Theoretical Population Biology 80 (2011) 80–99

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Generalized population models and the nature of genetic drift

Ricky Der [a,b,*], Charles L. Epstein [a], Joshua B. Plotkin [b]

[a] Department of Mathematics, University of Pennsylvania, Philadelphia, United States

[b] Department of Biology, University of Pennsylvania, Philadelphia, United States

ARTICLE INFO

Article history:
Received 14 September 2010
Available online 28 June 2011

Keywords:
Wright–Fisher model
Cannings model
Λ-Fleming–Viot process
Λ-coalescent
Selection amplification
Genetic drift

ABSTRACT

The Wright–Fisher model of allele dynamics forms the basis for most theoretical and applied research in population genetics. Our understanding of genetic drift, and its role in suppressing the deterministic forces of Darwinian selection has relied on the specific form of sampling inherent to the Wright–Fisher model and its diffusion limit. Here we introduce and analyze a broad class of forward-time population models that share the same mean and variance as the Wright–Fisher model, but may otherwise differ. The proposed class unifies and further generalizes a number of population-genetic processes of recent interest, including the Λ and Cannings processes. Even though these models all have the same variance effective population size, they encode a rich diversity of alternative forms of genetic drift, with significant consequences for allele dynamics. We characterize in detail the behavior of standard population-genetic quantities across this family of generalized models. Some quantities, such as heterozygosity, remain unchanged; but others, such as neutral absorption times and fixation probabilities under selection, deviate by orders of magnitude from the Wright–Fisher model. We show that generalized population models can produce startling phenomena that differ qualitatively from classical behavior — such as assured fixation of a new mutant despite the presence of genetic drift. We derive the forward-time continuum limits of the generalized processes, analogous to Kimura's diffusion limit of the Wright–Fisher process, and we discuss their relationships to the Kingman and non-Kingman coalescents. Finally, we demonstrate that some non-diffusive, generalized models are more likely, in certain respects, than the Wright–Fisher model itself, given empirical data from Drosophila populations.

1. Introduction

Mathematical descriptions of allele dynamics are typically built upon the Wright–Fisher model (Fisher, 1958; Wright, 1931) or its diffusion limit (Kimura, 1962; Ewens, 2004). Most theoretical and applied research rests implicitly on this framework, including Kimura's work on fixation probabilities (Kimura, 1955), Ewens' sampling formula (Ewens, 1972; Lessard, 2007), Kingman's coalescent (Kingman, 1982), tests of neutrality (Hudson et al., 1987; Tajima, 1989; McDonald and Kreitman, 1991; Fu and Li, 1992; Fay and Wu, 2000), and techniques for inferring mutation rates and selection pressures (Sawyer and Hartl, 1992; Yang and Bielawski, 2000; Bustamante et al., 2001). Moreover, the development of the neutral theory of molecular evolution, and our understanding of the role of genetic drift in suppressing the deterministic forces of Darwinian selection, has relied on the specific form of the Wright–Fisher model and its diffusion limit.

Kimura's diffusion is remarkably robust to modifications of the model's underlying assumptions. Indeed, many population-genetic models, including the Moran process (Moran, 1958), and a number of Karlin–Taylor and Cannings processes (Karlin and McGregor, 1964; Cannings, 1974; Ewens, 2004), share the same diffusion limit as the Wright–Fisher model (Möhle, 2001). As a result, Kimura's diffusion has had an enormous impact on the development of theoretical and applied population genetics.

Despite its robustness, the Wright–Fisher model (or its diffusion limit) is not appropriate in all circumstances. The Wright–Fisher model is founded on a set of assumptions about how organisms reproduce, and the way in which the constraint of finite-population size enforces dependencies among individuals. Mathematically, in its simplest form, the process is defined as a time-homogeneous Markov chain describing the evolution of the frequencies of two neutral, non-mutating alleles. Suppose X_k denotes the number of type-one individuals in generation k among a population of fixed size N. Under the Wright–Fisher model the transition matrix $\mathbf{P}_{ij}$ for the probability of j alleles in generation $k+1$, given i alleles in generation k is

$$\mathbf{P}_{ij} = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}. \tag{1}$$

The specific, binomial form of this transition matrix suggests an underlying mode of reproduction in which individuals each

* Corresponding address: University of Pennsylvania, Lynch Laboratories, Philadelphia, PA 19104, United States.
E-mail address: rickyder@sas.upenn.edu (R. Der).

0040-5809/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2011.06.004

produce an extremely large number of gametes, from which N are then sampled to form the subsequent generation. This implicit mechanism of reproduction provides a good description of some biological organisms, such as semelparous wildflowers, but, it may not be suitable for all organisms. For example, the reproductive mode of many pelagic organisms (Beckenbach, 1994; Hedgecock, 1994), including oysters, allows for the possibility that the offspring of a single individual may form a substantial fraction of the entire population in a single generation — a possibility of negligible probability in the binomial model.

Given the specific mode of reproduction inherent to the Wright–Fisher model, population geneticists must resolve a fundamental question: how do the predictions and inferences of the Wright–Fisher model change when its underlying assumptions are relaxed or generalized to accommodate a larger range of biological situations? Recent developments in coalescent theory (Pitman, 1999; Schweinsberg, 2003; Sagitov, 1999) have uncovered a diversity of biologically relevant models in the backward-time framework, of which the Wright–Fisher model or, equivalently, Kingman's coalescent serves as only a single example. While much work remains in exploring the implications and applications of these generalized models, the initial task of mathematically classifying such coalescents is mostly complete (Möhle and Sagitov, 2001), in the neutral, exchangeable case. In this paper, by contrast, we define and systematically analyze the forward-time population genetics of a large class of models that generalize the Wright–Fisher model, both with and without selection.

The forward-time equivalents for some non-Kingman coalescents have already been defined and studied, in an abstract setting. In particular, much is known about the so-called Λ and Ξ-Fleming–Viot processes, a class of generalizations of the Wright–Fisher process, especially concerning the duality between the forward and backward processes and their relation to various measure-valued processes (Donnelly and Kurtz, 1999b,a; Birkner and Blath, 2009). These results have been derived in the very general setting of an infinite type space; nonetheless, even in the simplest case of two genetic types, there has been no systematic analysis of the standard forward-time population-genetic quantities for such processes—such as the expected time to fixation of an allele, or the fixation probability of an allele under selection. In this paper we provide an analytic treatment of such quantities, and we investigate how they behave across a broad class of processes that encompasses the two-type Λ- and Ξ-processes.

The set of models we explore is large, but constrained to match the first- and second-order moments of the Wright–Fisher process (see below). We call this set of models the Generalized Wright–Fisher (GWF) processes. This broad definition includes the so-called Cannings exchangeable models (Cannings, 1974) as a strict subset. Nevertheless, in contrast to the approach of Cannings, and more generally any particle-based representation, we specify the space of GWF models in terms of conditions on the allele frequency dynamics, without specifying any lineage information. This change in focus allows us to analyze the properties of Cannings processes with greater ease than previously possible, even in the presence of selection. As it turns out, many probabilistic properties of Cannings processes are shared by the larger class of GWF models we study, and indeed they are more easily appreciated in this broader context.

Since the first- and second-order moments of a process completely define Kimura's diffusion and determine the variance effective population size, one might naively expect that all GWF processes will have the same continuum limit as the Wright–Fisher does. This is false. Subtle differences between processes, associated with their higher moments, can lead to radically different limiting behaviors. We will show that some population-genetic quantities (such as the expected absorption time of an allele initiated at an intermediate frequency) are robust or even impervious to violations of the Wright–Fisher assumptions; but other quantities can differ dramatically from the standard model. When selection is included, we discover the interesting general fact that most GWF processes tend to *amplify* selective effects, relative to the standard Wright–Fisher model. This amplification leads to some startling new population-genetic phenomena, including the possibility of *assured* fixation for an advantageous allele, despite the presence of non-trivial random drift. Finally, by comparison to empirical data on allele frequency dynamics from laboratory populations of fruit flies, we demonstrate that some of the non-standard population models we study are equally likely or, in some respects, more likely than the Wright–Fisher model.

Roughly speaking, by fixing the first and second moments of a GWF process we are imposing the constraint of neutrality and a choice of time scale on the evolution of the population. Given these moments, one can imagine two directions of deformation to the transition probabilities of the process: the third and higher moments may be increased relative to the second moment, or they may be decreased. According to classical diffusion theory, deformations decreasing the higher moments will improve the rate of convergence of the process to the classical Kimura diffusion; qualitatively different results can therefore only occur when the higher moments are enlargened. This simple intuition will lead to the viewpoint that the Wright–Fisher model, in many of its properties, is extreme, compared to the set of models with the same first and second moments.

One of our primary goals in this work is to deepen our understanding of genetic drift by exploring alternatives to the binomial drift inherent to the Wright–Fisher model. Although there has been much work on generalizing other aspects of population-genetic models, by including demographic variation or frequency-dependent selection, for example, there has not been widespread recognition that the nature of drift itself is conceivably more important than any of these other forces. Population geneticists typically characterize genetic drift by a single number – the variance effective population size – which simply scales all genetic quantities. As a result, there is widespread belief among many biologists that there is only a single reasonable model of genetic drift. Our analysis of generalized population processes reveals a rich diversity of alternative forms of genetic drift, with significant consequences for the dynamics of alleles in replicating populations.

2. Generalized population processes

The Wright–Fisher model describes the dynamics of two competing types in a population of fixed size, N. The model is formulated as a discrete-time Markov chain. Suppose X_k denotes the number of individuals of type one in generation k. The first two conditional moments of the Wright–Fisher process satisfy

$$E[X_{k+1}|X_k] = X_k \tag{2}$$

$$\mathrm{Var}[X_{k+1}|X_k] = X_k\left(1 - \frac{X_k}{N}\right). \tag{3}$$

Qualitatively, the condition on the mean (2) expresses neutrality, while the condition on the variance (3) is a gross, second-order characterization of the strength of genetic drift. Drift is strongest when an allele is at intermediate frequencies.

Cannings (1974) introduced a large family of general population processes, subsuming the Wright–Fisher model, centered around the concept of *exchangeability*—a type of invariance under relabeling of individuals. The first two conditional moments of any Cannings process coincide with those of the Wright–Fisher model (2)

and (3), up to a scaling factor. Viewed in this light, the specific linear and quadratic forms of these moments are natural characteristics of population processes. Taken together, these two conditions comprise a minimal set of desiderata for any neutral population process.

Motivated by these considerations, we define a larger class of population processes: those that coincide with the Wright–Fisher (and Cannings) models in their first two conditional moments.

Definition 1. A *pure-drift Generalized Wright–Fisher (GWF) Process* in a population of constant size N is a time-homogeneous Markov chain, $\{X_k, k = 1, 2, \ldots\}$, on states $\{0, \ldots, N\}$ whose conditional moments satisfy

$$E[X_{k+1}|X_k] = X_k \tag{4}$$

$$\mathrm{Var}[X_{k+1}|X_k] = \frac{N\sigma_N^2}{N-1} X_k(1 - X_k/N) \tag{5}$$

where $\sigma_N^2 = \mathrm{Var}(X_{k+1}|\{X_k = 1\})$.

Thus, a GWF process matches the first and second conditional moments of the Wright–Fisher process, up to a scaling constant, but it might otherwise differ in its higher moments. Loosely speaking, such processes describe alternative forms of genetic drift that are roughly similar to the binomial drift encoded by the Wright–Fisher model.

The scaling constant above, $\sigma_N^2 = \mathrm{Var}(X_{k+1}|\{X_k = 1\})$, is called the (finite-population) *offspring variance*. In the standard Wright–Fisher model $\sigma_N^2 = 1 - 1/N$. The offspring variance determines the so-called variance effective population size, so that $N_e = N/\sigma_N^2$. More generally, we call the conditional distribution $X_{k+1}|\{X_k = 1\}$ the (finite-population) *offspring distribution* for the process.

If we let $\mathbf{Q}$ denote the transition matrix of a pure-drift GWF process, so that $\mathbf{Q}_{ij} = P(X_{k+1} = j|X_k = i)$, then the above moment conditions are equivalent to

$$\sum_{j=0}^{N} j\mathbf{Q}_{ij} = i \tag{6}$$

$$\sum_{j=0}^{N} (j-i)^2\mathbf{Q}_{ij} = \frac{N\sigma_N^2}{N-1} i\left(1 - \frac{i}{N}\right). \tag{7}$$

In this notation, the first row $\mathbf{Q}_{1,j}$ corresponds to the (finite-population) offspring distribution. We will typically consider sequences of GWF processes $\mathbf{Q}^{(N)}$, one for each population size N, just as the Wright–Fisher model prescribes a process for each population size N.

Since our definition of a GWF process encompasses the Cannings models, it includes as special cases the conditional branching models of Karlin and McGregor (1964) and the sampling models of Schweinsberg (2003). Unlike those processes, however, an arbitrary GWF process does not prescribe an explicit mechanism of reproduction, such as the exchangeable vector of Cannings or the unconditional offspring distributions of Schweinsberg and Karlin-McGregor. Rather, a GWF process requires simply that, regardless of the offspring mechanism, the Markov chain describing the resulting dynamics of allele frequency be a martingale and have the same conditional variance as the Wright–Fisher model. As a result, an arbitrary GWF process does not come equipped, *a priori*, with an explicit genealogy – that is, information about which individuals in generation k share a common parent in generation $k - 1$ – such as would be required to define a corresponding backward-time coalescent. Nevertheless, as mentioned above, it is precisely because we discard genealogical information in the definition of a GWF process that we will achieve a more tractable forward-time analysis of standard population genetics quantities – including an understanding of how these quantities behave among the Cannings processes.

We should clarify that the offspring variances σ_N^2 associated with a sequence of GWF processes, and the limiting offspring variance $\sigma^2 = \lim_{N\to\infty} \sigma_N^2$, both differ from the variance in offspring number of an individual, before conditioning, in the Karlin–Taylor conditional branching process; or from the offspring variance of an individual, before sampling, in the sampling models of Schweinsberg (2003).

According to diffusion theory (Norman, 1975), a sequence of Markov chains, $X_k^{(N)}$, indexed by the population size N that each satisfy the mean and variance conditions above and for which $\sigma_N^2 \to$ const as N gets large, can be well approximated by Kimura's diffusion provided some conditional higher moment $E[|X_{k+1}^{(N)}/N - X_k^{(N)}/N|^{2+\delta}|X_k^{(N)}]$ is of order $o(1/N)$ uniformly. In other words, a GWF process with small higher moments (relative to its second moment) can be well approximated by the Wright–Fisher process itself, for large N. We wish to study population models that cannot necessarily be approximated by the standard model, and so our definition does not impose any constraints on the higher moments of the process.

In the following sections we analyze the range of behavior for important genetic quantities (e.g. absorption times, fixation probabilities etc.) exhibited by Generalized Wright–Fisher processes. Put another way, we explore the population genetics of alternative forms of genetic drift that are, in their large features, similar to Wright–Fisher, but which might otherwise depart from the standard model.

2.1. Examples of pure-drift GWF processes

Before analyzing GWF processes in general, we first describe a few examples of such processes whose behavior illustrates the broad range of possibilities. Below we use some of these example processes to illustrate our general results. In order to specify a pure-drift GWF process we must define the transition matrix $\mathbf{Q}^{(N)}$ and verify that it satisfies the first- and second-moment conditions (6) and (7).

2.1.1. Cannings models

All Cannings models qualify as GWF processes, as remarked above. Thus, the set of GWF processes contains a large and well-studied set of models.

We briefly review the definition of Cannings process. Such a model is based on a random exchangeable vector $\nu = (\nu_1, \ldots, \nu_N)$ that satisfies the fixed population constraint $\sum_{i=1}^{N} \nu_i = N$. If individual i contributes ν_i offspring of its own type to the next generation, then the number of individuals of a given type, X_k, is a Markov chain with transition probabilities

$$X_{k+1}|X_k \stackrel{d}{=} \sum_{i=1}^{X_k} \nu_i$$

where $\stackrel{d}{=}$ denotes equality in distribution. The Wright–Fisher model occurs as a special case when ν is a multinomial distribution. A simple computation of moments shows that such processes satisfy (6), (7) and are thus GWF processes.

When a GWF process X_k is generated from an underlying Cannings process, then the conditional distribution $X_{k+1}|\{X_k = 1\}$ coincides with the marginal offspring distribution ν_1. Furthermore, if a GWF process X_k is generated from an underlying Cannings process then the exchangeable vector ν contains information about the genealogical properties of a sample, and it can be used to study a corresponding coalescent (see Discussion).

2.1.2. The Λ_1-model

This process is in some sense the antipode to the Wright–Fisher process, and it will serve as a useful, even if biologically unrealistic, example. The Λ_1-model is a Cannings process defined by the transition matrix, when $i \neq j$:

$$\mathbf{Q}_{ij} = \begin{cases} \dfrac{N-i}{N^2}, & j = 0 \\ 1 - \dfrac{1}{N}, & j = i \\ \dfrac{i}{N^2}, & j = N. \end{cases} \tag{8}$$

At $i = 0, i = N$ the matrix has the standard absorbing states.

Viewed as a Cannings model, the underlying exchangeable vector ν is specified by the probabilities $P(\nu = (1, \ldots, 1)) = 1 - 1/N$, $P(\nu = (N, 0, \ldots, 0)) = 1/N^2$, $P(\nu = (0, N, 0, \ldots)) = 1/N^2, \ldots, P(\nu = (0, \ldots, 0, N)) = 1/N^2$. This process describes a population in which individuals typically produce exactly one offspring each generation, until a random time (of average length N generations) at which a single individual replaces the entire population. We use the nomenclature Λ_1, originating from the theory of Λ-coalescents (Pitman, 1999), a class of non-standard coalescents parameterized by measures on [0, 1]. Our discrete forward-time Λ_1-process corresponds, in the retrospective theory, to a generalized coalescent with Λ measure concentrated at the point 1. The Λ_1-Cannings model is unrealistic for most biological populations, but it will be useful for delimiting certain extreme aspects of GWF processes.

2.1.3. Eldon–Wakeley models

The Λ_1-Cannings model can be generalized to a process in which individuals produce exactly one offspring each generation, until a random time at which a single individual replaces a fraction λ of randomly chosen individuals from the entire population. The value $0 < \lambda \leq 1$ is a parameter of the model. This generalized model is also a Cannings process, and its corresponding exchangeable vector ν gives weight $1 - c/N$, for some positive value c, to the partition $N = 1 + \cdots + 1$, and weight c/N to the partition $N = \lambda N + 1 + \cdots + 1 + 0 + \cdots 0$ where the latter equation has $(1 - \lambda)N$ ones. Note that λN is required to be an integer for the (discrete) process. For any σ_N, there is a unique value c such that (5) is satisfied. This value determines how often the population goes through a "bottleneck"—i.e. a jump in allele frequency. These processes have been proposed by Eldon and Wakeley (2006, 2008, 2009) as a model for the population genetics of Pacific oysters. Eldon–Wakeley processes correspond, in the retrospective theory, to Λ-coalescents with Λ measure concentrated at the number λ. When $\lambda = 1$, this process reduces to the Λ_1-model above; the limit $\lambda \to 0$ can be shown to converge to the Wright–Fisher model.

To be entirely clear, the Eldon–Wakeley process defined here is slightly simpler than the definition given by Eldon and Wakeley (2006), which included a Moran-like process during the time intervals between bottleneck events. Nonetheless, the two definitions have the same continuum limit (see below), provided the parameter γ in Eldon and Wakeley (2006) is less than two.

2.1.4. Λ-models

The Eldon–Wakeley processes can be generalized to define a yet larger class of Cannings models. In an Eldon–Wakeley process the fraction of the population replaced in a bottleneck event is a fixed deterministic number, λ. We can alternatively assume that λ is randomly distributed according to some probability measure Λ on the interval [0, 1]. It can be shown that the resulting models are also Cannings processes, and, in fact, are the forward-time analogues of the so-called Λ-coalescents, parameterized by measures Λ and studied in Pitman (1999), Sagitov (1999) and Donnelly and Kurtz (1999b). We shall refer to these models as Λ-processes, since they are the two-dimensional projections of the Λ-Fleming–Viot measure-valued processes.

2.1.5. Power-law processes

The divide between the Wright–Fisher process and Λ_1-process can be partly spanned by considering a special, one-parameter family that we call the power-law processes. These are defined as the pure-drift Markov processes whose transition matrices have tails that decay at rate $\alpha > 0$:

$$\mathbf{Q}_{ij} = \begin{cases} \dfrac{c_i}{1 + b_{1,i}|j - i|^{\alpha}}, & j < i, 0 < i < N \\ \dfrac{c_i}{1 + b_{2,i}|j - i|^{\alpha}}, & j \geq i, 0 < i < N \\ \delta_{0,j}, & i = 0 \\ \delta_{N,j}, & i = N \end{cases} \tag{9}$$

where $b_{1,i}, b_{2,i} > 0$ are "slope" parameters and the normalization constant c_i is chosen so that $\sum_{j=0}^{N} \mathbf{Q}_{ij} = 1$. For any choice of population size N and decay rate $\alpha > 0$, there exist unique positive constants $b_{1,i}, b_{2,i}, c_i$ which satisfy the GWF conditions ((4) and (5)) (Der, 2010).

Unlike the Wright–Fisher model, which exhibits rapid, Gaussian decay in the rows of its transition matrix, the power-law processes exhibit slower, power-law decay. As a result, we may informally think of the Wright–Fisher process as a power-law process with $\alpha = \infty$ (i.e. very rapid decay). In fact, it is possible to prove that the Wright–Fisher model has the same diffusion limit as the power-law models for any $\alpha \geq 3$ (Der, 2010). By contrast, the power-law models with $\alpha < 3$ cannot be accurately approximated by Kimura's diffusion, and they exhibit behavior that is markedly different from that of the Wright–Fisher model, as we discuss below. We will call any GWF process, such as the power-law processes with $\alpha < 3$, for which the standard diffusion approximation does not hold, a "non-diffusive" population model. There are nevertheless continuum techniques for analyzing the behavior of such processes as $N \to \infty$, described below. In fact, as we discuss below, the continuum limits for the power-law processes $0 < \alpha < 2$ will turn out to coincide with certain families of Λ-processes.

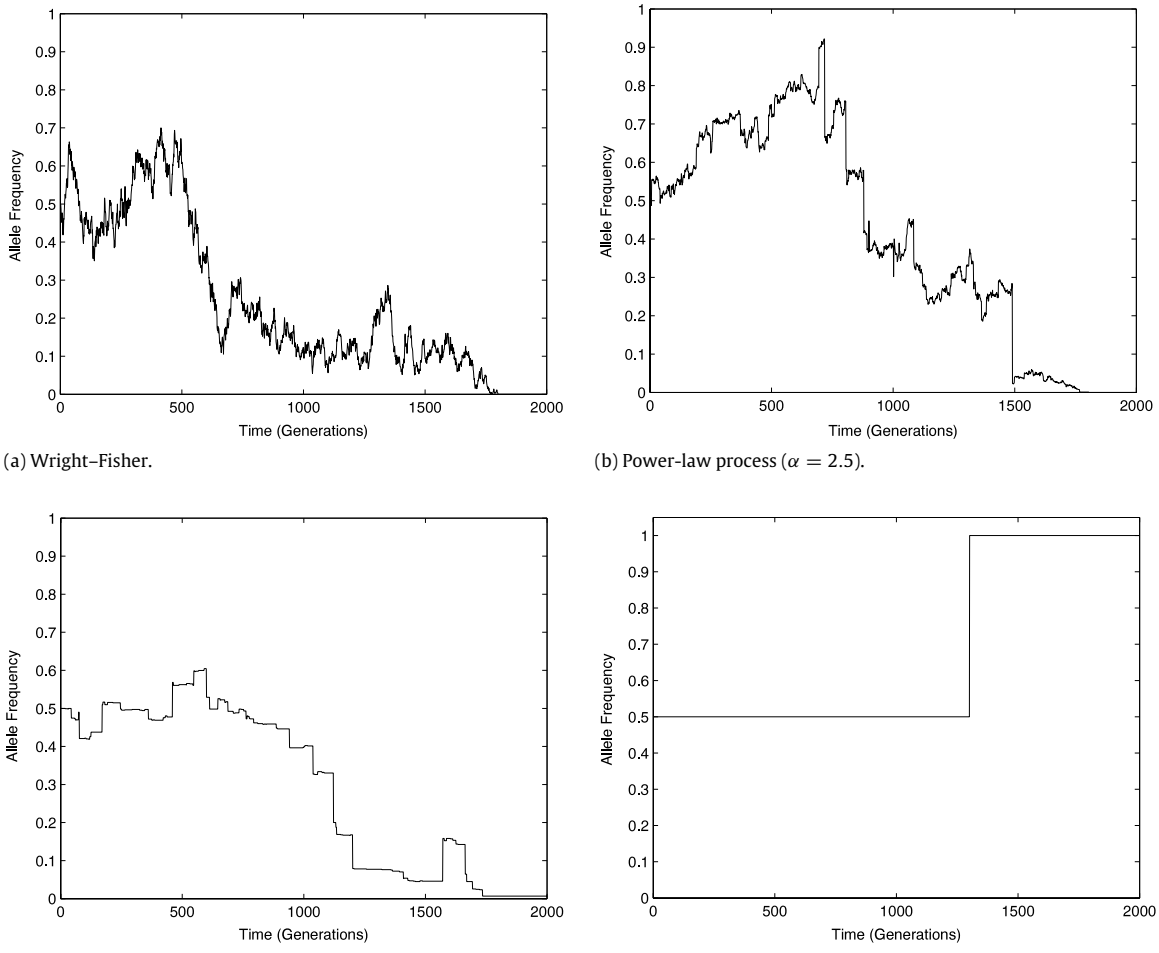
Fig. 1 shows simulated sample paths from several example GWF processes, including the Wright–Fisher model, power-law models, and the Λ_1-model. As the figure illustrates, only the Wright–Fisher model has continuous sample paths (in its continuum limit), while the other example processes experience discontinuous changes in allele frequencies, the sizes of which increase from panels (a) through (d). These paths illustrate the notion that the power-law models, as α varies from zero to three, span a range of behavior delineated by the Wright–Fisher model on one side and the Λ_1-model on the other.

2.1.6. Convex combinations

Many other GWF processes can be formed by taking mixtures of two GWF processes. If we have two GWF processes with a common offspring variance and transition matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, then their combination $\mathbf{P} = a\mathbf{P}_1 + (1 - a)\mathbf{P}_2$, for $0 < a < 1$, is also a GWF process of the same offspring variance, as is clear from the linearity of the conditions on first and second moments. The convexity structure of many classes of GWF processes can be useful for deriving their forward-time dynamics in terms of simpler models. Convex combinations also have the interpretation as a randomization of models; the Λ-processes, for example, can be obtained as the convex hull of the set of Eldon–Wakeley models.

2.2. Incorporating selection and mutation

Our analysis here is concerned with alternative forms of genetic drift – that is, stochastic effects on allele frequencies arising from



(a) Wright–Fisher.

(b) Power-law process ($\alpha = 2.5$).

(c) Power-law process ($\alpha = 1.5$).

(d) Λ_1 process.

Fig. 1. Simulated sample paths for several Generalized Wright–Fisher processes. The Wright–Fisher process has continuous sample paths in its diffusion limit, whereas non-Wright–Fisherian models experience discontinuous changes in allele frequencies (in their continuum limit). The Wright–Fisher model and the Λ_1-model can been seen as antipodes, with the power-law processes intermediate, as a function of the frequency and amplitude of their jump sizes.

the finite-population size. It is particularly interesting to ask how alleles subject to these forms of genetic drift behave in the presence of natural selection and mutation – a question that is not easily accommodated using the backward-time genealogical approach (see Wakeley (2009) and Discussion). Selection and mutation are typically modeled as deterministic forces on the allele frequencies, and they have a well-accepted mathematical formulation.

We can easily incorporate the standard forms of selection and mutation into the pure-drift processes defined above. Let $\mathbf{Q}^{(N)}$ be the transition matrix for a pure-drift GWF process. Selection and mutation may be incorporated in a natural way by multiplying $\mathbf{Q}^{(N)}$ by additional transition matrices denoted $\mathbf{S}^{(N)}$ and $\mathbf{M}^{(N)}$ respectively. The model with drift, mutation, and selection is then defined by the composite transition matrix

$$\mathbf{P}^{(N)} = \mathbf{S}^{(N)}\mathbf{Q}^{(N)}\mathbf{M}^{(N)}. \tag{10}$$

Conceptually, the product above partitions the order of events in a generation into three components: (1) alteration of the allele frequency based on differential fitness (i.e. survival until reproduction) ($\mathbf{S}$); (2) a reproduction stage that includes genetic drift ($\mathbf{Q}$); and (3) a mutational stage ($\mathbf{M}$).

In order to define the mutation and selection matrices we require only that $\mathbf{S}^{(N)}$ and $\mathbf{M}^{(N)}$ reduce $\mathbf{P}^{(N)}$ to the classical theory of deterministic evolution (Haldane, 1932) when the offspring variance equals zero (i.e. when $\mathbf{Q}^{(N)} = I$), as $N \to \infty$. Mathematically, under the scaling assumption $\sigma_N^2 \to$ const, this amounts to requiring that the transition matrices have generators that converge to the standard convection terms in Kimura's diffusion:

$$\lim_{N\to\infty} N(\mathbf{S}^{(N)} - I)u_N = \gamma x(1-x)\frac{\mathrm{d}u}{\mathrm{d}x} \tag{11}$$

$$\lim_{N\to\infty} N(\mathbf{M}^{(N)} - I)u_N = \frac{1}{2}(-\theta_1 x + \theta_2(1-x))\frac{\mathrm{d}u}{\mathrm{d}x} \tag{12}$$

where the allele of interest has selective advantage $\gamma = Ns$, mutation away from class i occurs at rates $\theta_i = 2N\mu_i$; u is any smooth function, and u_N is a vector, $(u(0), u(1/N), \ldots, u(N - 1/N))$, sampled from u. It can be shown that the order in which the matrices are multiplied in (10) is immaterial, as $N \to \infty$.

Appendix A includes a concrete construction of matrices $\mathbf{M}^{(N)}$ and $\mathbf{S}^{(N)}$, which satisfy the conditions above. When $\mathbf{Q}^{(N)}$ is the pure-drift Wright–Fisher model, these choices of mutation and selection matrices produce a model $\mathbf{P}^{(N)}$ that agrees extremely well with the standard Wright–Fisher model including selection and mutation (Ewens, 2004), even at very low population sizes. It should be emphasized, though, that none of our theoretical results depends upon the specific form of mutation and selection matrices, but only on the canonical asymptotic forms (11) and (12).

3. The population genetics of GWF processes

In the following sections we describe the behaviors of important genetic quantities (e.g. absorption times, fixation

probabilities) that are possible under Generalized Wright–Fisher processes. Aside from providing significant information about the Wright–Fisher model itself, our analysis uncovers a startlingly broad span of phenomena for GWF processes—despite the fact that all GWF processes share the same first- and second-order moments as the standard model.

In stating our results on absorption times and fixation probabilities, below, we focus on families of GWF processes for which $\sigma_N^2 \to \sigma^2 < \infty$—i.e. the offspring variance remains finite as population size grows. For such models we likewise assume that $s = O(1/N)$ and $\mu_i = O(1/N)$. This focus restricts our attention to models that are more similar to the standard Wright–Fisher model (for which $\sigma_N^2 \to 1$) than those exhibiting divergent σ_N. Nonetheless, our results on absorption times and fixation probabilities can be readily generalized to accommodate models in which σ_N^2 diverges with N or shrinks to zero. We elaborate on the relationships between time scaling, diffusiveness, and σ_N^2 at greater length in the Discussion.

3.1. Time evolution of moments

We start by studying the moments of GWF processes—including, for example, the heterozygosity. The definition of a pure-drift GWF process (4) and (5) implies, by taking expectations, that

$$EX_{k+1} = EX_k \tag{13}$$

$$EX_{k+1}^2 = \frac{N\sigma_N^2}{N-1} EX_k + \left(1 - \frac{\sigma_N^2}{N-1}\right) EX_k^2. \tag{14}$$

These recurrences can be solved to produce closed-form expressions for the time evolution of the first two moments:

$$EX_k = X_0 \tag{15}$$

$$EX_k^2 = X_0^2 \lambda_N^k + NX_0(1 - \lambda_N^k) \tag{16}$$

$$\operatorname{Var} X_k = X_0(N - X_0)(1 - \lambda_N^k) \tag{17}$$

where $\lambda_N = 1 - \frac{\sigma_N^2}{N-1}$. The behavior of these moments coincides exactly with those of the Wright–Fisher process. As a fundamental corollary, *any* function of the first two moments, including the heterozygosity $EX_k(N - X_k)$, is invariant over the entire set of GWF processes with a given offspring variance. Moreover, the invariance of the unconditional moments extends in the presence of non-zero mutation:

Theorem 1. *For any mutation rates $\theta_1 \geq 0, \theta_2 \geq 0$, the first two* unconditional *moments of any GWF process with $\sigma^2 = 1 - 1/N$ coincide, for all time, with the corresponding Wright–Fisher model. In particular, the first two moments of the equilibrium distribution for a GWF process converge (as $N \to \infty$) to*

$$E(X_\infty/N) \to \frac{\theta_2}{\theta_1 + \theta_2} \tag{18}$$

$$E(X_\infty/N)^2 \to \frac{(1+\theta_2)\theta_2}{(\theta_1+\theta_2)(1+\theta_1+\theta_2)}. \tag{19}$$

The preceding result follows from a simple calculation with the mutation matrix $\mathbf{M}^{(N)}$, which shows that even under mutation EX_k and EX_k^2 satisfy linear difference equations identical to those of the Wright–Fisher model.

No such closed system of equations is possible once selection is present, however, and so the first and second moments of the stationary distribution of a GWF process with both mutation and selection may differ from those of the corresponding Wright–Fisher model. This suggests an interesting interaction between selection and non-traditional forms of genetic drift—an idea that we pursue in greater detail below.

3.2. Neutral absorption time properties

The time until fixation or loss of a neutral allele is one of the simplest and most fundamental properties of interest to population geneticists. Absorption times for the Wright–Fisher process have long been understood and used to study empirical populations. For Generalized Wright–Fisher processes, the second-moment condition (7) implies that 0 and N are the only two absorbing states, as in the Wright–Fisher process. Moreover, elementary martingale theory implies that the eventual loss or fixation of an allele is assured. Furthermore, the fixation probability satisfies the following, classical result:

Theorem 2 (*Neutral Probability of Fixation*)**.** *For a GWF process without selection or mutation, an allele will fix with probability equal to its initial frequency.*

The mean absorption time of a GWF process may differ from that of the standard Wright–Fisher model. We have characterized the range of mean absorption times that are possible among all GWF models:

Theorem 3 (*Absorption Time Bounds*)**.** *Suppose that $\sigma_N \to \sigma$.*

1. *For a pure-drift GWF process, the mean absorption time (in generations) $E\tau_N$ of an allele initiated at a fixed fraction x of the total population size N satisfies (as $N \to \infty$):*

$$-\frac{N}{\sigma^2}(x\log x + (1-x)\log(1-x)) \leq E\tau_N \leq \frac{N}{\sigma^2}\log(Nx(1-x)). \tag{20}$$

2. *For a pure-drift GWF process, the mean absorption time $E\tau_N$ of an allele initiated at $X_0 = 1$ satisfies (as $N \to \infty$)*

$$\frac{1}{\sigma^2}\log N \leq E\tau_N \leq \frac{N}{\sigma^2}. \tag{21}$$

The asymptotic bounds described in this theorem are optimal in order N, and they are achieved by specific examples of GWF models. The proof of this result is sketched in Appendix C, and it produces stochastic estimates on the distribution of the absorption time itself. While not pursued here, this approach provides bounds on all moments of the absorption time, not only the mean.

Note that the Wright–Fisher process meets both of the lower bounds above, in order N: an allele initiated at a fixed fraction x of the population will absorb in $O(N)$ generations, whereas an allele initiated $X_0 = 1$ will absorb in $O(\log N)$ generation. Moreover, we have constructed a GWF process (the "maximal process", described in Appendix B) whose mean absorption time starting from a fixed fraction of N achieves the upper bound of $N \log N$ generations. Finally, the Λ_1-Cannings process meets the upper bound, $O(N)$ generations, for absorption when started at $X_0 = 1$. Thus, the bounds described by Theorem 3 represent the full range of absorption times among GWF models.

Perhaps surprisingly, Theorem 3 shows that the mean absorption time of an allele initiated at intermediate frequencies, e.g. $X_0 = N/2$, is a fairly stable quantity: this varies over a relatively narrow range, $O(N)$ to $O(N \log N)$, over all GWF processes. By contrast, when the initial frequency of a mutant allele is more extreme (near 0 or N), then the absorption time is highly sensitive to the form of genetic drift—and it ranges from $O(\log N)$ to $O(N)$ generations depending upon the underlying GWF model.

We observed above that the absorption times under the Wright–Fisher process are very close to the minimum absorption times achievable under any GWF model; for example, in the classical process, the absorption times for an allele initiated at $X_0 = N/2$ or $X_0 = 1$ are $(2\log 2)N$ and $2\log N$, respectively (Ewens,

2004). This result suggests that the form of genetic drift encoded by the Wright–Fisher model is somewhat unusual compared to the forms of genetic drift accommodated by the larger class of generalized models—a point we consider in greater detail below.

3.3. Neutral fixation time properties

The conditional fixation time of a neutral allele is also of interest to population geneticists. Once again, this quantity is well understood for the Wright–Fisher model, or any other model that shares the standard Kimura diffusion limit. Here we study the behavior of the conditional fixation time among the larger family of GWF processes.

Given a pure-drift GWF process, we consider only those paths which happen to fix (i.e. absorb at state N). What is the mean time to absorption among such paths, and how does it vary across underlying GWF models? Our next theorem answers this question:

Theorem 4 (*Conditional Fixation Time Bounds*)**.** *Suppose that* $\sigma_N \to \sigma$.

1. *Started at a fixed fraction x of the population size N, the mean time to fixation* $E\tau_N^*$ *of a pure-drift GWF process satisfies (as* $N \to \infty$*)*

$$\frac{N}{\sigma^2}(1-x) \le E\tau_N^* \le \frac{N}{\sigma^2}\log(N(1-x)). \tag{22}$$

2. *Started at* $X_0 = 1$*, the mean time to fixation* $E\tau_N^*$ *satisfies (as* $N \to \infty$*)*

$$\frac{N}{\sigma^2} \le E\tau_N^* \le \frac{N\log N}{\sigma^2}. \tag{23}$$

The proof is similar to that of Theorem 3. Once again, the asymptotic bounds described by these formulas are tight, with the Wright–Fisher process behaving near the lower bounds, whether initiated at a fixed fraction of the population or with a single individual. Similarly, the "minimal process" (described in Appendix B) meets the upper bounds on mean fixation time.

It is worth commenting that the *conditional* fixation time described here is significantly more stable (i.e. does not vary widely across GWF processes), than the unconditional absorption time, even when the allele frequency starts near the lower boundary. Explanations for this behavior are deferred to the Discussion section.

3.4. Absorption times under one-way mutation

If we now suppose that one-way mutation occurs from an allele of interest to the other allele at rate $\theta = 2N\mu$, then eventual loss of the allele is assured. Using techniques similar to those above, we can derive bounds on the mean time before loss. For brevity, we report here a simple, lower bound on the mean absorption time:

Theorem 5 (*Time to Loss for One-Way Mutation*)**.** *Suppose one-way mutation occurs at rate* θ*, then*

1. *Started at a fixed fraction x of the population size, the mean time to loss* $E\tau_N$ *of an allele satisfies (as* $N \to \infty$*)*

$$E\tau_N \ge \frac{xN\log\left(\frac{x(\sigma^2+\theta/2)}{\sigma^2}\right)}{(\theta/2+\sigma^2)\left(x-\frac{\sigma^2}{\sigma^2+\theta/2}\right)}. \tag{24}$$

2. *Started at* $X_0 = 1$*, the mean time to loss* $E\tau_N$ *satisfies (as* $N \to \infty$*)*

$$E\tau_N \ge \frac{\log\left(\frac{\sigma^2}{\sigma^2+\theta/2}N\right)}{\sigma^2}. \tag{25}$$

For comparison, Kimura's classical expression for the time to loss of a new mutant with one-way mutation is[1]

$$E\tau_N = \int_{1/N}^{1} y^{-1}(1-y)^{\theta-1}\,dy \sim 2\log N$$

where $\sigma^2 = 1$. Thus, the time to loss in the Wright–Fisher model is again close to the lower bound given by our Theorem 5. By contrast, the mean time until loss of a new allele can be significantly larger in general; the Λ_1-model, for example, has a mean loss time of $O(N)$ generations.

In all the cases presented thus far the Wright–Fisher process absorbs very quickly, compared to the full range of possible behaviors among all Generalized Wright–Fisher models, particularly when initiated at allele frequencies near a boundary. These results suggest that the binomial form of genetic drift encoded by the Wright–Fisher process is much stronger than the drift of most other GWF processes, despite the fact that the second moments are the same.

3.5. Probabilities of fixation under selection

We turn now to another classical quantity of interest: the probability of fixation of an allele under selection. This topic has received relatively little attention in the literature on generalized population models, because those models are typically studied in the coalescent framework, which does not easily accommodate selection (but see Krone and Neuhauser (1997), Neuhauser and Krone (1997), Etheridge and Griffiths (2009), Etheridge et al. (2010) and Discussion). In the neutral case, all GWF processes have the same probability of fixation, equal to the allele's initial frequency. However, when one allele is selectively favored over the other, we show that there is enormous variation in fixation probabilities, as well as some startling phenomena that differ qualitatively from standard results in population genetics.

We have derived effective bounds for fixation probabilities of GWF processes. We have also found specific processes that attain these bounds; we call them the "maximal" and "minimal" processes (see Appendix B). Specifically, we have solved the following extremal problem. Let $\pi(x)$ be the probability of fixation of an allele initiated at frequency x. For a fixed selection pressure Ns and offspring variance σ^2, what GWF process(es) maximize and minimize $\pi(x)$ for all x? A priori, it is unclear whether a single process can be an extremizer *for all initializing frequencies*. Remarkably, a solution does exist over the space of GWF processes, and the computation of their fixation probabilities leads to the following theorem.

Theorem 6. *Let* $\gamma = \frac{Ns}{\sigma^2}$*, and define*

$$\pi_-(x) = \begin{cases} \dfrac{(1+\gamma)x}{1+\gamma x}, & \gamma > -1 \\ 0, & \gamma \le -1 \end{cases} \tag{26}$$

$$\pi_+(x) = \begin{cases} \dfrac{x}{1-\gamma(1-x)}, & \gamma < 1 \\ 1, & \gamma \ge 1. \end{cases} \tag{27}$$

Then the probability of fixation $\pi(x)$ *of a GWF process started at allele frequency x lies between* π_- *and* π_+*:*

$$\pi_-(x) \le \pi(x) \le \pi_+(x), \quad \text{for all } 0 \le x \le 1. \tag{28}$$

[1] Note that the first term in the asymptotic expansion for the expected time is *independent* of θ.

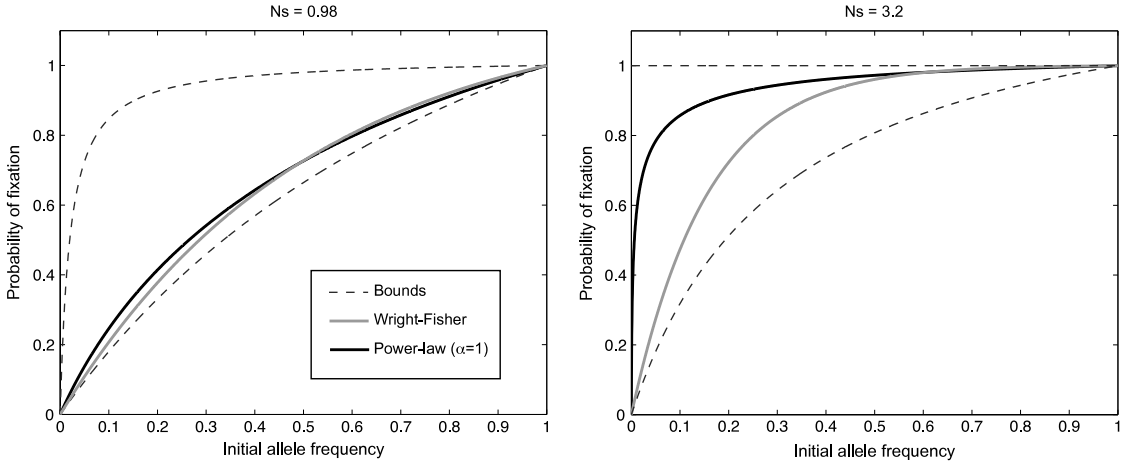


Fig. 2. Probabilities of fixation for Generalized Wright–Fisher processes. The upper and lower bounds for fixation probabilities, which are achieved by the maximal and minimal processes, are shown in dashed. Lighter curves correspond to the Wright–Fisher model, and darker curves correspond to the power-law model with $\alpha = 1$. The left panel shows $Ns = 0.98$, and the right panel $Ns = 3.2$. Note that when $Ns > 1$, the power-law process greatly amplifies the fixation probability compared to the Wright–Fisher model, particularly for an allele introduced at low frequencies. In all cases $\sigma^2 = 1$.

Fig. 2 plots the range of fixation probabilities among GWF processes for a few values of γ, along with the fixation probabilities for several example processes we have introduced. For every selection coefficient γ, at least one of the inequalities in (28) is non-trivial, and in the region $-1 < \gamma < 1$ both inequalities are non-trivial. The maximal and minimal processes, which achieve the extreme fixation probabilities π_+ and π_-, are described in Appendix B.

Many insights about the nature of selection and drift in generalized populations can be gleaned from Theorem 6, whose proof is sketched in the Appendix. We divide the discussion of these insights into three cases.

3.5.1. Fixation of a new mutant, s fixed

Theorem 6 provides bounds on fixation probabilities when an allele frequency starts at a fixed fraction of the population size. The theorem can also be used, heuristically, to derive bounds on the fixation probability π^* of a newly introduced mutant, by inserting the frequency $x = 1/N$ into (28). For large N, small $s > 0$, and $\sigma^2 = 1$, these amount to

$$s \le \pi^* \le 1, \quad s > 0. \tag{29}$$

There are a number of interesting consequences of (29). First among these is the observation that the Wright–Fisher fixation probability of $\pi^* \approx 2s$ is very close to the absolute possible minimum, $\pi^* = s$. Thus, even relative to the enormous class of GWF processes, which contains pathological models (the minimal process in particular), the Wright–Fisher model acts as a vigorous suppressor of selection. Once again, this supports the idea that the form of genetic drift encoded by the Wright–Fisher model (and those models with the same diffusion limit) is extremely strong. In the Wright–Fisher model, drift counteracts the deterministic force of selection more powerfully than in virtually any other generalized population model.

Second, the upper bound on the fixation probability (29) suggests that some processes can tremendously amplify minute selective differences. The Λ_1-model, in fact, is one such example. In Appendix D we show that the probability of fixation of an advantageous allele under Λ_1 is

$$\pi_1(x) = \frac{1}{Ns}\left(\frac{x}{1-x}\right)^{1/(Ns)} \int_x^1 y^{-1/(Ns)}(1-y)^{1/(Ns)-1}\,dy. \tag{30}$$

For a fixed value of $s > 0$, one calculates $\pi^* = \lim_{N\to\infty} \pi(1/N) = 1$. In other words, new mutants are *assured* fixation in the Λ_1-model, regardless of the size of s, in the limit as $N \to \infty$. By contrast, under the Wright–Fisher model fixation is never assured for any value of s, since $\lim_{N\to\infty} \pi(1/N) \approx 1 - \exp(-2s) \approx 2s$.

Remarkably, fixation is achieved with probability one in the Λ_1-model, despite the presence of stochasticity—i.e. genetic drift with $\sigma^2 = 1$. Although the Λ_1-model is an extreme example, similar phenomena occur in more realistic models, including the power-law processes with decay rate $\alpha < 2$. Fig. 3 compares fixation probabilities of a new mutant as a function of population size for a number of processes, some of which possess the extraordinary property of extreme selection amplification. The Wright–Fisher model does not possess this property because, as we have argued, the strong form of binomial genetic drift overwhelms the force of selection even as $N \to \infty$.

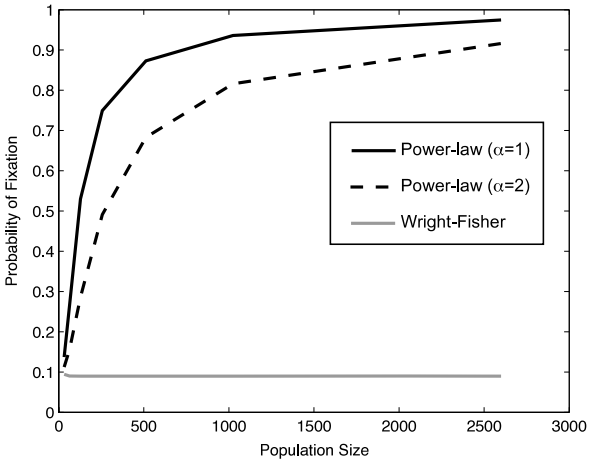


Fig. 3. Fixation probability of a new, advantageous mutant for several Generalized Wright–Fisher processes, as a function of population size N. In all cases $s = 0.05$ fixed. Note that the power-law processes achieve assured fixation as population size gets large, whereas the Wright–Fisher process has an asymptotic fixation probability of approximately $2s$.

The limiting fixation probability of a new mutant as the population becomes large, with s fixed, has been studied by a few other authors. Lieberman et al. (2005) studied a Moran-like processes in a graph-structured population, and found graphs in which π^* exceeded the classical value of $2s$. In particular, those authors constructed a sequence of graphs (the analogue of $N \to \infty$) for which the probability of fixation of a newly introduced type was assured. Our results indicate that an identical phenomenon

can occur even in a well-mixed population, requiring only higher-order modifications to the form of genetic drift.

In the setting of structured populations, processes that suppress selection also exist, for which fixation probabilities do not differ from neutrality even in the presence of large fitness differentials (Lieberman et al., 2005). In a well-mixed GWF population, by contrast, such possibilities are precluded by the lower bounds in Theorem 6 and (29).

3.5.2. Fixation probability from a fixed initial frequency, Ns fixed

The content of Theorem 6 reveals several other striking features about the nature of selection and drift in GWF populations. In this section we consider the fixation probability for processes initialized at a fixed fraction, x, of the total population size. In particular, we study the behavior of such a mutant as the population size gets large, with Ns held constant. The standard diffusion approximation of the Wright–Fisher model indicates that such an allele fixes with probability $\pi(x) = \frac{1-\exp(-2Nsx)}{1-\exp(-2Ns)}$. This expression is strictly less than one (because Ns is finite), so that such an allele is not guaranteed fixation even in very large population sizes.

Non-standard models can produce very different behavior. According to the optimal upper bounds in Theorem 6, $\pi_+(x) \to 1$ as $Ns \uparrow \sigma^2$. Thus an allele starting at any frequency x will be assured fixation in certain continuum GWF models for some *finite* value of Ns (as compared to the limit of the previous section, which required $Ns \to \infty$). Interpreted from the point of view of the discrete models, there exist GWF processes that produce assured fixation of advantageous alleles, as the population size increases, *even with Ns constant*. An example of this behavior is shown in Fig. 4.

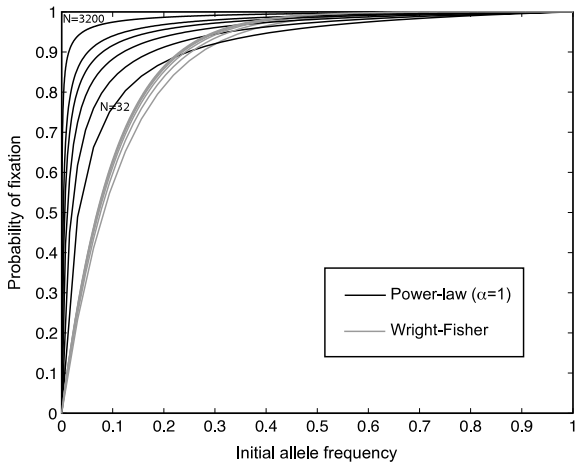


Fig. 4. Some GWF processes produce the assured fixation of an allele, initially at any frequency, as $N \to \infty$ with Ns fixed. Here we plot the probability of fixation as a function of initial frequency for the power-law process with $\alpha = 1$ (dark) and for the Wright–Fisher process (light). In both cases $\sigma^2 = 1$ and $Ns = 5$. The curves, from bottom to top, correspond to $N = 32, 64, 128, 256, 512, 3200$. For the power-law process, the fixation probability approaches one, for any initial frequency x, as N gets large; whereas for the Wright–Fisher process, fixation is not assured for any initial frequency, and the curves converge to function strictly less than 1 as N gets large.

The maximal process is the ultimate such amplifier of selection and it features this behavior of assured fixation. But this phenomena appears in more realistic models as well. In particular, some power-law models exhibit this behavior. For any power-law model with $\alpha < 2$, there exists a critical selective value $\gamma^* < \infty$, such that alleles initially at any frequency x with selective advantage γ, larger than γ^*, will assuredly fix, as $N \to \infty$. When $\gamma < \gamma^*$, fixation probabilities are strictly less than one, as in the Wright–Fisher model. For the power-law processes with $0 < \alpha < 2$ this critical value of selection satisfies (see Appendix E):

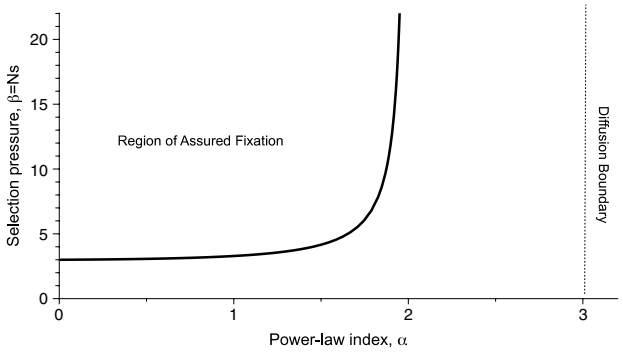


Fig. 5. Regime of assured fixation in power-law models. Processes with index and selection parameters α, γ lying above the curve $\gamma^* = (3-\alpha)\int_0^1 (1-t)^{-\alpha}|\log t|\,dt$ experience assured fixation of an allele initially at any frequency x, as $N \to \infty$ with $\gamma = Ns$ fixed. Models below the curve have strictly sub-unital fixation probabilities. For $\alpha \geq 3$, the power-law processes are well approximated by the standard Kimura diffusion, and assured fixation does not occur.

$$\gamma^* = (3-\alpha)\int_0^1 (1-t)^{-\alpha}|\log t|\,dt. \tag{31}$$

A graph of this expression is shown in Fig. 5. In fact, a more general formula for the critical selection pressure can be established for the Λ-processes; we defer that exposition to a future publication. We note that γ^* ranges from 3 to approximately 3.3 in the range $0 < \alpha < 1$. In other words, despite the presence of non-trivial genetic drift in such models, fixation is virtually assured in large populations, even for only moderately advantageous alleles.

The critical value γ^* for assured fixation is an increasing function of α, the power-law decay rate of the underlying GWF process. This makes intuitive sense: as α increases, the power-law process becomes more similar to the Wright–Fisher model, for which fixation is never assured. In fact, for $\alpha > 3$ the power-law process has the same diffusion limit as the Wright–Fisher, and so $\gamma^* = \infty$ in those cases.

The upper bound in Theorem 6 also implies a minimum possible value for γ^*: namely $\gamma^* \geq 1$. Thus, whenever $Ns \leq \sigma^2$, an advantageous allele always faces some positive probability of loss, even as $N \to \infty$, in any GWF model. Naturally, this condition degenerates at $\sigma^2 = 0$, in agreement with the well-known behavior of deterministic evolution without any random genetic drift.

3.5.3. Fixation of new mutants, Ns fixed

Aside from the exotic phenomena described above, it is also possible that a GWF population may exhibit a non-zero asymptotic fixation probability for a new mutant, even when Ns is held constant. Under the same assumptions, the Wright–Fisher model has an asymptotic fixation probability of zero. We do not have an explicit theoretical capability to predict this behavior, because it would rely on a more detailed analysis than provided by continuum limits (below). Nevertheless, our numerical investigations (Fig. 6) strongly suggest that this phenomenon holds for at least those power-law processes with $0 < \alpha < 2$. In these cases, whenever $\gamma > \gamma^*$ there is a positive probability that a newly arisen mutant will fix, as $N \to \infty$, even with Ns fixed. We have not yet identified the general conditions under which this form of selection amplification occurs.

Lessard and Ladret (2007) have also studied the problem of fixation probability under selection for a subset of GWF processes: the discrete exchangeable Cannings models. In particular, those authors derived an expression for the derivative, with respect to

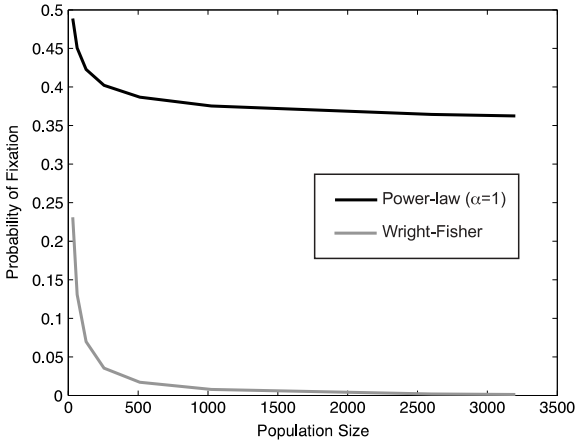


Fig. 6. Probability of fixation of a new mutant, as a function of population size, with $Ns = 5$ fixed. The top curve corresponds to the power-law process with $\alpha = 1$, and the bottom curve to the Wright–Fisher process. The power-law process amplifies selection, and it has a strictly positive asymptotic fixation probability ≈ 0.35, whereas the Wright–Fisher process has an asymptotic fixation probability of zero. In both cases $\sigma^2 = 1$.

selection pressure s, of the fixation probability of a single mutant, under the assumption of weak selection $s \to 0$. By contrast, our results are free from the assumption of weak selection and they provide information about the fixation probability initiated at any frequency, not only at $1/N$.

3.5.4. Deleterious mutations

If a selective strength $s < 0$ is inserted into Theorem 6, one discovers an asymptotic upper bound on the fixation probability π^* of a newly introduced deleterious mutant:

$$\pi^* \leq \frac{1}{N^2|s|}. \tag{32}$$

This upper bound on the fixation probability is achieved by the Λ_1-process.

For finite-population sizes, the quantity above can be larger or smaller than the corresponding classical value $\pi^*_{WF} \approx (1 - \exp(-2|s|)) \cdot \exp(-2N|s|)$ in the Wright–Fisher case, depending upon $|s|$. In particular, the fixation probability of a slightly deleterious new mutation may exceed that of the Wright–Fisher process. By contrast, when selection is sufficiently strong, the loss of a newly introduced disadvantageous mutant is assured under some GWF processes, unlike for Wright–Fisher. The same considerations hold for deleterious alleles initiated at a positive frequency of the population.

4. Continuum theory for GWF processes

Many quantitative questions in population genetics are amenable to diffusion theory, as developed by Kimura and others. Loosely speaking, the diffusion approximation of the Wright–Fisher model consists of rescaling the state space and time domain of the discrete Markov chain, resulting in a continuous Markov process in the limit $N \to \infty$. Such continuum limits have a number of attractive properties. It is usually easier to obtain qualitative insight into the continuum process, since models that differ only in higher-order effects are collapsed onto the same continuum attractor. The space of continuum limits is thus smaller and more amenable to analysis than the original space of discrete models. In many cases it is possible to obtain closed-form solutions to exit-time problems, such as the absorption times and probabilities discussed above, which are intractable in the discrete model.

A wide range of processes (including, for example, the Moran process) share the same diffusion limit as the Wright–Fisher model itself—and they are thus seen as equivalent in most practical applications of the theory. In fact, as mentioned above, any sequence of GWF processes whose offspring variance converges to a constant, i.e. $\sigma_N^2 \to \sigma$, can be well approximated by Kimura's standard diffusion, provided *some* conditional higher moment, $E[|X_{k+1}/N - X_k/N|^{2+\delta}|X_k]$, (for any $\delta > 0$) is of order $o(1/N)$. Thus, passing to a continuum limit greatly reduces the number of distinct GWF processes, up to a rescaling of time.

Nonetheless, many GWF processes do not converge to Kimura's diffusion. Indeed, we can classify GWF processes by the distinct continuum limits they attain. The mathematical formalism required to study the continuum limits of GWF processes is more complicated than that of standard diffusion theory. We have developed the necessary mathematical tools to formulate and productively use continuum limits of GWF processes. While the details of this theory will receive their full exposition in future publications (see also the thesis Der (2010)), it is pertinent here to provide an introduction and outline the essential ingredients of, and insights gained, from continuum approximations to Generalized Wright–Fisher processes.

The continuum limit for the Wright–Fisher model, and more generally for any GWF model, can be studied via any number of different state-space scalings. In Kimura's original approach, allelic numbers are scaled by $1/N$, so that the limiting diffusion is a process on the unit interval $[0, 1]$, which is appropriate for examining the process when initialized at a fixed fraction, x, of the population size. On the other hand, if fluctuations at low allelic frequencies are the object of interest, one might employ, for example, a $1/\sqrt{N}$ scaling; here the continuum limit is a process (different from the Kimuran diffusion) on $[0, \infty)$, and it can be used to study the process initiated at some frequency $[x\sqrt{N}]$ (Karlin and Taylor, 1981). In either case, once a state-space scaling has been chosen, there is typically only a single time scaling, depending on the asymptotics of the offspring variance σ_N^2, for which there exists a non-trivial convergent limit. When $\sigma_N \to \sigma < \infty$, and $\sigma > 0$, this time scaling coincides precisely with the state scaling.

By analogy with the standard diffusion theory of the Wright–Fisher model, we focus here on continuum approximations of GWF processes when both time and space are scaled by the population size N, under the assumption of a convergent offspring variance, i.e. $X(t) = \lim_{N\to\infty} \frac{1}{N} X_{[Nt]}$, whenever it exists. It is important to note, however, that the assumption $\sigma_N \to \sigma < \infty$ is not essential to the continuum theory: divergent σ_N patterns can be accommodated simply by changing the time scaling from N to some (for example) power N^α, $\alpha \neq 1$. In such cases, the natural scaling regime for the selection pressure s and mutation rates μ_i must also be modified to reflect the non-unital exponent of N.

Associated with the convergence of the actual stochastic processes X_k is the convergence of the generator matrices: $\mathbf{P}^{(N)} - \mathbf{I}$. Under the scaling assumptions we have made, the existence of a continuum approximation amounts to asking whether there is an operator G acting on functions u for which

$$Gu = \lim_{N\to\infty} N(\mathbf{P}^{(N)} - \mathbf{I})u_N \tag{33}$$

where u_N is vector of N samples from u. For many sequences of GWF processes such a generator does indeed exist, and we have proven the following representation theorem, informally stated here.

Theorem 7. *Let G be a limiting operator as in (33) for a sequence of GWF Markov chains $\mathbf{P}^{(N)} = \mathbf{S}^{(N)}\mathbf{Q}^{(N)}\mathbf{M}^{(N)}$ with asymptotic offspring*

Table 1
Continuum limits for various GWF processes.

Model	Kernel $d\Omega_x(y)$	Generator $Gu(x)$
Wright–Fisher	$\delta_x(y)$	$\frac{1}{2}x(1-x)u''(x)$
Λ_1 process	$x\delta_0(y)+(1-x)\delta_1(y)$	$(1-x)u(0)-u(x)+xu(1)$
Power-law processes, $\alpha \geq 3$	$\delta_x(y)$	$\frac{1}{2}x(1-x)u''(x)$
Power-law processes, $2<\alpha<3$	$\frac{(3-\alpha)\,dy}{(x^{3-\alpha}+(1-x)^{3-\alpha})\lvert y-x\rvert^{\alpha-2}}$	$x(1-x)\int_0^1 \frac{u(y)-u(x)-u'(x)(y-x)}{(y-x)^2}\,d\Omega_x(y)$
Cauchy process ($\alpha=2$)	$1\cdot dy$	$x(1-x)\int_0^1 \frac{u(y)-u(x)-u'(x)}{(y-x)^2}\,dy$
Power-law processes, $0<\alpha<2$	$\frac{(3-\alpha)\,dy}{x^{2-\alpha}\lvert y-x\rvert^{\alpha-2}}1_{[0,x)}(y)+\frac{(3-\alpha)\,dy}{(1-x)^{2-\alpha}\lvert y-x\rvert^{\alpha-2}}1_{[x,1]}(y)$	$x(1-x)\int_0^1 \frac{u(y)-u(x)-u'(x)(y-x)}{(y-x)^2}\,d\Omega_x(y)$
Eldon–Wakeley processes	$(1-x)\delta_{x+(1-x)\lambda}(y)+x\delta_{x-\lambda x}(y)$	$\frac{1}{\lambda^2}(xu(x+(1-x)\lambda)-u(x)+(1-x)u(x-\lambda x))$
Λ-processes	$xd\left(\Lambda\left(\frac{x-y}{x}\right)\right)1_{[0,x]}(y)+(1-x)d\left(\Lambda\left(\frac{y-x}{1-x}\right)\right)1_{[x,1]}(y)$	$\int_0^1 \frac{1}{\lambda^2}(xu(x+(1-x)\lambda)-u(x)+(1-x)u(x-\lambda x))\,d\Lambda(\lambda)$
Minimal process	$\delta_0(y)$	$(1-x)u'(x)+\frac{1-x}{x}(u(0)-u(x))$
Maximal process	$\delta_1(y)$	$-xu'(x)+\frac{x}{1-x}(u(1)-u(x))$

variance σ^2, selection pressure γ, and mutation rates θ_1, θ_2. Then G has the form

$$Gu(x)=\gamma x(1-x)u'(x)+\frac{1}{2}(-\theta_1 x+\theta_2(1-x))u'(x)+\sigma^2 x(1-x)\int_0^1 \frac{u(y)-u(x)-u'(x)(y-x)}{(y-x)^2}\,d\Omega_x(y) \tag{34}$$

for some family of probability measures $\{\Omega_x, 0<x<1\}$. Additionally, for any G of the above form, there is a sequence of GWF chains $\mathbf{P}^{(N)}$ converging to G in the sense of (33).

Theorem 7 establishes that the continuum limit of a GWF process is characterized by a family of probability measures $\{\Omega_x\}$ on (0, 1), typically called a Levy kernel. In this representation Ω_x depends only on the pure-drift matrices $\mathbf{Q}^{(N)}$, and hence it reflects the contribution due solely to genetic drift. The decomposition of $\mathbf{P}^{(N)}$ in (10) is reflected in the limiting generator, G, as a sum of three terms each encoding separately the effects of selection, mutation, and drift.

The generator G encapsulates all information concerning the forward-time dynamics of a GWF process, and it can be used, at least in principle, to solve for genetic quantities of interest, including absorption times, fixation probabilities, stationary distributions, and the evolution of allele densities. The general theory of continuous-time Markov processes (Dynkin, 1965) provides the connection between these quantities and solutions to certain operator equations involving G. In this more general theory, the Kolmogorov forward and backward equations read

$$\frac{\partial \mu}{\partial t}=G^*\mu(x,t)$$
$$\frac{\partial u}{\partial t}=Gu(x,t) \tag{35}$$

where G^* is the adjoint operator, and they reduce to the standard diffusion equations only if G is a second-order differential operator. As an example, the absorption time $\tau(x)$ and probability of fixation $\pi(x)$, both started at a fixed fraction x of population size, are solutions to

$$G\tau(x)=-1 \tag{36}$$
$$G\pi(x)=0. \tag{37}$$

Besides giving a representation for the generators of GWF continuum limits, the proof of Theorem 7 also gives a practical method of computing the family of measures Ω in terms of the sequence of pure-drift transition matrices $\mathbf{Q}^{(N)}$. We have computed the generators for all of the GWF models discussed above, including the Wright–Fisher model, the Λ_1-process, the Eldon–Wakeley processes, the Λ-processes, and the power-law processes. The generators associated with each of these processes are shown in Table 1.

As Table 1 suggests, there is a close relationship between power-law processes and Λ-processes. In particular, a power-law process with $0<\alpha\leq 2$ has the same continuum limit as the Λ-process with probability density $\frac{d\Lambda}{d\lambda}=(3-\alpha)\lambda^{2-\alpha}$. In practice, this implies that the discrete power-law processes can be well approximated by the appropriate discrete Λ-models, which themselves are examples of Cannings processes. This result also demonstrates how GWF models constructed purely on the basis of its transition probabilities on the level of allele frequencies can be sometimes revealed to have a genealogical (i.e. particle) construction.

Theorem 7 provides a complete and powerful characterization of the continuum limits that arise from GWF processes. The set of such continuum limits comprise a fairly large space. For example, Kimura's classical diffusion model arises only in the extreme case of a family of delta measures: $d\Omega(x)=\delta_x(y)$. Roughly speaking, $d\Omega_x(y)$ has the physical interpretation as a family of "jump measures", which are propensities for the processes to make discontinuous changes from an originating state x to another state y. Since only the delta family of measures fails to assign mass to states other than x, it follows that Kimura's diffusion is the only continuum limit of a GWF process (with a given offspring variance) with continuous sample paths. This result is stated in the following theorem.

Theorem 8. *A continuum limit of a GWF process has continuous sample paths if and only if it is the Kimura diffusion.*

Fig. 1 provides visualizations of this theorem. In each of the three non-Wright–Fisherian processes, the sample paths clearly show discontinuities. We elaborate on an intuitive explanation for this result in the Discussion.

When specialized to the case of a continuum Cannings process, Theorem 7 provides a formula for the generator which complements Möhle's infinite-series representation of a Cannings generator (Möhle, 2001). Our formula also gives an alternative representation, in terms of a Levy kernel, for the two-type Λ-Fleming–Viot process (see Table 1), instead of the usual Choquet decomposition into Eldon–Wakeley processes. Each of these representations may have its advantages, depending on the application. For example, sample-path properties such as the continuity Theorem 8 are most easily established using the representation of Theorem 7. Our results on the fixation probability under selection required additional representations of the generator (see Appendix E for an example).

It may be natural to ask which GWF generators, posed in terms of the kernel Ω, are Cannings generators? While we cannot give a complete characterization at the present time, one necessary condition is that any putative Cannings generator G must map the space of polynomials to polynomials. From this, it can be seen that our power-law processes are Cannings if and only if $0<\alpha\leq 2$.

Equations such as (35)–(37) rarely admit closed-form solutions. Nonetheless, significant information can be obtained about the

nature of solutions without explicitly solving these equations. Indeed, many of our results on fixation probabilities in the preceding sections were obtained by an indirect analysis of these generalized Kolmogorov equations (see Appendices C–E).

5. Empirical plausibility of non-diffusive GWF processes

Surprisingly few studies have been concerned with direct empirical verification of the Wright–Fisher model. Of these, the most extensive and widely cited is the long-term experiment conducted by Peter Buri, a student of Sewall Wright. Buri maintained 107 independent laboratory populations of Drosophila, in which he tracked the frequencies of the bw^{75}/bw allele over each of 20 successive generations (Buri, 1956). In each population, successive generations were initiated with 16 flies, 8 male and 8 female, drawn randomly from among the F1 individuals in the preceding generation, with samples taken every 12 days (Buri's "Series I"). The frequency of the bw^{75} allele, initiated at 50% in generation 1, was subsequently recorded in generations 2 through 20 of each population. The time-series of these allele frequencies may be viewed as 107 independent sample paths of some underlying population process.

Buri designed his experiments using the bw^{75} allele, which was presumed to be neutral and non-mutating over the course of the experiment, to quantify the dynamics of genetic drift and to compare them to the Wright–Fisher model. Buri concluded that the dynamics observed were largely consistent with neutral drift, in the sense of Wright–Fisher. In fact, Buri's data have become the standard, textbook example of empirical support for the Wright–Fisher model of genetic drift (Hartl and Clark, 2006; Pierce, 2007). Here we revisit Buri's data and ask if they truly support the Wright–Fisher model to the exclusion of alternative forms of genetic drift such as those introduced above.

Buri restricted most of his analysis to the first and second moments of his empirical time series—that is, the dynamics of the mean allele frequency and the heterozygosity. From these statistics, Buri concluded that his data conformed to a neutral Wright–Fisher model, with an effective population size of $N_e = 18.0$ chromosomes. (A variance effective population size N_e is equivalent to $\sigma_N^2 = N/N_e$ in the notation above.) However, since his analysis was based on only the first two moments, whose dynamics are invariant across GWF models, all GWF models would be equally compatible with the data from Buri's perspective. In order to distinguish among GWF processes, it is necessary to inspect higher moments of Buri's data or, ideally, the full likelihood of the data under alternative models.

Genetic data are often presented in the form of allele counts from a sample of individuals in the population, at one point in time. In this scenario a coalescent analysis with the Kingman coalescent as a null model, or the exact coalescent of Fu (2006), would be appropriate. For example, Birkner and Blath (2007) inferred the exponent in the Beta-family of coalescents from empirical data in this manner. Buri's data, by contrast, comprise a time-series of the allele frequencies in the *entire* population, over subsequent generations. In this case, a coalescent analysis is not appropriate. Rather, such time-series data enable a more direct and more powerful likelihood analysis based on the transition matrix of the forward-time population process.

It is easy to compute the likelihood of independent sample paths drawn from a given Markov chain. Unfortunately, Buri did not report the actual sample path for each of his 107 populations, but rather only histograms summarizing the states of the (unmarked) populations in each generation. It is computationally unfeasible to compute the exact likelihood of these histograms, H_k ($k = 1, 2, \ldots, 20$), because of the astronomical number of sets of sample paths consistent with the histograms. Approximate Monte Carlo methods, including importance sampling techniques, are also problematic in this case. Nonetheless, for some functionals of the histograms we can still explore the relative fit of various GWF models via minimum-norm metrics, such as least squares.

Let dH_k be the measure on the set of possible allele frequencies $\{0, 1/32, \ldots, 1\}$ induced by the histogram at generation k. We consider the following two natural classes of functionals, for $p > 0$:

$$m_k(p) = \int_0^1 |x - 1/2|^p \, \mathrm{d}H_k(x) \tag{38}$$

$$h_k(p) = \int_0^1 x^p (1-x)^p \, \mathrm{d}H_k(x). \tag{39}$$

The quantities $m_k(p)$ and $h_k(p)$ are complementary: $m_k(p)$ are higher-order central moments and thus quantify the monomorphicity of a population, whereas $h_k(p)$ are higher-order measures of heterozygosity. If $\mathbf{Q}$ denotes the pure-drift transition matrix for a given GWF model, then[2] $\tilde{H}_k = (\mathbf{Q}^*)^k q_0$ gives the expected histogram at the kth generation, given the initial state q_0 of the populations. We define the functionals $\tilde{m}_k(p)$, $\tilde{h}_k(p)$ relative to $\tilde{H}_k$ similarly to (38), (39), and we consider the minimum Euclidean distance between the data and the expected functionals:

$$d_m = \sum_{k=1}^{T} |m_k(p) - \tilde{m}_k(p)|^2 \tag{40}$$

$$d_h = \sum_{k=1}^{T} |h_k(p) - \tilde{h}_k(p)|^2. \tag{41}$$

Here T denotes the total number of generations in the experiment, so that $T = 20$ in Buri's case. The residuals d_m and d_h serve as proxies for the likelihood, quantifying the fit between the empirical histograms and the expected model histograms, with respect to monomorphicity and heterozygosity.

For any given GWF model, there is a one-parameter family of similar models differing only in their variance effective population size, N_e. Since we are concerned with exploring precisely those aspects of Buri's data which lie beyond second-order statistics, we define the residual of an *equivalence class* of GWF processes differing only in N_e as the minimum residual over all possible variance scalings:

$$D_m = \min_{N_e} d_m \tag{42}$$

$$D_h = \min_{N_e} d_h. \tag{43}$$

We are interested in exploring the fit between Buri's data and various GWF processes. In particular, we consider the power-law processes, which are characterized by a parameter $\alpha > 0$ that quantifies the decay of the transition probability. As discussed above, the value $\alpha = \infty$ corresponds precisely to the Wright–Fisher model, but we have shown that processes with $\alpha > 3$ have the same, diffusive continuum limit as the Wright–Fisher model. Power-law processes with $\alpha < 3$, by contrast, are non-diffusive and qualitatively different from Wright–Fisher.

Table 2 shows the values of the residuals D_m and D_h for the Wright–Fisher process, as well as for the best-fit power-law process. Generally, the associated minimizing N_e's for these models lie in the range $N_e = 16$ to $N_e = 21$, roughly consistent with Buri's estimate of $N_e = 18$ based on the conditional second moment. For the functionals $m(2)$ and $h(1)$, only second-order statistics of the data are involved, hence all GWF models achieve

[2] $*$ is the adjoint operator.

Table 2
Residuals for Buri Histogram functionals.

Functional	Parameter p	Wright–Fisher residual	Best α	Best α residual	Best-fit type
Monomorphicity	0.5	8.7×10^{-3}	2.2	4.8×10^{-3}	Non-diffusive
Monomorphicity	1	5.9×10^{-3}	1.7	3.8×10^{-3}	Non-diffusive
Monomorphicity	2	1.8×10^{-3}	$\alpha > 0$	1.8×10^{-3}	All equivalent
Monomorphicity	3	4.7×10^{-4}	∞	4.7×10^{-4}	Diffusive
Heterozygosity	0.5	2.7×10^{-2}	∞	2.7×10^{-2}	Diffusive
Heterozygosity	1	7.1×10^{-3}	$\alpha > 0$	7.1×10^{-3}	All equivalent
Heterozygosity	2	4.3×10^{-4}	0.7	1.8×10^{-4}	Non-diffusive
Heterozygosity	3	2.5×10^{-5}	1.1	1.2×10^{-5}	Non-diffusive

the same fit. As Table 2 shows, in some respects (i.e. the generalized central moment for p large) Buri's data are most consistent with the Wright–Fisher model; whereas in other respects (i.e. other central moments, and higher-order measures of heterozygosity) they are more consistent with non-diffusive power-law models.

In addition to the technique summarized in Table 2, we have also analyzed Buri's data by inspecting the time-series of absorption events (that is, the times at which populations became monomorphic). In this case, we can perform an exact likelihood analysis by computing the probability of the absorption time-series for any GWF model. For Buri's data, we find that the best-fit Wright–Fisher model ($N_e = 19.8$) has a higher likelihood than any non-diffusive power-law model, but the difference in log-likelihoods ($\Delta ll = 0.61$) between the Wright–Fisher model and the diffusive boundary $\alpha = 3$ is too small to reject one model in favor of the other with a tolerable false positive rate or reasonable Bayes factor.

Even though Buri's experiment was designed to test the validity of the Wright–Fisher model, and it has long been interpreted as having done so, we have seen that the data are equally, or, in certain features, more consistent with non-diffusive models that are fundamentally different than the Wright–Fisher model. In retrospect, this result is not completely unexpected. Indeed, Buri's experiment does not exactly match the conceptual, Wright–Fisherian mechanism of sampling with replacement; rather, it is more similar to Schweinsberg's model of sampling without replacement (Schweinsberg, 2003). The Schweinsberg model is known to produce non-Kingman coalescents (and hence non-diffusive forward processes) when the offspring distribution is heavy-tailed. It is plausible that in each generation of Buri's experiment, some flies, regardless of their genotype, may have had a disproportionally large number of offspring, perhaps because they encountered a greater amount of resources, for example. The resulting heavy-tailed offspring distribution (relative to population size) may have resulted in higher-order features that are better described by non-diffusive processes than by the Wright–Fisher process.

Our results suggest that neither the standard Wright–Fisher model nor any single power-law process consistently provides the best fit to all functionals of Buri's data. Rather, Buri's data may be best explained by some mixture of non-diffusive and diffusive components.

6. Discussion

We have introduced a broad class of models that describe the dynamics of alleles in a population of constant size. The set of Generalized Wright–Fisher models share the same conditional first and second moments as the standard, Wright–Fisher model, but they may otherwise differ. We have characterized the behavior of classical quantities, such as absorption times and fixation probabilities, across the full range GWF models. Specific examples of generalized models (e.g. the power-law processes) produce strikingly different behavior from the classical models—including, for example, the assured asymptotic fixation of new mutants, even introduced at low frequencies, provided the selective advantage, Ns, exceeds a finite threshold. We have also characterized the continuum limits of GWF processes, analogous to Kimura's diffusion limit of the Wright–Fisher process. Finally, we have demonstrated that some of the non-diffusive GWF processes are equally likely, or more likely, than the Wright–Fisher model, given empirical data on allele frequency dynamics from Drosophila experiments, which were originally designed to test the validity of the Wright–Fisher model. Taken together, these results compel us to critically re-assess the widespread reliance on standard, diffusive models in theoretical and applied population genetics.

Part of our motivation for studying generalized population processes has been to determine which aspects of the Wright–Fisher model are robust to alternative forms of genetic drift or offspring mechanism—as encoded, for example, by the exchangeable variable ν in a Cannings model. For neutral processes without mutation, many classical quantities of interest remain completely unchanged by such perturbations: those quantities that are functions solely of the lower-order conditional moments, including fixation probabilities and the first two unconditional moments. These invariances even hold in the presence of mutation.

Other quantities, however, vary across the set of GWF models. Neutral absorption times for an allele initiated at intermediate frequencies and conditional fixation times arbitrarily initiated vary in a relatively narrow range; whereas absorption and extinction times of a new mutant vary dramatically. Furthermore, enormous deviations from the standard model arise in the presence of selection, regardless of the initial allele frequency, as evidenced by variation in fixation probabilities over many orders of magnitude.

These phenomena have a unified, intuitive explanation, which we describe via a thought-experiment. Consider the probability distribution comprised by a row of the drift matrix, $\mathbf{Q}^{(N)}$. The center row of the Wright–Fisher process, for example, is a discrete symmetric distribution around $N/2$, approximately Gaussian in shape. We can imagine deforming this probability distribution while maintaining its mean. Two directions of deformation are possible: probability mass may be moved to the tails of the distribution, or further concentrated near the mean. The additional constraint of a fixed variance (which is enforced by the definition of the GWF process) implies that for any mass moved to the extremes, a compensating mass must be concentrated near its center, and conversely. Thus, the two requirements of a GWF process (6) and (7) imply a range of distributions, some with heavier tails than the Wright–Fisher model, and correspondingly more strongly peaked about their mean, and some with lighter tails and with less probability near their mean. These two directions of deformation are associated with increases and decreases, respectively, of the higher central moments of the distribution.

Deforming the Wright–Fisher model by reducing the size of higher moments will have virtually no effect on the resulting behavior, because this only improves the accuracy of Kimura's standard diffusion approximation. Hence non-standard behavior occurs only for deformations that broaden the tails, which require

Table 3
Selected bounds and examples.

Genetic quantity	Bounds	Wright–Fisher model	Λ_1-model	Power-law models		
				$\alpha < 2$	$\alpha = 2$	$2 < \alpha < 3$
Mean Abs. Time, $X_0 = N/2$	$O(N)$ to $O(N \log N)$	$(2 \log 2)N$	N	?	?	?
Mean Abs. Time, $X_0 = 1$	$O(\log N)$ to $O(N)$	$2 \log N$	N	$\geq O(N/\log N)$	$O(N/\log N)$	$\geq O(N^{3-\alpha})$
Con. Fix. Time, $X_0 = N/2$	$O(N)$ to $O(N \log N)$	$(2 \log 2)N$	N	?	?	?
Con. Fix. Time, $X_0 = 1$	$O(N)$ to $O(N \log N)$	$2N$	N	?	?	?
Asymp. Prob. Fix., $X_0 = 1$	s to 1	$\frac{1-\exp(-2s)}{1-\exp -2Ns} \approx 2s$	1	1	?	?
Asymp. Prob. Fix., $X_0 = [xN]$	$\pi_-(x)$ to $\pi_+(x)$	$\frac{1-\exp(-2Nsx)}{1-\exp(-2Ns)}$	$\pi_1(x)$	$\geq x^{p_r}$	?	?
Regime of Assured Fixation	$1 \leq Ns \leq \infty$	$Ns = \infty$	$Ns = \infty$	$Ns \geq \gamma_0$	$Ns = \infty$?	$Ns = \infty$?

For the definitions of $\pi_\pm$, π_1, p_r, γ_0, refer to Theorem 6, Eq. (30), Theorem 9 (Appendix E) and Eq. (E.5) respectively. Entries with question marks represent unknown or conjectural properties at the time of writing.

a concomitant concentration of mass near a distribution's mean. Fig. 7 illustrates this principle by plotting the center row of the matrix $\mathbf{Q}^{(N)}$ associated with several GWF processes. In each case, the distributions of non-diffusive models are both more strongly peaked and have heavier tails than the Wright–Fisher model.

Strongly peaked transition distributions, as shown in Fig. 7, produce longer holding times. That is, the time that an allele spends at a given frequency before moving to another frequency is lengthened. The holding time ranges from $O(1)$ generations in the Wright–Fisher model to $O(N)$ generations in the Λ_1-model. This enormous variation accounts, in large part, for the great sensitivity of the mean absorption time when initialized near an absorbing state, since such times are dominated by the holding period in the initial state. Conversely, when started at intermediate frequencies, initial holding times contribute only a small fraction to the overall absorption time—which explains why absorption times are more stable across models when initiated at intermediate frequencies. This intuition also explains the robustness of the conditional fixation times for new mutants (Theorem 4): despite initialization near the lower boundary, conditioning on fixation effectively removes the absorbing state at zero, and once again the initial holding period contributes negligibly to the overall fixation time.

What does a lengthy holding time in a neutral model correspond to in a model with selection? When selection is added, a model with strongly peaked transition distributions, in $\mathbf{Q}^{(N)}$, will feature lengthy periods of nearly deterministic, logistic growth of the advantageous allele, in $\mathbf{P}^{(N)}$. This behavior explains why non-diffusive GWF processes amplify the effects of selection. In the Wright–Fisher model, a newly introduced advantageous mutant must reach appreciable frequency quickly (within $O(1)$ generations) in order to escape elimination by drift. In a power-law model, by contrast, such a mutant enjoys at least $O(N^{3-\alpha})$ generations (for $2 < \alpha < 3$) before drift threatens extinction; and the Λ_1-process provides $O(N)$ generations of deterministic, logistic growth. These lengthy periods of nearly deterministic growth for a new mutant serve to amplify the effects of selection, increasing the chance of fixation compared to the Wright–Fisher model.

The same intuition about the shape of transition distributions has important implications for the continuum limits of GWF processes. Some subsets of GWF models converge to Kimura's diffusion, whereas others converge to alternative attractors. Whereas Kimura's diffusion has *continuous* sample paths and is parameterized only by the first two moments, all non-standard GWF continuum limits are non-diffusive—i.e. their sample paths contain jumps. Thus, non-diffusive GWF models feature two unusual elements: lengthier holding times that produce periods of stasis or near-determinism, punctuated occasionally by large changes in allele frequency arising from the heavy tail.

Biologically, such discontinuities arise whenever a few individuals contribute to an appreciable fraction of the total population in the following generation. This observation has its

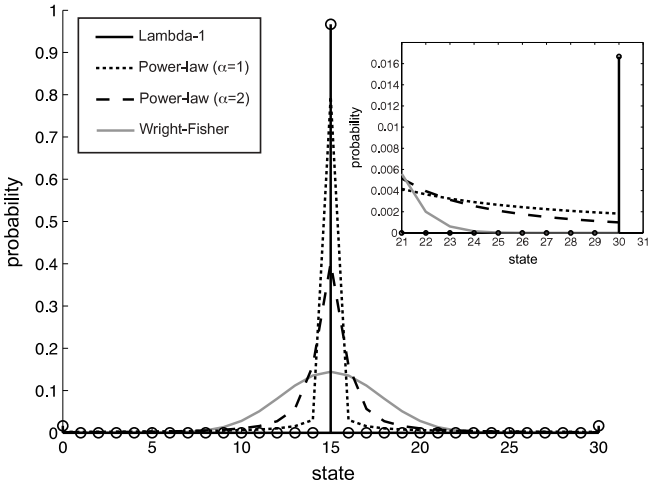


Fig. 7. The center row in the transition matrix for various GWF processes. The graph shows the values of $Q_{16,j}$ for $j = 0, 1, \ldots, N = 30$, for the Wright–Fisher model, power-law model with $\alpha = 2$, power-law model with $\alpha = 1$, and Λ_1-model. The Wright–Fisher model has the lightest tails and it is least centrally peaked, while the Λ_1-process is most centrally peaked with the heaviest tails. Power-law processes are intermediate.

correlate in the retrospective theory, where the merging of more than two ancestral lines in a sample becomes possible in non-Kingman coalescents. The Λ_1-process is the most extreme in this respect, because its sample path begins with a constant allele frequency and terminates in a violent jump to absorption. Correspondingly, its backward-time version, the $\Lambda_1 n$-coalescent, consists of n individuals who, after some exponential waiting period, merge into a single lineage. The power-law processes are intermediate to these (see Fig. 7): as mentioned above, for $0 < \alpha < 2$ they may be interpreted as Λ models, and thus are mixtures of Eldon–Wakeley models, where the replacement fraction λ is random and drawn from the probability density $d\Lambda/d\lambda = (3-\alpha)\lambda^{2-\alpha}$. For small α, this density has more mass near one than near zero, exhibiting behavior similar to the Λ_1-process; as α increases to $\alpha = 3$ the behavior approaches the Wright–Fisher model, eventually coinciding with the standard model in the continuum limit.

6.1. The Wright–Fisher process is an extreme model

Table 3 summarizes several bounds we have derived for standard population-genetic quantities under GWF processes, as well as the values attained in particular cases. This once again validates the notion that Wright–Fisher and Λ_1-processes are at antipodes, with other GWF processes intermediate.

A surprising picture emerges from the results summarized in Table 3: the Wright–Fisher process is the actual- or near-minimizer for neutral absorption times, and the strongest suppressor of

selection among all GWF models. In other words, far from being located in the center of all generalized population processes, the classical model falls at the extreme of this space. Roughly speaking, the Wright–Fisher process exhibits the strongest form of genetic drift amongst all models of a given offspring variance. This accounts for its extremely short absorption times near the boundaries, both with and without mutation, the continuity of its sample paths, the very short holding time of states, and its strong suppression of selection.

The extremity of the Wright–Fisher process can be explained from the point of view of continuum theory and the representation result of Theorem 7. The Levy kernel associated to Kimura's diffusion is a family of delta measures, and hence it is "pure" in the sense that it cannot be written as a non-trivial convex combination of other kernels; this also implies that its transition matrix likewise cannot be written as a mixture of two other matrices (in the limit of large population size). Thus, the Wright–Fisher process is extreme in a precise geometrical way: it lies on the boundary of the whole space of GWF processes. It can be shown that the maximal and minimal processes are also extreme in this sense.

Whereas the generator G for Kimura's diffusion is a second-order differential operator, this is not necessarily so for GWF processes in general. The power-law decay processes, for example, correspond to generators of fractional order $\alpha - 1$ (for $0 < \alpha \leq 3$). When $\alpha < 2$, this order is smaller than one, meaning that the selective term of order one dominates the strength of drift in (34). This provides a mathematical explanation for the phenomenon of assured fixation that occurs in these models: the strength of drift, as measured by the order of its generator, becomes weaker than the deterministic force of selection.

6.2. Relationship to non-diffusive coalescents

Our present work can, in part, be seen as the construction and detailed analysis of forward-time analogues of general reproductive models whose backward-time properties have recently received much attention. As mentioned above, an arbitrary GWF process contains information only about allele frequencies, and it does not include any information about parentage as would be required for constructing a coalescent theory. However, GWF processes include the large class of Cannings processes, which come equipped with an exchangeable offspring distribution and whose genealogical properties can therefore be studied.

The genealogical properties of Cannings' processes are more diverse than those of the Wright–Fisher process alone, whose geneologies are described by Kingman's classical coalescent in the limit of large N. In the Kingman coalescent, at most two lineages can merge at a given point in time. The probabilistic construction of coalescents that admit multiple (three or more) mergers was introduced in Pitman (1999) via the so-called Λ-coalescents. These processes were recognized by Sagitov (1999) as the continuum limits, in backward time, of certain discrete Cannings processes. In addition, Sagitov, Moehle and Schweinsberg classified all the limiting coalescents that arise from Cannings models, resulting in the so-called Ξ coalescents that allow simultaneous multiple mergers. A satisfying and complete duality in the space of Cannings processes was demonstrated by Möhle (2001), who proved that a Cannings process converges in forward time to Kimura's diffusion if and only if its backward-time genealogies converge to Kingman's coalescent.

In Theorem 8, we proved that Kimura's diffusion is the only continuum limit of a GWF process that exhibits continuous sample paths. In conjunction with Moehle's duality, this establishes an additional correspondence: the presence of jumps in the allele frequency of a forward-time Cannings process is equivalent to the possibility of multiple mergers in its backward-time coalescent. In fact, these two phenomenon have a common source: the ratio of higher moments in the offspring distribution to the second moment. This ratio measures the heaviness of the tails of the offspring distribution; when the ratio is large enough, the forward-time continuum limit is not a diffusion and the backward-time coalescent features multiple mergers. Biologically, both phenomena are caused by a reproductive mechanism in which very large family sizes are occasionally possible.

It is worth emphasizing that the qualitative properties of the continuum limit for a Cannings process – that is, whether the forward-time limit is Kimuran or not, and whether the backward-time limit is Kingman or not – depend on the *ratio* of its higher moments to its second moment, rather than on whether σ_N^2 converges or not with large N. The convergence of the offspring variance determines only the natural time scale under which to take the continuum limit: when $\sigma_N^2 \to \infty$, the natural time scale becomes shorter than N generations, and conversely when $\sigma_N \to 0$.

Indeed, each one of our GWF continuum limits, and every generalized Ξ coalescent, can be realized as a limit of discrete processes whose sequence of σ_N^2 behaves arbitrarily, by appropriate choice of time scale when taking the continuum limit. From the biological point of view, then, the important quantity in deciding between the use of a classical model versus a generalized model is not that of large offspring variances, but rather the relative size of higher moments of the offspring distribution to its variance.

Furthermore, we note that the optimal bounds we have derived for the behavior of GWF processes are often achieved by specific Cannings processes. In other words, even though the Cannings' processes are a strict subset of GWF processes, they are fairly representative of the full range of behavior possible across the broader class of GWF processes. Indeed, many (symmetric) GWF processes are either Cannings themselves, or they can be closely approximated by a Cannings process (for example, the continuum power-law processes with $0 < \alpha \leq 2$ coincide with the Λ-Cannings models with an appropriate choice of Λ-measure).

6.2.1. Coming down from infinity

A problem of interest in contemporary coalescent theory is the question of whether a given generalized (continuum) coalescent "comes down from infinity"—i.e. whether there are only finitely many blocks in the partition almost surely for all positive times, when the coalescent is initialized in a partition with an infinite number of blocks. In particular, it is known that the Kingman coalescent does come down from infinity, while the Bolthausen–Sznitman coalescent (corresponding to our power-law model $\alpha = 2$), does not (Schweinsberg, 2000). It is often presumed that those coalescents that do not come down cannot form plausible biological models, since this indicates that the time to most recent common ancestor for the entire population of N individuals is infinite (Birkner and Blath, 2009). The analysis conducted in this paper suggests that this view may be exaggerated, resulting from a conflation between continuum and discrete models. In a generalized process for which a time-reversibility argument is available, the conditional fixation time of the forward process started at one individual corresponds to the time to most recent common ancestor for the whole population. Theorem 4 shows that not only is the former finite in the discrete process, but has expectation bounded by $O(N \log N)$ for any model with asymptotically positive and finite offspring variance. This suggests that those coalescents that do not come down from infinity are simply those whose conditional fixation times scale faster than the population size N. However, the coalescence times could only range between $O(N)$ and $O(N \log N)$—a relatively narrow range, in practice, which is unlikely to serve as the basis for questioning the biological plausibility of a model.

6.3. Fixation of an allele under selection

Our results on the fixation probability of a mutant allele under selection represent a striking departure from what is otherwise known about fixation probabilities in population-genetic models. In particular, we have shown that, across a broad range of models, a new mutant introduced even at low frequency may be assured fixation, asymptotically, despite finite Ns and despite the presence of non-trivial genetic drift (i.e. despite a finite variance effective population size). The derivation of these results required an understanding of the continuum limits of generalized forward-time processes.

It may eventually be possible, however, to understand some of these results within a coalescent framework. In general, coalescent theory cannot easily incorporate selection. The problem is fundamental to genealogical frameworks, which rely on characterizing the space of possible genealogical trees before considering the possibility of mutations at various locations on these trees. When selection operates, the probabilities of particular trees cannot be defined independently of the mutations, and the approach breaks down (Tavare, 2004; Wakeley, 2009). The ancestral selection graph of Neuhauser and Krone (1997) and Krone and Neuhauser (1997) provides an elegant formal solution to this problem, but unfortunately it requires extensive numerical calculations that limit the intuition we can draw (Przeworski et al., 1999). However, very recent work of Etheridge and Griffiths (2009) and Etheridge et al. (2010) have produced coalescent dual processes in Moran models with selection and continuous-time Moran models with viability selection. When generalized to accommodate a broader class of underlying population processes, these techniques may also shed light on the fixation of alleles under selection in generalized population processes.

6.4. Remarks on the neutral theory

Having introduced a large class of generalized population models, with and without selection, it is useful to reconsider the development of Kimura's neutral theory of molecular evolution. Kimura's theory emphasized the importance of genetic drift, as opposed to Darwinian selection, in shaping observed patterns of genetic variation within populations and substitutions between divergent species. However, Kimura's neutral theory, and much of the intuition behind it, relied heavily on the specific form of genetic drift inherent to the Wright–Fisher model and its diffusion limit. Consideration of the broader class of GWF processes introduced here suggests that drift is typically much weaker in alternative, plausible models of biological populations. In particular, we have described a large class of biologically reasonable, non-diffusive processes that contain a relatively weak form of genetic drift and that exhibit long periods of quasi-determinism. Such models have the effect of amplifying even minute selective forces. Thus, to the extent that these alternative models are sometimes more accurate depictions of real populations, our investigation can be seen as diminishing the role of genetic drift in shaping the fate of mutant alleles.

Nevertheless, the strongest argument for the neutrality of most observed genetic substitutions has centered on the perceived existence of a "molecular clock"—a nearly constant rate of genetic substitutions observed, per unit of time (in years), across a diverse range of species with different population sizes (Kimura, 1994). Generation-time issues aside, this kind of clock-like behavior would be predicted by the Wright–Fisher model for neutral sites, because their substitution rate equals the mutation rate, μ, independent of N. By contrast, the corresponding rate of substitution for selectively advantageous mutants is $\approx 2N\mu s$, a quantity that is unlikely to remain constant over time or across species with different population sizes.

These arguments can be re-examined in the context of our more general population models. As with the Wright–Fisher model, the rate of neutral substitution in any GWF process is simply $N\mu\frac{1}{N} = \mu$. As discussed above, deleterious mutations in GWF processes may have larger probabilities of survival than in the classical scenario, but only marginally so, and certainly fix with rates that are orders of magnitude lower than the neutral rate μ. If novel, advantageous mutations with fitness advantage s fix with probability $\pi(N, s)$ under some GWF model, then the rate of such substitutions is then $N\mu\pi(N, s)$. If this quantity is to remain constant, independent of population size, we must assume $\pi = g/N$, for g with no functional dependence on N. But the lower bound in (29) reads $\pi \geq s$, or $g \geq Ns$, which contradicts the independence of g on N. Thus, the theory of generalized processes would only further support the hypothesis of pre-dominantly neutral or near-neutral evolution, at least to the extent that the molecular clock exists and is independent of population size and species (but see Li (1993)).

6.5. Implications for the inference of selection and mutation

Population geneticists often infer selection pressures and mutation rates from sampled genetic data by comparison to the steady state distribution of allele frequencies under mutation, selection, and drift. Such comparison typically assumes an infinite number of unlinked loci (Sawyer and Hartl, 1992; Bustamante et al., 2001; Desai and Plotkin, 2008), and may utilize the Poisson Random Field approximation. We have developed an analogous theory of steady state distributions for GWF processes, and its application to inference techniques will be described elsewhere.

Nonetheless, a few qualitative insights about inference are implied already in the present work. We have seen that the Wright–Fisher model is extreme in the sense that, amongst a large class of population processes, it tends to minimize selective differences among alleles due to the powerful nature of its genetic drift. Correspondingly, if the underlying drift process is not necessarily Wright–Fisherian, then estimates obtained under a Wright–Fisher assumption are expected to produce *only upper bounds* on the true strength of selection. Any particular pattern of data suggesting the presence of strong selection, under Wright–Fisher, could conceivably be explained by a smaller value of s and a weaker, alternative form of drift. Therefore, one expects that replacing Wright–Fisherian drift with other types of GWF drift will have the consequence of moving estimated selection values closer to zero. This further suggests that under all non-Wright–Fisher models, a greater standard of evidence would be required to reject the null assumption of neutrality.

Acknowledgments

We thank Warren Ewens for many detailed and encouraging discussions, as well as insightful input throughout the course of this research. The authors also recognize the close reading of the manuscript by two anonymous reviewers, and their valuable commentary which improved the manuscript. JBP acknowledges funding from the Burroughs Wellcome Fund, the David and Lucile Packard Foundation, the James S. McDonnell Foundation, and the Alfred P. Sloan Foundation. RD, CLE, and JBP also acknowledge support from the Defense Advanced Research Projects Agency (HR0011-05-1-0057 and HR0011-09-1-0055).

Appendix A. Mutation and selection matrices

The mutation matrix $\mathbf{M}^{(N)}$ can be specified as follows: after reproduction, each individual mutates to the opposing type independently with respective probabilities μ_1, μ_2. If we define $\mathbf{M}_{jk}$ as the probability that, starting with j individuals of type A, one obtains k individuals of type A after mutation, then $\mathbf{M}_{jk}$ is distributed according to a sum of independent binomial random variables X_j and Y_j:

$$P(X_j = r) = \binom{j}{r}(1-\mu_1)^r \mu_1^{j-r} \tag{A.1}$$

$$P(Y_j = r) = \binom{N-j}{r}\mu_2^r (1-\mu_2)^{N-j-r} \tag{A.2}$$

$$\mathbf{M}_{jk} = P(X_j + Y_j = k). \tag{A.3}$$

The selection matrix $\mathbf{S}^{(N)}$ can be chosen with the property that its ith row is a binomial distribution supported on states $\{i, \ldots, N\}$, and a probability parameter such that the mean of the distribution coincides with the standard selected binomial Wright–Fisher model of parameter s. It is simple to see, by Taylor expansion, that the preceding choices for $\mathbf{M}$ and $\mathbf{S}$ satisfy the constraint (11) and (12).

Appendix B. The minimal and maximal GWF processes

We define $\mathbf{Q}^-$ as the pure-drift transition matrix:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & & & 0 \\ x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & 0 & \cdots & & 0 \\ x_1^{(2)} & 0 & x_2^{(2)} & x_3^{(2)} & 0 & \cdots & 0 \\ \vdots & & & & & & \\ 0 & \cdots & & & & 0 & 1 \end{pmatrix} \tag{B.1}$$

whose entries are

$$x_1^{(i)} = \frac{N-i}{N(i+1)}, \qquad x_2^{(i)} = \frac{i}{N}, \qquad x_3^{(i)} = \frac{N-i}{N} - \frac{N-i}{N(i+1)}. \tag{B.2}$$

Define $\mathbf{Q}^+$ as the transition matrix:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ x_1^{(1)} & x_2^{(1)} & 0 & \cdots & 0 & x_3^{(1)} \\ 0 & x_1^{(2)} & x_2^{(2)} & 0 & \cdots & x_3^{(2)} \\ 0 & 0 & x_1^{(3)} & \cdots & & \\ \vdots & & & & & \\ 0 & \cdots & & & 0 & 1 \end{pmatrix} \tag{B.3}$$

with entries

$$x_1^{(i)} = \frac{i}{N} - \frac{i}{N(N-i+1)}, \qquad x_2^{(i)} = 1 - \frac{i}{N}, \qquad x_3^{(i)} = \frac{i}{N(N-i+1)}. \tag{B.4}$$

We refer to the GWF processes induced by the transitions $\mathbf{Q}^{\pm}$ as the maximal and minimal processes, respectively. They play an important technical role in our continuum theory, by delimiting two extreme boundaries. For example, the proof sketch to Theorem 6 below shows that these two models minimize and maximize the mean, and thus the fixation probability, given an initial allele frequency, across all GWF processes with selection; they operate as the ultimate selection amplifiers and suppressors. Additionally, it can be demonstrated that their neutral absorption time is $O(N \log N)$, and thus achieve the upper bounds in Theorems 3 and 4.

Appendix C. Proofs of theorems

C.1. Theorem 2

This is a consequence of the martingale property of neutral processes without mutation, and is immediate from the stopping-time theorem.

C.2. Theorem 3

We begin by proving the lemma:

Lemma 1. *For a Generalized Process X_k, started at X_0, and with $\sigma_N^2 = 1 - 1/N$, we have the estimates for the absorption time τ:*

$$P(\tau > k) \geq \frac{X_0(N-X_0)\alpha^k(2-\alpha^k)}{(N-\alpha^k(N-X_0))(N-X_0\alpha^k)} \tag{C.1}$$

$$P(\tau > k) \leq \min\left\{1, \frac{X_0(N-X_0)}{N-1}\alpha^k\right\} \tag{C.2}$$

where $\alpha = 1 - \frac{1}{N}$.

Proof. We start with the lower bound, beginning with:

$$P(\tau \leq k) = P(X_k = N \cup X_k = 0) = P(X_k = N) + P(X_k = 0). \tag{C.3}$$

The first term can be estimated by the Glivenko–Chebyshev inequality: if Y is a random variable of mean zero, and $\lambda > 0$,

$$P(Y \geq \lambda) \leq \frac{EY^2}{EY^2 + \lambda^2}. \tag{C.4}$$

Putting $Y = X_k - X_0$ and $\lambda = N - X_0$, one determines

$$P(X_k = N) \leq \frac{Var(X_k)}{Var(X_k) + (N-X_0)^2}. \tag{C.5}$$

From the Cauchy–Schwarz inequality applied to $X_k 1_{X_k > 0}$ the second term of (C.3) satisfies:

$$P(X_k = 0) \leq \frac{Var(X_k)}{Var(X_k) + X_0^2}. \tag{C.6}$$

Inserting this into (C.3), and substituting the moment formulas (15) and (16), gives

$$P(\tau > k) \geq \frac{X_0^2(N-X_0)^2 - [Var(X_k)]^2}{(X_0^2 + Var(X_k))((N-X_0)^2 + Var(X_k))}, \tag{C.7}$$

which is algebraically equivalent to (C.1).

For the upper bound, Markov's inequality implies that if $|Y| \leq K, 0 \leq \lambda < K$, then

$$P(|Y| \leq \lambda) \leq \frac{K^2 - EY^2}{K^2 - \lambda^2}. \tag{C.8}$$

Since $Y_k = X_k - N/2$ is bounded by $K = N/2$, we have

$$P(|X_k - N/2| \leq \lambda) \leq \frac{N^2/4 - E(X_k - N/2)^2}{N^2/4 - \lambda^2} = \frac{-EX_k^2 + N \cdot EX_k}{N^2/4 - \lambda^2}. \tag{C.9}$$

Inserting $\lambda = N/2 - 1$, and applying the recurrence solutions (15) and (16), one obtains the desired outcome:

$$P(\tau > k) = P(|X_k - N/2| \leq N/2 - 1) \leq \frac{X_0(N-X_0)}{N-1}\alpha^k. \quad \square \tag{C.10}$$

By summing over the estimates (C.1), (C.2), and evaluating the asymptotics of the resulting expressions, one finds the mean absorption time bounds described in Theorem 3.

C.3. Theorem 4

The Markov chain X_k^* *conditioned on ultimate fixation*, defined on the state space $\{1, \ldots, N\}$ has the transition matrix:

$$\mathbf{Q}_{ij}^* = \mathbf{Q}_{ij} \frac{j}{i} \tag{C.11}$$

where $\mathbf{Q}_{ij}$ is the unconditioned transition. X^* has a single absorbing state at N; let τ^* be the time to fixation, when started at X_0^*. We first demonstrate the following estimates on the distribution of τ^*:

Lemma 2.

$$\frac{N - X_0}{N-1}\alpha^k \le P(\tau^* > k) \le \min\left\{1, (N - X_0^*)\alpha^k\right\} \tag{C.12}$$

where $\alpha = 1 - 1/N$.

Proof. From (C.11) we derive the conditional mean relation:

$$\sum_j j\mathbf{Q}_{ij}^* = \frac{1}{i}\sum_j j^2\mathbf{Q}_{ij} = \frac{1}{i}(i(1 - i/N) + i^2) = 1 + \alpha i \tag{C.13}$$

where $\alpha = 1 - 1/N$. It follows by taking expectations that the mean of the process satisfies the recurrence

$$EX_{k+1}^* = 1 + \alpha EX_k^* \tag{C.14}$$

whose solution is

$$EX_k^* = \frac{1 - \alpha^k}{1 - \alpha} + \alpha^k X_0^*. \tag{C.15}$$

Applying Markov's inequality on the positive process $X_k^* - 1$, one derives

$$P(\tau^* \le k) = P(X_k^* = N) \le \frac{EX_k^* - 1}{N - 1}. \tag{C.16}$$

Finally, after substituting the recurrence solution (C.15) and taking complements, find the lower bound

$$P(\tau > k) \ge \alpha^k \frac{N - X_0^*}{N - 1}. \tag{C.17}$$

To discover an upper bound, apply Markov's inequality on the positive process $N - X_k^*$:

$$P(N - X_k^* \ge 1) \le N - EX_k^* \tag{C.18}$$

so that

$$\begin{aligned} P(\tau^* > k) &= 1 - P(X_k^* > N - 1) = P(X_k^* \le N - 1) \\ &\le N - EX_k^* = \alpha^k(N - X_0^*) \end{aligned} \tag{C.19}$$

which implies the upper bound of the theorem. □

Again, as in Theorem 3, by summing over these estimates, we derive the mean conditional fixation bounds of Theorem 4.

C.4. Theorem 5

Let one-way mutation occur at rate $\mu > 0$. A computation with the mutation matrix of (A.3) shows

$$E(X_{k+1}|X_k) = X_k(1 - \mu) \tag{C.20}$$

$$E(X_{k+1}^2|X_k) = cX_k + \left(1 - \frac{\sigma^2}{N-1}\right)(1 - \mu)^2 X_k^2 \tag{C.21}$$

where $c = \mu + N\sigma^2/(N-1) \cdot (1 - 2\mu)$. Taking the expectation of both sides, the lower moments satisfy the following recurrences:

$$EX_{k+1} = vEX_k \tag{C.22}$$

$$EX_{k+1}^2 = cEX_k + v^2\left(1 - \frac{\sigma_N^2}{N-1}\right)EX_k^2 \tag{C.23}$$

where $v = 1 - \mu$. Solving the recurrences produces

$$EX_k = v^k X_0 \tag{C.24}$$

$$EX_k^2 = c\frac{v^{k-1}X_0}{1 - v\lambda_N}(1 - (v\lambda_N)^k) + (v^2\lambda_N)^k X_0^2 \tag{C.25}$$

with $\lambda_N = 1 - \frac{\sigma_N^2}{N-1}$. Using the Cauchy–Schwarz inequality, one finds

$$P(X_k = 0) \le \frac{EX_k^2 - (EX_k)^2}{EX_k^2}. \tag{C.26}$$

Thus if τ is the time to absorption at 0,

$$P(\tau > k) \ge \frac{(EX_k)^2}{EX_k^2} = \frac{v^k X_0}{C(1 - (v\lambda_N)^k) + X_0(v\lambda_N)^k} \tag{C.27}$$

where $C = \frac{cv^{-1}}{1 - v\lambda_N}$. The preceding expression has the lower bound

$$\int_0^\infty \frac{X_0(v\lambda_N)^x}{C(1 - (v\lambda_N)^x) + (v\lambda_N)^x X_0}\,dx = \frac{X_0(\log(X_0/C))}{-[\log(v\lambda_N)](X_0 - C)}.$$

For large N, this bound reads

$$E\tau \ge \frac{NX_0\log(X_0/(kN))}{(\theta/2 + \sigma^2)(X_0 - kN)} \tag{C.28}$$

where $k = \frac{\sigma^2}{\theta/2 + \sigma^2}$. This in turn reduces to the asymptotic expressions in Theorem 5, in the special cases $X_0 = 1$ and $X_0 = [xN]$.

C.5. Theorem 6

Only an informal outline of the proof is given here. We assume the representation for the generator G of a GWF process in Theorem 7. Define the operator $Lu(x,t) = Gu(x,t) - \frac{\partial u(x,t)}{\partial t}$. As discussed in the section on continuum theory, the mean $m(x,t)$ of a process, as a function of initial allele frequency x and time t, satisfies the generalized Kolmogorov equation $Lm(x,t) = 0$, with the boundary conditions $m(x,0) = x$, $m(0,t) = 0$, and $m(1,t) = 1$.

Basic to the proof is an application of the maximum principle, which is a property of generators of random processes. We look at functions $u(x,t)$ in the rectangle $R = \{(x,t) \in [0,1] \times [0,T]\}$, and label three of the respective sides $S_1 = [0,1] \times 0$, $S_2 = 0 \times [0,T]$, $S_3 = 1 \times [0,T]$. It can be shown that the following holds (Der, 2010).

Lemma 3 (*Parabolic Maximum Principle*)**.** *Let* u *be continuous on* R *and suppose* $Lu \ge 0$ *throughout the interior of* R*. Then*

$$\max_R u = \max_{S_1 \cup S_2 \cup S_3} u. \tag{C.29}$$

That is, the maximum value that any function u takes which satisfies the operator inequality $Lu \geq 0$ is achieved on the "parabolic boundary" $S_1 \cup S_2 \cup S_3$ of the rectangle. A similar "minimal principle", obtained by reflection, also holds: if for some function $Lu \leq 0$ throughout R, then the minimum of u is also taken on the parabolic boundary. Next, define the following functions:

$$m_-(x,t) = \frac{(1+\gamma)x}{1+\gamma x+\gamma(1-x)e^{-(1+\gamma)t}} \tag{C.30}$$

$$m_+(x,t) = \frac{x(1-\gamma e^{-(1-\gamma)t})}{1-\gamma+\gamma x(1-e^{-(1-\gamma)t})}. \tag{C.31}$$

The functions $m_-(x,t)$ and $m_+(x,t)$ can be shown, respectively, to be the means of the marginal distributions of the minimal and maximal processes, defined earlier in Appendix B. Now we can establish:

Lemma 4. *Let $u(x,t)$ be a continuous solution on $[0,1]\times[0,\infty)$ to $Lu = 0$, satisfying the boundary conditions $u(x,0) = x$, and $u(0,t) = 0$, $u(1,t) = 1$ for all time. Then*

$$m_-(x,t) \leq u(x,t) \leq m_+(x,t),$$
$$\forall (x,t) \in [0,1]\times[0,\infty). \tag{C.32}$$

The foregoing can be proved by verifying that $Lm_- \geq 0$, and $Lm_+ \leq 0$ on R, and then applying the maximal and minimum principles to $L(u-m_\pm)$ and noting $u - m_\pm$ vanishes on the parabolic boundary. This suffices already to prove a very strong result: the mean $m(x,t)$ of any GWF process is bounded between $m_-(x,t)$ and $m_+(x,t)$. By taking $t \to \infty$ in Lemma 4, we obtain the probability of fixation, and the content of Theorem 6.

C.6. *Theorems 7 and 8*

We do not give proofs of these mathematical representation theorems here. Their precise statement and complete proofs can be found in the thesis Der (2010), and will receive exposition in future publications.

Appendix D. Fixation probability in the Λ_1-Cannings model

By Theorem 7 and Table 1, the generator for the Λ_1-model with selection is

$$Gu = \gamma x(1-x)u'(x) + xu(1) + (1-x)u(0) - u(x) \tag{D.1}$$

where $\gamma = Ns$. The fixation probability $\pi_1(x)$ is then an equilibrium solution of the generalized backward Kolmogorov equation; that is, $G\pi_1 = 0$. Solving this first-order differential equation subject to the boundary conditions $\pi_1(0) = 0$, $\pi_1(1) = 1$ leads to the expression (30).

Appendix E. Assured fixation in power-law models

In this section we establish the property of almost-sure fixation in some of the power-law models. This property can be extended to many of the Λ-processes, which will receive exposition in a later publication. For simplicity we consider the index range $0 < \alpha < 1$. Using Table 1 and the expression in Theorem 7, and integrating by parts it can be shown that the generator for this process under selective coefficient $\gamma = Ns/\sigma^2$ is

$$\begin{aligned} G_{\gamma,\alpha}u(x) = {} & \gamma x(1-x)u'(x) + \frac{3-\alpha}{1-\alpha}\Bigg(-u(x) \\ & + (1-\alpha)x^{\alpha-1}(1-x)\int_0^x (x-z)^{-\alpha}u(z)\,dz \\ & + (1-\alpha)x(1-x)^{\alpha-1}\int_x^1 (z-x)^{-\alpha}u(z)\,dz\Bigg). \end{aligned} \tag{E.1}$$

Now we study solutions to

$$G_{\gamma,\alpha}u = 0, \qquad u(0) = 0, \quad u(1) = 1 \tag{E.2}$$

or equivalently, $Lu = 0$, where $L = (k/\gamma)G$, with $k = \gamma\frac{1-\alpha}{3-\alpha}$.

It is important to first analyze the auxiliary operator defined by

$$L_0u = kxu' - u + (1-\alpha)x^{\alpha-1}\int_0^x \frac{u(z)}{(x-z)^\alpha}\,dz. \tag{E.3}$$

By a change of variables, $u(x) = x^p$ for $p > -1$ is an eigenvector of L_0 associated to the eigenvalue $\lambda_p = kp - 1 + (1-\alpha)B(p+1, 1-\alpha)$, where $B(\cdot,\cdot)$ the Beta-function. Since we expect the operator L_0 to mimic that of L near $x = 0$, we examine the critical points in the spectrum. Define

$$S(p) = \gamma\frac{1-\alpha}{3-\alpha}p - 1 + (1-\alpha)B(p+1, 1-\alpha) \tag{E.4}$$

for values $p > -1, 0 < \alpha < 1, \gamma > 0$. The function S has two roots, one at 0, and another at a generally non-trivial root p_r, depending on γ. It can further be shown that there is exactly one value of γ for which the double-root $p_r = 0$ occurs, and it happens precisely when $S'(0) = 0$. Solving this produces the critical value for γ:

$$\gamma_0 = -(3-\alpha)\int_0^1 (1-t)^{-\alpha}\log t\,dt. \tag{E.5}$$

Next, note that the operator L can be written as

$$\begin{aligned} Lu = {} & L_0u - x(L_0u + u) + (1-\alpha)x(1-x)^{\alpha-1} \\ & \times \int_x^1 (z-x)^{-\alpha}u(z)\,dz \end{aligned} \tag{E.6}$$

Using the previous observations, one can prove the following inequality for functions associated to the root p_r:

Lemma 5. *Let $p = p_r$ be the non-trivial root of $kp - 1 + (1-\alpha)B(p+1, 1-\alpha) = 0$, where $k = \gamma\frac{1-\alpha}{3-\alpha}$. Setting $u(x) = x^{p_r}$, one has $Lu \geq 0$.*

Finally, if $\pi_{\alpha,\gamma}(x)$ is the probability of fixation for the power-law process of index α and selection pressure γ, initialized at allele frequency x, we can apply the maximum principle in coordination with the previous lemma to show that it must satisfy

Theorem 9. *Let $p = p_r$ be the positive root of the equation $\gamma\frac{1-\alpha}{3-\alpha}p - 1 + (1-\alpha)B(p+1, 1-\alpha) = 0$, with selective pressure $0 < \gamma < \gamma_0$, where γ_0 is given by (E.5). Then*

$$x^{p_r} \leq \pi_{\alpha,\gamma}(x), \quad 0 < x < 1. \tag{E.7}$$

Theorem 9 reveals an upper bound on the selective pressure which ensures assured fixation, for every positive initializing frequency. Since $p_r \downarrow 0$ as $\gamma \uparrow \gamma_0$, the critical value of assured fixation γ^* satisfies $\gamma^* \leq \gamma_0$, establishing half of the result (31). The other inequality $\gamma^* \geq \gamma_0$ is established in a similar way.

References

Beckenbach, A., 1994. Mitochondrial haplotype frequ encies in oysters: neutral alternatives to selection models. Genetics and Evolution of Aquatic Organisms 122–134.

Birkner, M., Blath, J., 2007. Inference for lambda coalescents. WIAS Preprint 1211.

Birkner, M., Blath, J., 2009. Measure-valued diffusions, general coalescents and population genetic inference. Trends in Stochastic Analysis 351, 329–363.

Buri, P., 1956. Gene frequency in small populations of mutant drosophila. Evolution 10, 367–402.

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

Cannings, C., 1974. The latent roots of certain markov chains arising in genetics: a new approach, i. haploid models. Advances in Applied Probability 6, 260–290.

Der, R., 2010. A theory of generalised population processes, Ph.D. Thesis ProQuest, Philadelphia.

Desai, M.M., Plotkin, J.B., 2008. The polymorphism frequency spectrum of finitely many sites under selection. Genetics 180, 2175–2191.

Donnelly, P., Kurtz, T., 1999a. Genealogical processes for fleming-viot models with selection and recombination. Annals of Applied Probability 9, 1091–1148.
Donnelly, P., Kurtz, T., 1999b. Particle representations for measure-valued population models. Annals of Probability 27, 166–205.
Dynkin, E., 1965. Markov Processes. Springer-Verlag, Berlin.
Eldon, B., Wakeley, J., 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172, 2621–2633.
Eldon, B., Wakeley, J., 2008. Linkage disequilibrium under skewed offspring distribution among individuals in a population. Genetics 178, 1517–1532.
Eldon, B., Wakeley, J., 2009. Coalescence times and FST under a skewed offspring distribution among individuals in a population. Genetics 181, 615–629.
Etheridge, A., Griffiths, C., 2009. A coalescent dual process in a moran model with genic selection. Theoretical Population Biology 75, 320–330.
Etheridge, A., Griffiths, C., Taylor, J., 2010. A coalescent dual process in a moran model with genic selection, and the lambda coalescent limit. Theoretical Population Biology 78, 77–92.
Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theoretical Population Biology 3, 87–112.
Ewens, W.J., 2004. Mathematical Population Genetics, second ed., Springer, New York.
Fay, J.C., Wu, C.I., 2000. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.
Fisher, R.A., 1958. The Genetical Theory of Natural Selection. Dover Publications, New York.
Fu, Y., 2006. Exact coalescent for the wright–fisher model. Theoretical Population Biology 69, 385–394.
Fu, Y.X., Li, W.H., 1992. Statistical tests of neutrality of mutations. Genetics 133, 693–709.
Haldane, J., 1932. A mathematical theory of natural and artificial selection, part ix, rapid selection. Proceedings of the Cambridge Philosophical Society 28, 244–248.
Hartl, D.L., Clark, A.G., 2006. Principles of Population Genetics, fourth ed., Sinauer Associates, Inc.
Hedgecock, D., 1994. Does variance in reproductive success limit effective population size of marine organisms? Non-neutral Evolution 188–198.
Hudson, R.R., Kreitman, M., Aguad, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159.
Karlin, S., McGregor, J., 1964. Direct product branching processes and related markov chains. Proceedings of the National Academy of Science of the United States of America 51, 598–602.
Karlin, S., Taylor, H., 1981. A Second Course in Stochastic Processes. Academic Press, San Diego.
Kimura, M., 1955. Solution of a process of random genetic drift with a continuous model. Proceedings of the National Academy of Science of the United States of America 41, 144–150.
Kimura, M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47, 713–719.
Kimura, M., 1994. Population Genetics, Molecular Evolution, and the Neutral Theory: Selected Papers. The University of Chicago Press, Chicago.
Kingman, J.F.C., 1982. The coalescent. Stochastic Process and their Applications 13, 235–248.
Krone, S.M., Neuhauser, C., 1997. Ancestral processes with selection. Theoretical Population Biology 51, 210–237.
Lessard, S., 2007. An exact sampling formula for the WrightFisher model and a solution to a conjecture about the finite-island model. Genetics 177, 1249–1254.
Lessard, S., Ladret, V., 2007. The probability of fixation of a single mutant in an exchangeable selection model. Journal of Mathematical Biology 54, 721–744.
Li, W.H., 1993. So, what about the molecular clock hypothesis? Current Opinion in Genetics and Development 3, 896–901.
Lieberman, E., Hauert, C., Nowak, M., 2005. Evolutionary dynamics on graphs. Nature 433, 312–316.
McDonald, J., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in drosophila. Nature 351, 652–654.
Möhle, M., 2001. Forward and backward diffusion approximations for haploid exchangeable population models. Stochastic Process and their Applications 95, 133–149.
Möhle, M., Sagitov, S., 2001. A classification of coalescent processes for haploid exchangeable population models. Annals of Probability 29, 1547–1562.
Moran, P.A.P., 1958. Random processes in genetics. Proceedings of the Cambridge Philosophical Society 54, 60–71.
Neuhauser, C., Krone, S.M., 1997. The genealogy of samples in models with selection. Genetics 145, 519–534.
Norman, F., 1975. Diffusion approximation of non-Markovian processes. Annals of Probability 3, 358–364.
Pierce, B., 2007. Genetics: A Conceptual Approach, third ed., W.H. Freeman.
Pitman, J., 1999. Coalescents with multiple collisions. Annals of Probability 27, 1870–1902.
Przeworski, M., Charlesworth, B., Wall, J., 1999. Genealogies and weak purifying selection. Molecular Biology and Evolution 16, 246–252.
Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. Journal of Applied Probability 36, 1116–1125.
Sawyer, S.A., Hartl, D.L., 1992. Population genetics of polymorphism and divergence. Genetics 132, 1161–1176.
Schweinsberg, J., 2000. A necessary and sufficient condition for the λ-coalescent to come down from infinity. Electronic. Communications in Probability 5, 1–11.
Schweinsberg, J., 2003. Coalescent processes obtained from supercritical Galton-Watson processes. Stochastic Processes and their Applications 106, 107–139.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.
Tavare, S., 2004. Ancestral inference in population genetics. Lectures on Probability Theory and Statistics 1837, 1–188.
Wakeley, J., 2009. Coalescent Theory: An Introduction. Roberts and Co., New York, NY.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.
Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 15, 496–503.