an amount of time t using the semigroup associated with the generator $\mathscr{G}$.

More generally, if the initial configuration of frequencies is random (so that it can be thought of as a point process on $(0,1)$), then the frequency spectrum at time t is obtained by taking the measure that is the intensity of that point process and again moving it forwards an amount of time t using this semigroup.

For $t > 0$ the resulting measure will be absolutely continuous with respect to Lebesgue measure and have a density $f^0(y,t)$ at frequency $y \in (0,1)$. We will also refer to this density as the frequency spectrum. It is immediate that f^0 satisfies the Kolmogorov forward equation, equations that go with the generator $\mathscr{G}$ (with initial conditions corresponding to the intensity measure of the point process of initial frequencies). That is,

$$\frac{\partial}{\partial t}f^0(y,t) = -\frac{\partial}{\partial y}[a(y)f^0(y,t)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(y)f^0(y,t)], \tag{6}$$

with $\lim_{y\downarrow 0} f^0(y,t)$ and $\lim_{y\uparrow 1} f^0(y,t)$ both finite and appropriate boundary conditions at $t = 0$ (in particular, if the point process of initial frequencies has intensity $h(y)\,dy$, then $\lim_{t\downarrow 0} f^0(y,t) = h(y)$).

Now we want to introduce mutation from the ancestral type as time progresses. In the Markov chain model, this corresponds to new mutants arising in the population at rate $(\theta/2)\rho$ per unit of Markov chain time, where θ is independent of N. The initial number of mutants at a locus is 1. This is equivalent to mutants appearing at rate $2N(\theta/2)\rho$ per unit of rescaled time, with the initial proportion of mutants at a locus being $1/(2N\rho)$.

Imagine now that we pass to the diffusion limit for the allele frequencies, but for the moment still work with a finite N for the description of the appearance of new mutants. That is, we think of our evolving point process as having new points added at location $1/(2N\rho)$ at rate $(\theta/2)(2N\rho)$, and that the locations of these points then evolve as independent copies of the diffusion with generator $\mathscr{G}$.

We will make substantial use of the theory of entrance laws for one-dimensional diffusions laid out in Section 3 of Pitman and Yor (1982). Write $P_t(x,dy)$ for the semigroup associated with $\mathscr{G}$. This is the semigroup of the 0-*diffusion* in the terminology of Pitman and Yor (1982). The contribution to the frequency spectrum from mutations that appear after time 0 is

$$2N\frac{\theta}{2}\rho\int_0^t P_{t-s}\left(\frac{1}{2N\rho},dy\right)ds. \tag{7}$$

Choose a scale function s for the 0-diffusion such that $s(0)=0$ (so that s is then unique up to a positive multiple). As Pitman and Yor (1982) observe,

$$P_u^{\uparrow}(x,dy) := \frac{1}{s(x)}P_u(x,dy)s(y), \quad 0<x,y\leqslant 1 \tag{8}$$

is the semigroup of a diffusion that never hits 0 (this $\uparrow$-*diffusion* is the Doob h-transform that corresponds to the naive idea of conditioning the 0-diffusion never to hit 0). Moreover, this semigroup can be extended to allow starting at 0 by setting

$$P_u^{\uparrow}(0,dy)=\lim_{x\downarrow 0}P_u^{\uparrow}(x,dy). \tag{9}$$

The resulting extended process can start at 0 but it will never return to 0.

Assume now that $s'(0)>0$, which will be the case in the diffusions that are of interest to us. We can choose the free multiplicative constant in the definition of the scale function s so that $\lim_{y\downarrow 0}s(y)/y=s'(0)=1$. Then

$$\lim_{N\to\infty}2N\rho P_u\left(\frac{1}{2N\rho},dy\right)=\frac{P_u^{\uparrow}(0,dy)}{s(y)}=:\lambda_u(dy) \tag{10}$$

in the notation of Pitman and Yor. Thus

$$\begin{aligned}&\lim_{N\to\infty}2N\frac{\theta}{2}\rho\int_0^t P_{t-s}\left(\frac{1}{2N\rho},dy\right)ds\\ &\quad=\frac{\theta}{2}\int_0^t\lambda_{t-s}(dy)\,ds=:\Phi_t(dy),\end{aligned} \tag{11}$$

say.

As observed in Pitman and Yor (1982), the family $(\lambda_u)_{u>0}$ is an entrance law for the semigroup of the 0-diffusion, and so it has densities that satisfy the Kolmogorov forward equation associated with the generator $\mathscr{G}$—intuitively, $(\lambda_u)_{u>0}$ describes that injection of an infinite amount of mass at location 0 at time 0, with this mass subsequently evolving in $(0,1)$ according to the dynamics of the 0-diffusion. Consequently, the family $(\Phi_t)_{t>0}$ also satisfies the Kolmogorov forward equation associated with the generator $\mathscr{G}$—again intuitively, $(\Phi_t)_{t>0}$ describes a continuous-in-time injection of mass at location 0, with this mass again subsequently evolving in $(0,1)$ according to the dynamics of the 0-diffusion. That is, if we write ϕ_t for the density of Φ_t, we have that

$$\frac{\partial}{\partial t}\phi_t(y)=-\frac{\partial}{\partial y}[a(y)\phi_t(y)]+\frac{1}{2}\frac{\partial^2}{\partial y^2}[b(y)\phi_t(y)]. \tag{12}$$

It remains to work out what the boundary conditions for ϕ_t are. Following Pitman and Yor (1982), introduce the $\downarrow$-*diffusion*, which is the 0-diffusion conditioned to hit 0 before 1. The $\downarrow$-diffusion has the Doob h-transform semigroup

$$P_t^{\downarrow}(x,dy)=\left(1-\frac{s(x)}{s(1)}\right)^{-1}P_t(x,dy)\left(1-\frac{s(y)}{s(1)}\right). \tag{13}$$

From Williams (1974), the $\uparrow$-diffusion started at 0 and killed at the last time it visits $y>0$ is the time-reversal of the $\downarrow$-diffusion started at y and killed when it first hits 0. Write $(Q_t^{\downarrow})_{t\geqslant 0}$ for the semigroup of this killed process. Since we have normalized the scale function s so that $s(y)\approx y$ for y close to 0,

$$\begin{aligned}\lim_{y\downarrow 0}y\phi_t(y)&=\lim_{y\downarrow 0}s(y)\phi_t(y)\\ &=\lim_{y\downarrow 0}s(y)\int_0^t\frac{\theta}{2}\frac{\lambda_{t-s}(dy)}{dy}\,ds\\ &=\frac{\theta}{2}\lim_{y\downarrow 0}\int_0^t\frac{P_{t-s}^{\uparrow}(0,dy)}{dy}\,ds\\ &=\frac{\theta}{2}\lim_{y\downarrow 0}\int_0^t\frac{P_{s}^{\uparrow}(0,dy)}{dy}\,ds\\ &=\frac{\theta}{2}\lim_{y\downarrow 0}\int_0^\infty\frac{P_{s}^{\uparrow}(0,dy)}{dy}\,ds\\ &=\frac{\theta}{2}\lim_{y\downarrow 0}\int_0^\infty\frac{Q_{s}^{\downarrow}(y,dy)}{dy}\,ds.\end{aligned} \tag{14}$$

Note that if $(B_t)_{t\geqslant 0}$ is a standard Brownian motion and $T:=\inf\{t\geqslant 0: B_t=0\}$, then, by Eq. (3.2.1) and Section 3.1 of Knight (1981), and

$$\begin{aligned}\int_0^\infty\frac{\mathbb{P}^y\{B_s\in dy, T>s\}}{dy}\,ds&=\int_0^\infty\frac{1}{\sqrt{2\pi s}}-\frac{1}{\sqrt{2\pi s}}e^{-(2y)^2/2s}\,ds\\ &=\lim_{\lambda\downarrow 0}\frac{1}{\sqrt{2\lambda}}-\frac{1}{\sqrt{2\lambda}}\exp(-\sqrt{2\lambda}2y)\\ &=2y.\end{aligned} \tag{15}$$

Observe also that a scale function for the $\downarrow$-diffusion is

$$\sigma(x):=\frac{s(x)s(1)}{s(1)-s(x)}. \tag{16}$$

By standard one-dimensional diffusion theory, if we compose the killed $\downarrow$-diffusion with σ, then the resulting process is a time-change of standard Brownian motion killed when it first hits 0, with the time-change given by the corresponding speed measure (see, for example, V.7 of Rogers and Williams, 2000). Moreover, since $\sigma(x)\sim x$ for x close to 0, the speed measure m for the $\downarrow$-diffusion satisfies $m(dx)\sim dx/b(x)$ for x close to 0 (beware that the definitions of the speed measure can vary from author to author by multiplicative constants, we are using the definition of Rogers and Williams, 2000). Therefore

$$\frac{\theta}{2}\lim_{y\downarrow 0}\int_0^\infty\frac{Q_s^{\downarrow}(y,dy)}{dy}\,ds=\frac{\theta}{2}\lim_{y\downarrow 0}\frac{2y}{b(y)}=\theta\lim_{y\downarrow 0}\frac{y}{b(y)}. \tag{17}$$

Write $f(x,t)$ for the frequency spectrum of the model with mutation from ancestral type. We have $f(x,t) = f^0(x,t) + \phi_t(x)$, where f^0 is defined for the appropriate initial conditions at $t = 0$. If we want to start with all alleles ancestral type, then the initial conditions at $t = 0$ are null and $f^0 \equiv 0$. Combining what we have obtained above, we find that

$$\frac{\partial}{\partial t} f(x,t) = -\frac{\partial}{\partial x}[a(x)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x)f(x,t)] \tag{18}$$

with appropriate boundary conditions at $t = 0$ (in particular, $\lim_{t\downarrow 0} f(x,t) = 0$ if we start will all alleles being ancestral), and further boundary conditions $\lim_{x\downarrow 0} xf(x,t) = \theta \lim_{x\downarrow 0} x/b(x)$ and $\lim_{x\uparrow 1} f(x,t)$ finite.

Now consider a time-inhomogeneous diffusion with generator $a(x,t)d/dx + \frac{1}{2}b(x,t)d^2/dx^2$ and suppose also that ρ is now a function $\rho(t)$ of time. By first considering the case where a, b and ρ are piecewise constant, using the above analysis, and then taking limits, we get that the frequency spectrum solves

$$\frac{\partial}{\partial t} f(y,t) = -\frac{\partial}{\partial y}[a(y,t)f(y,t)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(y,t)f(y,t)] \tag{19}$$

with appropriate boundary conditions at $t = 0$ and further boundary conditions $\lim_{y\downarrow 0} yf(y,t) = \theta \lim_{x\downarrow 0} x/b(x,t)$ and $\lim_{y\uparrow 1} f(y,t)$ finite.

For the purposes of a numerical solution, it is more convenient to consider the function $g(x,t) := x(1-x)f(x,t)$ which satisfies

$$\begin{aligned}\frac{\partial}{\partial t} g(x,t) = & -x(1-x)\frac{\partial}{\partial x}\left[\frac{a(x,t)}{x(1-x)} g(x,t)\right] \\ & + \frac{x(1-x)}{2}\frac{\partial^2}{\partial x^2}\left[\frac{b(x,t)}{x(1-x)} g(x,t)\right]\end{aligned} \tag{20}$$

with appropriate boundary conditions at $t = 0$ and further boundary conditions $\lim_{x\downarrow 0} g(x,t) = \theta \lim_{x\downarrow 0} x/b(x,t)$ and $\lim_{x\uparrow 1} g(x,t) = 0$.

As an example, consider the case where $a(x) = Sx(1-x)$ and $b(x) = x(1-x)/\rho(t)$. This is a Wright–Fisher diffusion with selection and varying population size. The corresponding forward equation is

$$\frac{\partial}{\partial t} g(x,t) = -Sx(1-x)\frac{\partial}{\partial x}[g(x,t)] + \frac{x(1-x)}{2\rho(t)}\frac{\partial^2}{\partial x^2}[g(x,t)] \tag{21}$$

with boundary conditions

$$\lim_{x\downarrow 0} g(x,t) = \theta \lim_{x\downarrow 0} \frac{x\rho(t)}{x(1-x)} = \theta\rho(t). \tag{22}$$

4. Equilibrium solution

When $\rho(t) \equiv 1$ is a constant and the coefficients a and b do not depend on time, then we expect $f(\cdot,t)$ to converge to an equilibrium $\hat{f}$ as $t \to \infty$.

From Eq. (19), the function $\hat{f}$ should satisfy

$$0 = -\frac{\partial}{\partial x}[a(x)\hat{f}(x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x)\hat{f}(x)] \tag{23}$$

with boundary conditions $\lim_{x\downarrow 0} x\hat{f}(x) = \theta \lim_{x\downarrow 0} x/b(x)$ and $\lim_{x\uparrow 1} f(x)$ finite.

In order to solve this equation, suppose first that the diffusion process is on natural scale in $[0,1]$, so that $a \equiv 0$. In that case, Eq. (23) becomes

$$0 = \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(x)\hat{f}(x)], \tag{24}$$

so that $b(x)\hat{f}(x) = c_0 + c_1 x$ for some constants c_0 and c_1. Assume that $0 < \lim_{x\uparrow 1}(1-x)/b(x) < \infty$. Satisfying the boundary conditions requires that $\hat{f}(x) = \theta(1-x)/b(x)$.

Suppose now that the diffusion process is not on natural scale and that s is a scale function with $s(0) = 0$ and $s(1) = 1$, so that the image of the diffusion under s is a diffusion on $[0,1]$ in natural scale. The generator of the image diffusion is

$$\frac{1}{2}[(s' \circ s^{-1})(x)]^2 (b \circ s^{-1})(x)\frac{\partial^2}{\partial x^2}, \tag{25}$$

and hence the image diffusion has the associated frequency spectrum

$$\hat{h}(x) := \frac{\theta(1-x)}{[(s' \circ s^{-1})(x)]^2 (b \circ s^{-1})(x)} \tag{26}$$

from what we have just observed. It follows from the usual change of variable formula for densities that the original diffusion has the associated frequency spectrum

$$\hat{f}(x) = (\hat{h} \circ s)(x)s'(x) = \frac{\theta(1-s(x))}{s'(x)b(x)}. \tag{27}$$

In terms of coefficients,

$$s(x) = \frac{\int_0^x \exp(-2\int_0^y a(z)/b(z)\,dz)\,dy}{\int_0^1 \exp(-2\int_0^y a(z)/b(z)\,dz)\,dy}. \tag{28}$$

Also,

$$\frac{1}{s'(x)b(x)} = \frac{dm}{dx}(x), \tag{29}$$

where m is the speed measure corresponding to the scale measure s (recall that we are using the normalization of the speed measure in Rogers and Williams, 2000). Thus

$$\hat{f}(x) = \theta P_0(x)\frac{dm}{dx}(x), \tag{30}$$

where $P_0(x)$ is the probability the diffusion will be absorbed at 0 given that it starts at x. This well-known equation can be established directly via an ergodic argument using time-reversal—see Griffiths (2003) for a discussion of the history of this technique. Note that our derivation of Eq. (19) also used time-reversal.

For example, when $a = 0$ and $b(x) = x(1-x)$ (the neutral Wright–Fisher diffusion), $\hat{f}(x) = \theta/x$, agreeing with Eq. (9.18) of Ewens (2004). Similarly, when $a(x) = Sx(1-x)$ and $b(x) = x(1-x)$ (the Wright–Fisher diffusion

with selection and no dominance),

$$\hat{f}(x) = \theta \frac{e^{2S}(1 - e^{-2S(1-x)})}{(e^{2S} - 1)x(1-x)}, \tag{31}$$

agreeing with Eq. (9.23) of Ewens (2004).

5. A system of ODEs for the moments in a Wright–Fisher diffusion with varying population size

Suppose in this section that $a(x) = Sx(1-x)$ and $b(x) = x(1-x)/\rho(t)$. This is a Wright–Fisher diffusion model with time-varying population size and constant selection and without dominance. Eq. (21) applies.

Put $\mu_n(t) := \int_0^1 x^n g(x,t)\,dx$ for $n = 0, 1, 2, \ldots$. Integrating by parts, we get

$$\begin{aligned}
&\int_0^1 x^n x(1-x)\frac{\partial}{\partial x} g(x,t)\,dx \\
&\quad = [(x^{n+1} - x^{n+2})g(x,t)]_0^1 \\
&\qquad - \int_0^1 ((n+1)x^n - (n+2)x^{n+1})g(x,t)\,dx \\
&\quad = (n+1)\mu_n - (n+2)\mu_{n+1}(t).
\end{aligned} \tag{32}$$

Similarly,

$$\begin{aligned}
&\int_0^1 x^n x(1-x)\frac{\partial^2}{\partial x^2} g(x,t)\,dx \\
&\quad = \left[(x^{n+1} - x^{n+2})\frac{\partial}{\partial x} g(x,t)\right]_0^1 \\
&\qquad - \int_0^1 ((n+1)x^n - (n+2)x^{n+1})\frac{\partial}{\partial x} g(x,t)\,dx \\
&\quad = -\int_0^1 ((n+1)x^n - (n+2)x^{n+1})\frac{\partial}{\partial x} g(x,t)\,dx \\
&\quad = -[((n+1)x^n - (n+2)x^{n+1})g(x,t)]_0^1 \\
&\qquad + \int_0^1 ((n+1)nx^{n-1}1\{n \neq 0\} - (n+2)(n+1)x^n)g(x,t)\,dx \\
&\quad = [1\{n=0\}\theta\rho(t)] + [(n+1)n\mu_{n-1}(t)1\{n\neq 0\} \\
&\qquad - (n+2)(n+1)\mu_n(t)].
\end{aligned} \tag{33}$$

We thus get the coupled system of ODEs

$$\mu_0'(t) = \frac{\theta}{2} - \frac{1}{\rho(t)}\mu_0(t) + S(\mu_0(t) - 2\mu_1(t)) \tag{34}$$

and

$$\begin{aligned}
\mu_n'(t) = {} & \frac{1}{2\rho(t)}[(n+1)n\mu_{n-1}(t) - (n+2)(n+1)\mu_n(t)] \\
& + S((n+1)\mu_n(t) - (n+2)\mu_{n+1}(t)), \quad n \geqslant 1.
\end{aligned} \tag{35}$$

When $S = 0$ (that is, the neutral case) this system is lower triangular and we can be solved explicitly. The ODE for μ_0 has solution

$$\mu_0(t) = \mu_0(0)\exp\left(-\int_0^t \frac{1}{\rho(s)}\,ds\right) + \frac{\theta}{2}\frac{\int_0^t \exp(\int_0^s (1/\rho(u))\,du)\,ds}{\exp(\int_0^t (1/\rho(s))\,ds)}. \tag{36}$$

Given μ_{n-1}, the ODE for μ_n has solution

$$\begin{aligned}
\mu_n(t) = {} & \mu_n(0)\exp\left(-\binom{n+2}{2}\int_0^t \frac{1}{\rho(s)}\,ds\right) + \binom{n+1}{2} \\
& \times \frac{\int_0^t (1/\rho(s))\mu_{n-1}(s)\exp\left(\binom{n+2}{2}\int_0^s (1/\rho(u))\,du\right)ds}{\exp\left(\binom{n+2}{2}\int_0^t (1/\rho(s))\,ds\right)}.
\end{aligned} \tag{37}$$

We can draw some conclusions from these equations about the effect of ρ on the asymptotic behavior of μ_n. For example, recall that the measure $f(x,t)\,dx$ is the intensity of the point process on $(0,1)$ that records the frequencies of derived alleles at all the loci at which derived alleles are present with non-zero frequency at time t. Hence $2\mu_0(t) = 2\int_0^1 x(1-x)f(x,t)\,dx$ is the expected value of the total of the heterozygosities summed over all loci. For any initial conditions and any ρ such that $\int_0^\infty (1/\rho(t))\,dt < \infty$, the expected total heterozygosity $2\mu_0(t)$ is asymptotically equivalent to θt as $t \to \infty$.

6. Explicit recurrences for the moments in a Wright–Fisher diffusion with exponentially increasing population size

Suppose in this section that $\rho(t) = e^{Rt}$ with $R > 0$, $a(x) = 0$, and $b(x) = x(1-x)/e^{Rt}$. This is a neutral Wright–Fisher diffusion model with constant selection and exponentially increasing population size. We will obtain an explicit recursive recipe for the moments $\mu_n(t)$. For simplicity, suppose that $f(x,0) \equiv 0$, so that $\mu_n = 0$ for all n. A similar development holds for other initial conditions.

Recall that the *exponential integral function* Ei is given by $\mathrm{Ei}(x) = -\int_{-x}^\infty t^{-1}e^{-t}\,dt$, where the principal value is taken if $x > 0$ (although we are only interested in the case $x < 0$). For $x < 0$, $\mathrm{Ei}(x) = -\Gamma(0, -x)$, where Γ is the usual upper incomplete gamma function. For $x > 0$,

$$\begin{aligned}
\mathrm{Ei}(-x) &= \gamma + \log(x) + \int_0^x \frac{e^{-t} - 1}{t}\,dt \\
&= \gamma + \log(x) + \sum_{j=1}^\infty \frac{(-x)^j}{j \cdot j!},
\end{aligned} \tag{38}$$

where γ is Euler's constant (Gradshteyn and Ryzhik, 2000).

For $n = 0$,

$$\mu_0(t) = \frac{\theta}{2R}e^{e^{-Rt}/R}\left(\mathrm{Ei}\left(-\frac{1}{R}\right) - \mathrm{Ei}\left(-\frac{e^{-Rt}}{R}\right)\right). \tag{39}$$

For $n = 1, 2, \ldots$ define a linear operator Φ_n by

$$\Phi_n h(t) := \binom{n+1}{2}\frac{\int_0^t e^{-Rs}h(s)\exp\left(\binom{n+2}{2}\int_0^s e^{-Ru}\,du\right)ds}{\exp\left(\binom{n+2}{2}\int_0^t e^{-Rs}\,ds\right)}, \tag{40}$$

so that $\mu_n = \Phi_n\mu_{n-1} = \cdots = \Phi_n\Phi_{n-1}\cdots\Phi_1\mu_0$. Set

$$h_k(t) = \exp\left(\binom{k+2}{2}\frac{e^{-Rt}}{R}\right) \times \left[\mathrm{Ei}\left(-\binom{k+2}{2}\frac{1}{R}\right) - \mathrm{Ei}\left(-\binom{k+2}{2}\frac{e^{-Rt}}{R}\right)\right], \tag{41}$$

so that $\mu_0 = (1/2R)h_0$. It follows from a straightforward integration that

$$\Phi_n h_k(t) = \frac{\binom{n+1}{2}}{\binom{n+2}{2} - \binom{k+2}{2}}[h_k(t) - h_n(t)]. \tag{42}$$

Hence

$$\mu_n(t) := \sum_{k=0}^{n} c_{n,k}h_k(t), \tag{43}$$

where $c_{0,0} = 1/(2R)$ and the other $c_{n,k}$ are given recursively by

$$c_{n,k} = \frac{\binom{n+1}{2}}{\binom{n+2}{2} - \binom{k+2}{2}}c_{n-1,k}, \quad 1 \leqslant i \leqslant n-1 \tag{44}$$

and

$$\begin{aligned} c_{n,n} &= -\sum_{k=0}^{n-1}\frac{\binom{n+1}{2}}{\binom{n+2}{2} - \binom{k+2}{2}}c_{n-1,k} \\ &= -\sum_{k=0}^{n-1} c_{n,k}. \end{aligned} \tag{45}$$

For example,

$$\begin{aligned} c_{1,0} &= \frac{1}{3-1}\frac{1}{2R} = \frac{1}{4R}, \\ c_{1,1} &= \frac{-1}{4R} \end{aligned} \tag{46}$$

and

$$\begin{aligned} c_{2,0} &= \frac{3}{6-1}\frac{1}{4R} = \frac{3}{20R}, \\ c_{2,1} &= \frac{3}{6-3}\frac{(-1)}{4R} = -\frac{1}{4R}, \\ c_{2,2} &= -\left(\frac{3}{20R} - \frac{1}{4R}\right) = \frac{1}{10R}. \end{aligned} \tag{47}$$

Set $\psi(x) := \sum_{j=1}^{\infty} x^j/(j \cdot j!)$. Then

$$\begin{aligned} h_k(t) &= \exp\left(\binom{k+2}{2}\frac{e^{-Rt}}{R}\right) \\ &\quad \times \left[\mathrm{Ei}\left(-\binom{k+2}{2}\frac{1}{R}\right) - \mathrm{Ei}\left(-\binom{k+2}{2}\frac{e^{-Rt}}{R}\right)\right] \\ &= \exp\left(\binom{k+2}{2}\frac{e^{-Rt}}{R}\right) \\ &\quad \times \left[Rt + \psi\left(-\binom{k+2}{2}\frac{1}{R}\right) - \psi\left(-\binom{k+2}{2}\frac{e^{-Rt}}{R}\right)\right]. \end{aligned} \tag{48}$$

It follows from this that

$$\mu_0(t) \approx \frac{\theta t}{2} \tag{49}$$

when R is large (recall that $2\mu_0(t)$ is the expected total heterozygosity summed over all loci). For $n \geqslant 1$, the two observations that $\exp\left(-\binom{k+2}{2}e^{-Rt}/R\right)$ will be very close to 1 for even moderate values of R and that $\sum_k c_{n,k} = 0$ show that the contribution to $\mu_n(t)$ from the Rt term will almost cancel out and the primary contribution will be from the ψ terms.

7. Frequency spectrum in a finite sample

The function $f(x,t)$ approximates the frequency spectrum in a very large population. In a sample of n chromosomes, we can observe only the finite spectrum, $f_i(t)$, which is the distribution of the number of chromosomes at which there are i derived alleles ($0 < i \leqslant n$). In this context, $f_i(t)$ is similar to $f_i(t)$ defined for the Markov chain formulation, but here t is continuous. The finite spectrum is obtained from $f(x,t)$ by assuming sampling with replacement at each locus independently

$$f_i(t) = \binom{n}{i}\int_0^1 x^i(1-x)^{n-i}f(x,t)\,dx \tag{50}$$

Griffiths (2003). We can express this equation in terms of the moments of $g(x,t) = x(1-x)f(x,t)$ about $x = 0$:

$$f_i(t) = \binom{n}{i}\sum_{j=0}^{n-i-1}(-1)^j\binom{n-i-1}{j}\mu_{j+i-1}(t). \tag{51}$$

This expression involves an alternating sum and so, from a numerical point of view, it might be preferable to have a system of ODEs for the f_i themselves rather than for the μ_i. Unfortunately, a system of ODEs for the f_i appears to be rather complicated: it is doubly indexed by i and the suppressed index n and, moreover, the alternation present in Eq. (51) is just "transferred" to the coefficients of the ODEs.

For use later in Section 9 we make the following small observation. Suppose that $x \mapsto f(x,t)$ is decreasing. Observe

for $1 \leqslant i \leqslant n-1$ that, by an integration by parts,

$$\begin{aligned} f_{i+1}(t) - f_i(t) &= \frac{n!}{(i+1)!(n-i)!}\int_0^1 [x^{i+1}(n-i)(1-x)^{n-i-1} \\ &\quad - (i+1)x^i(1-x)^{n-i}]f(x,t)\,dx \\ &= -\frac{n!}{(i+1)!(n-i)!}\int_0^1 \frac{\partial}{\partial x}[x^{i+1}(1-x)^{n-i}]f(x,t)\,dx \\ &= \frac{n!}{(i+1)!(n-i)!}\int_0^1 x^{i+1}(1-x)^{n-i}\,\frac{\partial}{\partial x}f(x,t)\,dx \\ &< 0 \end{aligned} \tag{52}$$

(there are no boundary terms because of the boundary conditions on $f(\cdot,t)$). Thus, $f_i(t)$ is decreasing in i when $x \mapsto f(x,t)$ is decreasing.

8. Numerical analysis

If the population size or selection intensity vary only somewhat with time, numerical solutions to Eq. (21) can be obtained with standard methods. The function `NDSolve` in *Mathematica* (version 5) and the function `pdepe` of *MATLAB* (version 7) both provide solutions using default settings for those programs. With extreme population growth, as in the model of human history we consider in Section 9, neither of these programs provides accurate solutions, so it was necessary to write a program tailored to the problem.

Large gradients at the boundary $x = 0$ make it necessary to introduce a non-uniform grid with N internal points. With the view toward integrating the numerical solution in order to compute moments and the finite spectrum we select the grid with the smallest spatial increment $\Delta_0 = (x_1 - 0)$ of the order 10^{-8} and the largest $\Delta_N = (1 - x_N) \sim 10^{-3}$. The spacing is kept constant for a few nodes $x_i, \ldots, x_{i+k}$ and then doubled so that

$$x_{i+k+1} - x_{i+k} = \Delta_{i+k} = 2\Delta_{i+k-1} = 2(x_{i+k} - x_{i+k-1}).$$

This process is repeated until the spacing reaches the maximum size $\Delta_N \sim 10^{-3}$. The numerical domain is thus separated into sub-domains within each of which the spacing is uniform. This guarantees that centered difference schemes used in the sub-domains give a second-order accurate approximations to the corresponding differential operators. For the right end-point of the uniform sub-domains, x_{i+k}, (referred to as the edge-node) the values at x_{i+k-2} and x_{i+k+1} can be used, since $x_{i+k} - x_{i+k-2} = x_{i+k+1} - x_{i+k}$.

If $t = \Delta_t k$ and $x = x_i$ we denote the numerical approximation to $g(x,t)$ by G_i^k. A standard centered second-order differencing is used to approximate the diffusive term and one-sided first-order difference with the direction depending on the sign of S (so-called *upwinding*) is used for the advective term. For the time-stepping an implicit backward-Euler scheme with a fixed step-size $\Delta_t = \Delta_N$ is used. The following system of algebraic equations is obtained:

$$\begin{aligned} G_i^{k+1} &= G_i^k + \Delta_t\left(\frac{x_i(1-x_i)}{2\rho(t)(\Delta_i)^2}[G_{i-Edge}^{k+1} - 2G_i^{k+1} + G_{i+1}^{k+1}]\right) \\ &\quad - S\Delta_t\left(\frac{x_i(1-x_i)}{\Delta_i}[G_i^{k+UwR} - G_i^{k+UwL}]\right), \end{aligned} \tag{53}$$

where $Edge = 2$ if x_i is edge-node and $Edge = 1$ otherwise; $UwR = 0, UwL = -1$ if $S > 0$ and $UwR = 1$, $UwL = 0$ if $S < 0$. The boundary conditions are $G_0^k = g(\Delta_t * k)$ and $G_{N+1}^k = 0$.

Overall the truncation error is of the order Δ_N. In each time step, in order to solve the system (53) it is necessary to invert a sparse matrix with a large condition number ($cN \approx 10^8$), hence a gain in accuracy that would come from decreasing the time step is offset by the loss of precision in the inversion of such a matrix.

9. Model of recent human population growth

We considered a highly simplified model of the history of population sizes of modern humans similar to that used by Reich and Lander (2001) and Williamson et al. (2005) but not requiring the assumption of an instantaneous change in population size. We assumed a stable population containing $N_0 = 10,000$ individuals until 150,000 years ago, which we took as time $t = 0$. We assumed a generation time of 25 years. We measured time in units of $2N_0$, so the present is at $t = 6000/20,000 = 0.3$. At $t = 0$ the population began to increase in size exponentially at a scaled rate $R = 2N_0 r$, where r is the exponential rate per generation. We assumed additive selection with heterozygous fitness $1 + s$ relative to **aa** homozygotes, with the selection coefficient is also scaled by N_0: $S = 2N_0 s$. The value $R = 40$ corresponds to a current size of 1.63×10^9. Reich and Lander assumed a current size of 6×10^9, but that did not take account of the

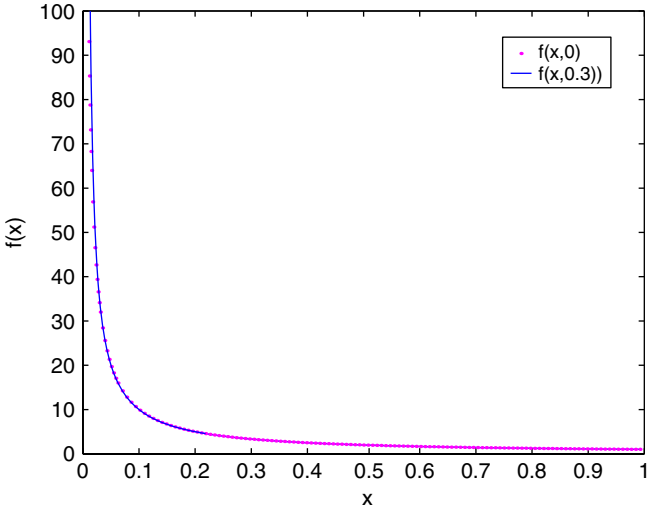

Fig. 1. Frequency spectrum $f(x,t) = g(x,t)/(x(1-x))$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = 0$. Obtained by numerically integrating the PDE. The values of f are restricted to the interval $[0, 100]$.

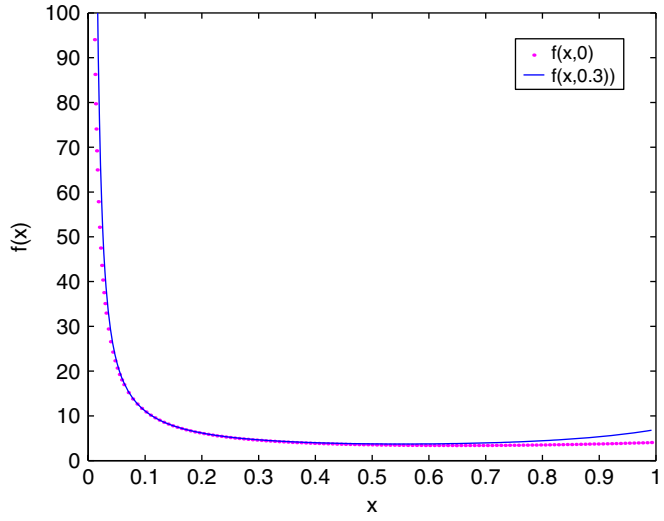

Fig. 2. Frequency spectrum $f(x,t) = g(x,t)/(x(1-x))$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = +2$. Obtained by numerically integrating the PDE. The values of f are restricted to the interval $[0, 100]$.

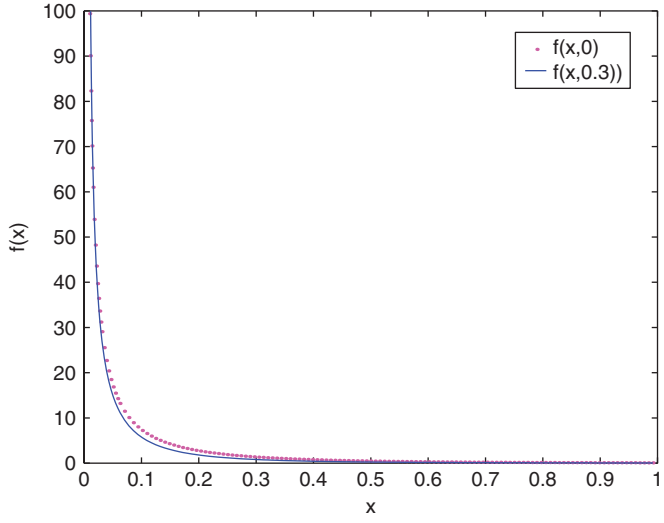

Fig. 3. Frequency $f(x,t) = g(x,t)/(x(1-x))$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = -2$. Obtained by numerically integrating the PDE. The values of f are restricted to the interval $[0, 100]$.

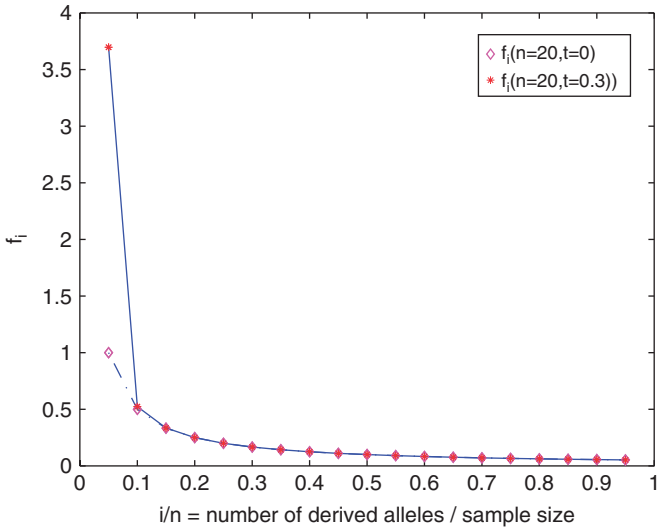

Fig. 4. Finite spectrum for a sample size $n = 20$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = 0$. Obtained by numerically integrating the system of ODEs for the moments of $g(\cdot, t)$.

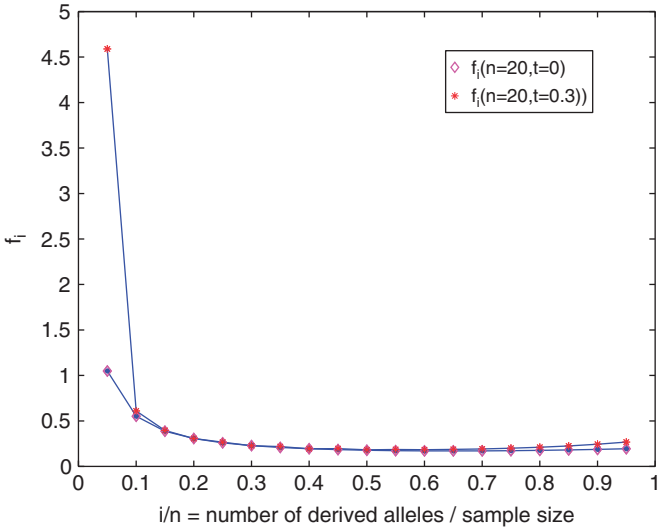

Fig. 5. Finite spectrum for a sample size $n = 20$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = +2$. Obtained by numerically integrating the system of ODEs for the moments of $g(\cdot, t)$.

fact that the effective size of human populations is roughly $\frac{1}{3}$ of the census size (Hill, 1972).

We assume that the spectrum at $t = 0$ is the equilibrium spectrum:

$$g(x,0) = \theta(1-x) \tag{54}$$

for $S = 0$ and

$$g(x,0) = \theta \frac{e^{2S}\left(1 - e^{-2S(1-x)}\right)}{e^{2S} - 1} \tag{55}$$

otherwise. The numerical solutions for $f(x,t)$ at times $t = 0$ and 0.3 are plotted in Figs. 1–3 for the respective choices $0, +2, -2$ of the selection parameter S (the mutation parameter is taken to be $\theta = 1$—a different choice of θ merely rescales the spectrum).

For neutral alleles, the equation for the 0th moment of $g(x,t)$, which is half the expected total heterozygozity, can be solved exactly. It is of interest to separate the solution into two parts, one representing alleles present at $t = 0$ (old alleles) and the other representing alleles that arose by mutation after $t = 0$ (new alleles). For old alleles,

$$\frac{d\mu_{0,o}}{dt} = -e^{-Rt}\mu_{0,o} \tag{56}$$

with initial condition $\mu_{0,o}(0) = \theta/2$ and for new alleles

$$\frac{d\mu_{0,n}}{dt} = \frac{\theta}{2} - e^{-Rt}\mu_{0,n} \tag{57}$$

with initial condition $\mu_{0,n}(0) = 0$. Eqs. (56) and (57) can be solved in terms of exponential integrals. With $R = 40$, we

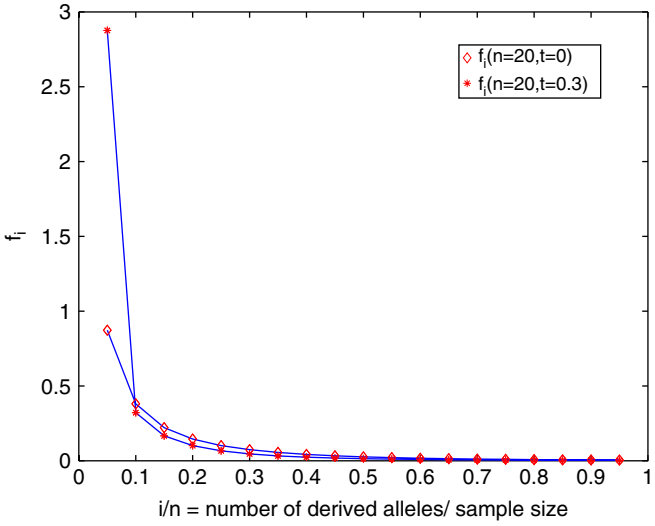

Fig. 6. Finite spectrum for a sample size $n = 20$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = -2$. Obtained by numerically integrating the system of ODEs for the moments of $g(\cdot, t)$.

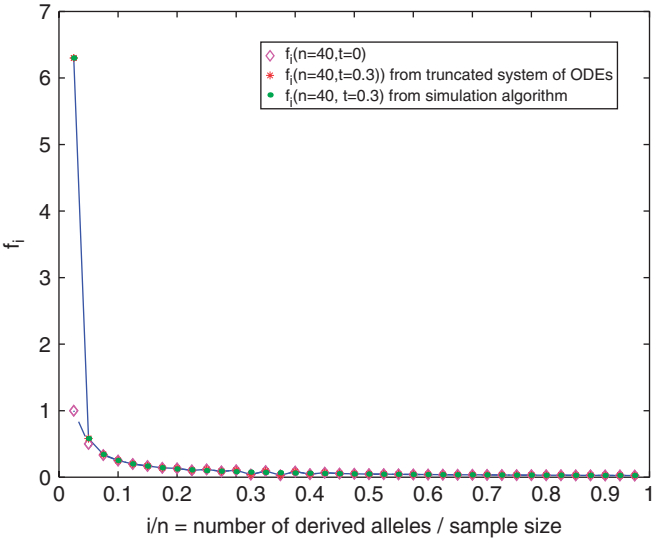

Fig. 7. Finite spectrum for a sample size $n = 40$ at times $t = 0$ and 0.3 with parameter values $R = 40$ and $S = 0$. Obtained by numerically integrating the system of ODEs for the moments of $g(\cdot, t)$. Compared at time $t = 0.3$ with results from the simulation algorithm of Griffiths and Tavaré.

find $\mu_{0,o}(0.3) = 0.49\theta$ and $\mu_{0,n}(0.3) = 0.15\theta$, which implies that under this model, roughly 30% of the expected heterozygosity at neutral sites is attributable to mutations that arose in the past 150,000 years.

For selected alleles ($S \neq 0$), the system for the moments is not closed. An approximation to the moments can be obtained by truncating the system and setting the first neglected term to its initial value. Alternatively, the moments can be computed by first solving for $g(x, t)$ and numerically integrating. Both approaches present certain numerical difficulties. The plots in Figs. 4–6 were obtained for a sample of size $n = 20$ by numerically solving the truncated system with 160 equations using *MATLAB*'s `ode45` (for $S = 0$) and `ode15s` (for $S = \pm 2$) routines. For the neutral case $S = 0$ there is a simulation algorithm due to Griffiths and Tavaré (1998) for approximating the finite spectrum, and we compare that approximation with the results from the numerical solution of the system of ODEs in Fig. 7 for a sample of size $n = 40$. The mutation parameter for Figs. 4–7 is taken to be $\theta = 1$. Again, a different choice of θ merely rescales the finite spectrum. It appears from Fig. 1 that the frequency spectrum at time $t = 0.3$ is a decreasing function and hence, from the remark at the end of Section 7, the corresponding finite spectrum plotted in Fig. 7 should also be decreasing. The undulations in the plot are therefore numerical artifacts.

References

Braverman, J., Hudson, R.R., Kaplan, N.L., Langley, C.H., Stephan, W., 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140, 783–796.

Bustamante, C.D., Wakeley, J., Sawyer, S., Hartl, D.L., 2001. Directional selection and the site-frequency spectrum. Genetics 159, 1779–1788.

Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Todd Hubisz, M., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M., Clark, A.G., 2005. Natural selection on protein-coding genes in the human genome. Nature 437, 1153–1157.

Ewens, W.J., 2004. Mathematical Population Genetics: I. Theoretical Introduction, second ed. Interdisciplinary Applied Mathematics. Springer, Berlin.

Fay, J.C., Wyckoff, G.J., Wu, C.-I., 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature 415, 1024–1026.

Fisher, R.A., 1930. The distribution of gene ratios for rare mutations. Proc. R. Soc. Edinburgh 50, 205–220.

Gradshteyn, I.S., Ryzhik, I.M., 2000. Table of Integrals, Series, and Products, sixth ed. Academic Press Inc., San Diego, CA translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger.

Griffiths, R.C., 2003. The frequency spectrum of a mutation, and its age, in a general diffusion model. Theor. Popul. Biol. 64, 241–251.

Griffiths, R.C., Tavaré, S., 1998. The age of a mutation in a general coalescent tree. Stochastic Models 14, 273–295.

Hill, W.G., 1972. Effective size of populations with overlapping generations. Theor. Popul. Biol. 3, 278–289.

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

Kimura, M., 1955. Solution of a process of random genetic drift with a continuous model. Proc. Natl. Acad. Sci. USA 41, 144–150.

Kimura, M., 1964. Diffusion models in population genetics. J. Appl. Probab. 1, 177–232.

Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61, 893–903.

Knight, F.B., 1981. Essentials of Brownian Motion and Diffusion. Mathematical Surveys, vol. 18. American Mathematical Society, Providence, RI.

Nei, M., Maruyama, T., Chakraborty, R., 1975. The bottleneck effect and genetic variability in populations. Evolution 29, 1–10.

Nielsen, R., 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154, 931–942.

Pitman, J., Yor, M., 1982. A decomposition of Bessel bridges. Z. Wahrsch. Verw. Gebiete 59, 425–457.

Polanski, A., Kimmel, M., 2003. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. Genetics 165, 427–436.

Reich, D.E., Lander, E.S., 2001. On the allelic spectrum of human disease. Trends Genet. 17, 502–510.

Rogers, L.C.G., Williams, D., 2000. Diffusions, Markov processes, and martingales, Vol. 2. Cambridge Mathematical Library, Cambridge University Press, Cambridge (reprint of the second (1994) edition).

Sawyer, S.A., Hartl, D.L., 1992. Population genetics of polymorphism and divergence. Genetics 132, 1161–1176.

Tajima, F., 1989. The effect of change in population size on DNA polymorphism. Genetics 123, 597–602.

Wakeley, J., Nielsen, R., Liu-Cordero, S.N., Ardlie, K., 2001. The discovery of single-nucleotide polymorphisms: and inferences about human demographic history. Am. J. Hum. Genet. 69, 1332–1347.

Williams, D., 1974. Path decomposition and continuity of local time for one-dimensional diffusions. I. Proc. London Math. Soc. 28, 738–768.

Williamson, S., Fledel-Alon, A., Bustamante, C.D., 2004. Population genetics of polymorphism and divergence for diploid selection models with arbitrary dominance. Genetics 168, 463–475.

Williamson, S.H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R., Bustamante, C.D., 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc. Natl. Acad. Sci. USA 102, 7882–7887.

Wooding, S., Rogers, A., 2002. The matrix coalescent and an application to human single-nucleotide polymorphisms. Genetics 161, 1641–1650.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

Wright, S., 1938. The distribution of gene frequencies under irreversible mutation. Proc. Natl. Acad. Sci. USA 24, 253–259.