Contents lists available at ScienceDirect

Theoretical Population Biology

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

The conditional ancestral selection graph with strong balancing selection

John Wakeley *, Ori Sargsyan

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Received 28 January 2009
Available online 14 April 2009

Keywords:
Ancestral selection graph
Structured coalescent
Balancing selection

ABSTRACT

Using a heuristic separation-of-time-scales argument, we describe the behavior of the conditional ancestral selection graph with very strong balancing selection between a pair of alleles. In the limit as the strength of selection tends to infinity, we find that the ancestral process converges to a neutral structured coalescent, with two subpopulations representing the two alleles and mutation playing the role of migration. This agrees with a previous result of Kaplan et al., obtained using a different approach. We present the results of computer simulations to support our heuristic mathematical results. We also present a more rigorous demonstration that the neutral conditional ancestral process converges to the Kingman coalescent in the limit as the mutation rate tends to infinity.

1. Introduction

Balancing selection is a phenomenon that fitness differences among individuals in a population tend to preserve genetic variation. It is a special case of frequency dependent selection in which, roughly speaking, rare alleles are favored over common ones. Theoretical studies of balancing selection date back to the beginning of population genetics (Fisher, 1930; Wright, 1931; Haldane, 1932; Wright, 1939) and more recently have focused on explaining the unusual patterns of variation observed at some genetic loci (Hudson and Kaplan, 1988; Takahata, 1990; Vekemans and Slatkin, 1994). These patterns include high levels of polymorphism and allelic variation shared between species; for an example from plants see Charlesworth et al. (2006).

It seems clear from recent genome-wide studies in humans (Asthana et al., 2005; Bubb et al., 2006) that long-term balancing selection is not as ubiquitous as purifying selection or even positive selection. Strong evidence has been found at only a handful of loci in humans, including the well known cases of the major histocompatibility loci (HLA) and the locus determining ABO blood type (Bubb et al., 2006). Loci have also been identified in other species. In the plant family Brassicaceae, the genes involved in self-incompatibility systems are under very strong balancing selection (Richman et al., 1996; Kamau et al., 2007). In the case of the alcohol dehydrogenase locus in *Drosophila melanogaster*, Hudson and Kaplan (1988) were able to explain variation near a codon under balancing selection by assuming that the frequencies of the two amino acids at that site were held constant over long periods of time.

Although balancing selection may be rare compared to other forms of selection, good examples do exist and these apparently exhibit strong selection. Thus, it is of interest to understand the properties of models of balancing selection, in particular when selection is strong. In this paper, we present a heuristic analysis of strong balancing selection in the simple case of two allelic types. In particular, we consider a model of symmetric heterozygote advantage in the context of the conditional ancestral selection graph (Krone and Neuhauser, 1997; Neuhauser and Krone, 1997; Stephens and Donnelly, 2003) and make a connection between this model and the model of Hudson and Kaplan (1988) in which the strength of selection is assumed to be infinite. We support our mathematical results using computer simulations.

2. Methods and results

We will focus on the special case of symmetric heterozygote advantage between two alleles, but we begin with the general diploid selection model described in Stephens and Donnelly (2003). Our notation differs slightly from theirs.

There are K possible alleles $(A_1, A_2, \ldots, A_K)$ at a single locus without recombination. Forward in time, the scaled rate of mutation from any allele to allele A_i is $\theta\alpha_i/2$, where $\sum_{i=1}^{K}\alpha_i = 1$. This 'parent-independent' mutation model can be used to describe any two-allele mutation model, but only some models with $K \geq 3$ alleles. There are $K(K-1)/2$ scaled selection parameters, one for each unique diploid combination of alleles, or genotype, and these are represented by $\sigma(A_i, A_j)$. Thus, $\sigma(A_i, A_j) = \sigma(A_j, A_i)$. Without loss of generality (Donnelly and Kurtz, 1999), we may assume that

$$0 \leq \sigma(A_i, A_j) \leq \sigma_{\max} \quad \text{for all } i \text{ and } j.$$

These parameters – θ, α_i, and $\sigma(A_i, A_j)$, for $i, j = 1, \ldots, K$ – are those of a continuous-time, continuous-allele-frequency

[*] Corresponding address: Harvard University, 4100 Biological Laboratories, 16 Divinity Ave., Cambridge, MA 02138, USA.
E-mail address: wakeley@fas.harvard.edu (J. Wakeley).

0040-5809/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2009.04.002

diffusion limit which holds for a broad class of discrete-time, finite-population-size models, including the Wright–Fisher model (Fisher, 1930; Wright, 1931) and the Moran model (Moran, 1958, 1962), the limit being as the population size tends to infinity with time rescaled appropriately; see Ewens (2004).

This is the most commonly used diffusion approximation in population genetics and is based on the assumption that the per-generation probability of mutation and the absolute fitness differences among individuals are on the order of the inverse of the population size, so that the rescaled parameters θ and $\sigma(A_i, A_j)$ are finite. Karlin and McGregor (1964) and Norman (1975) have described other diffusion approximations that are appropriate when the per-generation probability of mutation and the absolute fitness differences among individuals are much greater than the inverse of the population size, so that the rescaled parameters would tend to infinity as the population size tends to infinity. We will return to these other, "Gaussian" diffusion models below.

The frequency of allele A_i in the population is denoted x_i, and the state space of the forward-time diffusion process is the $K-1$-dimensional unit simplex

$$\Delta_K = \{(x_1, x_2, \dots, x_K) : x_i \geq 0, i = 1, 2, \dots, K,\\ x_1 + x_2 + \cdots + x_K = 1\}.$$

The stationary distribution of this diffusion process, with general diploid selection and parent-independent mutation, is known up to a normalizing constant (Wright, 1949, 1969), and is given by

$$\phi_{\sigma,\theta}(x_1, \dots, x_K) = C x_1^{\theta\alpha_1 - 1} \cdots x_K^{\theta\alpha_K - 1} e^{\sigma^*(x_1,\dots,x_K)/2}, \tag{1}$$

in which

$$\sigma^*(x_1, \dots, x_K) = \sum_{i=1}^{K} \sum_{j=1}^{K} \sigma(A_i, A_j) x_i x_j$$

is the scaled mean fitness of the population. In considering the ancestry of a sample from the population, we are interested in the sampling distribution

$$p_{\sigma,\theta}(n_1, \dots, n_K) = \int_{\Delta_K} x_1^{n_1} \cdots x_K^{n_K} \phi_{\sigma,\theta}(x_1, \dots, x_K) dx_1 \cdots dx_K, \tag{2}$$

which is the probability that an ordered sample of size n contains n_i copies of allele A_i, for $i = 1, \dots, K$. There are

$$\frac{n!}{n_1! n_2! \cdots n_K!}$$

such ordered samples and each one has the same probability, given by (2). The subscripts θ and σ denote the dependence of the sampling probability on these parameters, while the dependence on $\alpha_1, \dots, \alpha_K$ is implicit. Below, we consider the limits $\theta \to \infty$ (with $\sigma = 0$) and $\sigma \to \infty$ (with θ constant). The parameters $\alpha_1, \dots, \alpha_K$ are treated as constants throughout.

Using the above model, Stephens and Donnelly (2003) described a general version of the ancestral selection graph (ASG) of Krone and Neuhauser (1997). The ASG models the joint sampling distribution of allelic types and gene genealogies when selective differences exists among alleles. A gene genealogy is the genetic ancestry of a sample back to its most recent common ancestor. Under neutrality (i.e., without selection), ancestral processes describing gene genealogies are relatively simple because all genetic lineages are exchangeable (Kingman, 1982a,b,c). In the simplest model, each pair of lineages coalesces (reaches its common ancestor) independently with rate equal to 1, and the gene genealogy is a random-joining tree with associated coalescence times. Neutral models have been extended to include a range of biologically relevant complications. Notably for our purposes, the *structured coalescent* (Takahata, 1988; Notohara, 1990; Herbots, 1997) describes the movement of lineages between, and their coalescence within, subpopulations of constant size.

The ASG is one solution to the problem of non-exchangeability: that the rates of coalescence between genetic lineages depend on their allelic states when selection operates. Krone and Neuhauser solved this problem by constructing a two-layer model in which allelic types are initially unspecified and all lineages reproduce with the maximum fitness. Later, after the allelic states are specified, some reproduction events in which the parents have less than the maximum fitness are removed. Correspondingly, the ancestry of a sample whose allelic states are unknown initially includes some number of *virtual* lineages which proliferate in a large ancestral graph. Virtual lineages arise via branching events in which lineages split as they are followed back in time. Branching events correspond to the reproduction events in the population that may or may not be realized, depending on the allelic states of the parents.

In the ASG, the process of branching and coalescing is followed back to the first time there is only one lineage, the 'ultimate' ancestor of all the lineages. The ultimate ancestor is assigned an allelic type from the equilibrium distribution, allowing virtual lineages to be identified and removed from the graph. This leaves the gene genealogy of the *real* lineages together with the allelic types of the sample. For a basic introduction to the ancestral selection graph, see Section 7.1 in Wakeley (2008b). For a very general, mathematically rigorous treatment, see Donnelly and Kurtz (1999).

A second solution to the coalescent with selection is to explicitly model allele-frequency trajectories and gene genealogies backward in time (Barton et al., 2004; Barton and Etheridge, 2004). This was the method used by Kaplan et al. (1988), Hudson and Kaplan (1988) and Kaplan et al. (1989), who additionally assumed that the strength of selection was essentially infinite. When the strength of selection is very strong, allele frequencies over time will closely follow their predicted deterministic trajectories, with small Gaussian deviations whose magnitude becomes smaller if the population size and the rescaled parameters become larger (Karlin and McGregor, 1964; Norman, 1975). When balancing selection is exceedingly strong, the allele frequencies can be considered to be fixed, and the ancestral process has the same form as the structured coalescent mentioned above, with subpopulations represented by allelic states (Hudson and Kaplan, 1988).

The *conditional* ASG is an extension by Slade (2000a,b) to the case in which the allelic states of the sample are known. When this is true, the allelic states of all lineages, both virtual and real, are known during the entire ancestry of the sample. Then it is only necessary to follow the ancestry back to the most recent common ancestor of the real lineages (Slade, 2000a) rather than back to the ultimate ancestor of all the lineages. In addition, some number of virtual lineages may be ignored (Slade, 2000a; Fearnhead, 2002; Baake and Bialowons, 2008), greatly reducing the number of virtual branches that appear during the ancestry of the sample. This makes both analysis and simulation more practical. By following the minimum possible number of virtual lineages, the limiting $\sigma \to \infty$ ancestral process for directional, or genic, selection and mutation between two alleles was described in Wakeley (2008a), where it was also shown that simulations appear feasible for any value of σ.

Here we study the conditional ancestral selection graph in the case of strong balancing selection, in particular, symmetric heterozygote advantage between two alleles. Under genic selection, the simplifications of Slade (2000a) and Fearnhead (2002) lead to the annihilation of all virtual lineages in the limit $\sigma \to \infty$ (Wakeley, 2008a). Under balancing selection, however, virtual lineages still proliferate in the graph. Therefore, we approach the problem without using the simplifications of Slade (2000a) and Fearnhead (2002), and instead allow virtual lineages to grow in number, potentially without bound.

Using a heuristic analysis, we characterize the limiting ($\sigma \to \infty$) process back to the first coalescent event or mutation event among the real lineages in the sample. We treat the proliferation of virtual lineages using a "separation-of-time-scales" method based on that of Möhle (1998). The limiting process turns out to be identical to the structured coalescent (Takahata, 1988; Notohara, 1990; Herbots, 1997) process for strong balancing selection between a pair of alleles with constant allele frequencies, described previously by Kaplan et al. (1988) using a different approach. We support our analytical results with computer simulations. As an illustration of the separation-of-time-scales approach, we also present a more rigorous treatment of strong mutation under neutrality.

2.1. The conditional ASG for two alleles and symmetric balancing selection

Our starting point is the continuous-time conditional ancestral process given by equations 5 through 8 in Stephens and Donnelly (2003). We consider balancing selection in the form of symmetric heterozygote advantage, so that

$$\begin{aligned}&\sigma(A_i, A_i) = 0\\ &\sigma(A_i, A_j) = \sigma \quad i \neq j\\ &\sigma_{\max} = \sigma,\end{aligned}$$

and we restrict ourselves to the case of $K = 2$ alleles. The process is Markovian and the state space in Stephens and Donnelly (2003) is the set of all possible ordered sets of ancestral lineages with allelic types specified. We follow Slade (2000a,b) in decomposing the ancestral lines into real lineages (those we know are ancestral to the sample) and virtual lineages (those we know are not ancestral to the sample) and in using r_1, r_2, v_1, and v_2 to denote the numbers of real and virtual lineages of type A_1 and A_2.

A set of ancestral lines in state (r_1, r_2, v_1, v_2) makes transitions to

$$\begin{aligned}
&(r_1-1, r_2, v_1, v_2) \quad \text{with rate} \quad \binom{r_1}{2}\frac{p_{\sigma,\theta}(r_1-1, r_2, v_1, v_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2-1, v_1, v_2) \quad \text{with rate} \quad \binom{r_2}{2}\frac{p_{\sigma,\theta}(r_1, r_2-1, v_1, v_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1-1, r_2+1, v_1, v_2) \quad \text{with rate}\\
&\quad r_1\frac{\theta\alpha_1}{2}\frac{p_{\sigma,\theta}(r_1-1, r_2+1, v_1, v_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&\quad (r_1+1, r_2-1, v_1, v_2) \quad \text{with rate}\\
&\quad r_2\frac{\theta\alpha_2}{2}\frac{p_{\sigma,\theta}(r_1+1, r_2-1, v_1, v_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1, v_2) \quad \text{with rate} \quad r_1\frac{\theta\alpha_1}{2} + r_2\frac{\theta\alpha_2}{2}\\
&(r_1, r_2, v_1-1, v_2) \quad \text{with rate}\\
&\quad \left(r_1v_1 + \binom{v_1}{2}\right)\frac{p_{\sigma,\theta}(r_1, r_2, v_1-1, v_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1, v_2-1) \quad \text{with rate}\\
&\quad \left(r_2v_2 + \binom{v_2}{2}\right)\frac{p_{\sigma,\theta}(r_1, r_2, v_1, v_2-1)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1-1, v_2+1) \quad \text{with rate}\\
&\quad v_1\frac{\theta\alpha_1}{2}\frac{p_{\sigma,\theta}(r_1, r_2, v_1-1, v_2+1)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1+1, v_2-1) \quad \text{with rate}\\
&\quad v_2\frac{\theta\alpha_2}{2}\frac{p_{\sigma,\theta}(r_1, r_2, v_1+1, v_2-1)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1, v_2) \quad \text{with rate} \quad v_1\frac{\theta\alpha_1}{2} + v_2\frac{\theta\alpha_2}{2}\\
&(r_1, r_2, v_1+2, v_2) \quad \text{with rate}\\
&\quad (r_1+v_1+2r_2+2v_2)\frac{\sigma}{2}\frac{p_{\sigma,\theta}(r_1, r_2, v_1+2, v_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1+1, v_2+1) \quad \text{with rate}\\
&\quad (r_1+v_1+r_2+v_2)\frac{\sigma}{2}\frac{p_{\sigma,\theta}(r_1, r_2, v_1+1, v_2+1)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}\\
&(r_1, r_2, v_1, v_2+2) \quad \text{with rate}\\
&\quad (2r_1+2v_1+r_2+v_2)\frac{\sigma}{2}\frac{p_{\sigma,\theta}(r_1, r_2, v_1, v_2+2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}.
\end{aligned} \tag{3}$$

The transitions in lines 1, 2, 6, and 7 above are coalescent events, while transitions 11, 12, and 13 are branching events. The remaining six transitions are mutation events, including the "empty" mutation events (Baake and Bialowons, 2008) which do not change the allelic type (lines 5 and 10). These follow from the assumption of parent-independent mutation (Stephens and Donnelly, 2003), which is not really necessary here, but is needed for tractability when $K > 2$. We could filter these empty events out, and thereby reduce the total rate of events to those that actually affect the state of the lineages.

The total rate of events – the sum of the rates in (3) – is equal to

$$\binom{r_1+v_1+r_2+v_2}{2} + (r_1+v_1+r_2+v_2)\frac{\theta}{2} + (r_1+v_1+r_2+v_2)\frac{\sigma}{2}. \tag{4}$$

This can be verified by substituting the probability of an ordered sample of $r_1 + v_1$ A_1 alleles and $r_2 + v_2$ A_2 alleles into (3). Under symmetric heterozygote advantage between $K = 2$ alleles, the distribution of the frequency, x, of allele A_1 is given by a special case of (1), namely

$$\phi_{\sigma,\theta}(x) = Cx^{\theta\alpha_1-1}(1-x)^{\theta\alpha_2-1}e^{-\sigma(x^2+(1-x)^2)/2},$$

where C is defined so that $\int_0^1 \phi_{\sigma,\theta}(x)dx = 1$. Then, the sampling probability (2) becomes

$$p_{\sigma,\theta}(r_1, r_2, v_1, v_2) = C\int_0^1 x^{\theta\alpha_1+r_1+v_1-1}(1-x)^{\theta\alpha_2+r_2+v_2-1}e^{-\sigma(x^2+(1-x)^2)/2}dx, \tag{5}$$

with C the same as in $\phi_{\sigma,\theta}(x)$.

The rates in (3) are the rates of events in the ancestral graph, conditional on the states of the lineages at any given time. These conditional rates are derived using Bayes' rule (Stephens and Donnelly, 2003) and thus have the form (unconditional rate of Event) × $P\{\text{Data}|\text{Event}\}/P\{\text{Data}\}$, in which "Data" refers to an ordered sample. Note that, in contrast to the case of genic selection where each branching event produces one virtual lineage, under heterozygote advantage or general diploid selection, each branching event produces two virtual lineages. When the fitness depends on the diploid genotype, each reproduction event in the population involves three genetic lineages: the single allele removed from the population by a death event, and two others. We must keep all three as we follow the ancestry through a branching event.

While Stephens and Donnelly (2003) distinguish events depending on which particular lineages are involved, we have followed the fairly common practice of grouping events based on how they change the numbers of real and virtual lineages of each allelic type. We can use (3) to study times to events, but in order to specify the entire structure of the ancestry of a sample we would need the additional rule that every lineage is equally likely to be involved in every event. For example, if an event of the first type above were to occur, we would then need to choose a random pair of A_1 lineages to be the pair that coalesces.

2.2. Separation of time scales: Strong neutral mutation

Here we consider the limit $\theta \to \infty$ for $\sigma = 0$ in order to illustrate the separation-of-time-scales approach of Möhle (1998) that we will later apply heuristically to the case $\sigma \to \infty$. Since $\sigma = 0$, $\upsilon_1 = 0$, and $\upsilon_2 = 0$, we omit them in this section. For this well studied neutral case, the constant in $\phi_{\sigma,\theta}(x)$ can be evaluated, and we have

$$\phi_\theta(x) = \frac{\Gamma(\theta)}{\Gamma(\theta\alpha_1)\Gamma(\theta\alpha_2)} x^{\theta\alpha_1 - 1}(1-x)^{\theta\alpha_2 - 1}$$

and

$$p_\theta(r_1, r_2) = \frac{\Gamma(\theta)\Gamma(\theta\alpha_1 + r_1)\Gamma(\theta\alpha_2 + r_2)}{\Gamma(\theta\alpha_1)\Gamma(\theta\alpha_2)\Gamma(\theta + r_1 + r_2)}.$$

To gain some intuition about what follows, consider the behavior of $\phi_\theta(x)$ and $p_\theta(r_1, r_2)$ when θ is large. The limit of the sampling probability is straightforward to obtain, and is

$$\lim_{\theta\to\infty} p_\theta(r_1, r_2) = \alpha_1^{r_1}\alpha_2^{r_2}. \qquad (6)$$

Thus, as θ grows, each sample independently has probability α_1 of being type A_1 and probability $\alpha_2 = 1 - \alpha_1$ of being type A_2. We infer that the dependence of the allelic states of the samples on the underlying gene genealogy, which is captured in $p_\theta(r_1, r_2)$, disappears in the limit.

Correspondingly, the distribution $\phi_\theta(x)$ of the random variable X, which is the equilibrium frequency of allele A_1 under neutrality, becomes concentrated at $X = \alpha_1$ as θ grows. The shape of this distribution when θ is large may be obtained by the direct study of $\phi_\theta(x)$ above or by appealing to the general work of Karlin and McGregor (1964) and Norman (1975). These authors developed diffusion approximations for large populations in which the scaled strengths of evolutionary forces (here θ for mutation or σ for selection) are also large. In the case of strong mutation, X should exhibit Gaussian deviations around its deterministic equilibrium point, $X = \alpha_1$, with a smaller and smaller variance as θ grows. From $\phi_\theta(x)$, we obtain

$$E[X] = \alpha_1$$

and

$$\mathrm{Var}[X] = \alpha_1\alpha_2/(\theta + 1).$$

These are originally due to Wright (1931) – see page 123 – who also noted the approach of $\phi_\theta(x)$ to a normal density for large population sizes. Fig. 1 shows how the shape of $\phi_\theta(x)$ changes as θ increases for $\alpha_1 = 2/3$, and also illustrates the excellent agreement when $\theta = 100$ of an approximating normal distribution with the same mean and variance, given by the equations above. As θ increases to infinity, all of the probability mass does become concentrated at $X = \alpha_1$, and hence the sampling probability converges to (6).

The ancestral process under neutrality ($\sigma = 0$, $\upsilon_1 = 0$, and $\upsilon_2 = 0$) is greatly simplified compared to (3). Using the expression for $p_\theta(r_1, r_2)$ above, a set of ancestral lines in state (r_1, r_2) moves to state

$$\begin{aligned}
&(r_1 - 1, r_2) \quad \text{with rate} \quad \binom{r_1}{2}\frac{\theta + r - 1}{\theta\alpha_1 + r_1 - 1},\\
&(r_1, r_2 - 1) \quad \text{with rate} \quad \binom{r_2}{2}\frac{\theta + r - 1}{\theta\alpha_2 + r_2 - 1},\\
&(r_1 - 1, r_2 + 1) \quad \text{with rate} \quad r_1\frac{\theta\alpha_1}{2}\frac{\theta\alpha_2 + r_2}{\theta\alpha_1 + r_1 - 1},\\
&(r_1 + 1, r_2 - 1) \quad \text{with rate} \quad r_2\frac{\theta\alpha_2}{2}\frac{\theta\alpha_1 + r_1}{\theta\alpha_2 + r_2 - 1},\\
&(r_1, r_2) \quad \text{with rate} \quad r_1\frac{\theta\alpha_1}{2} + r_2\frac{\theta\alpha_2}{2},
\end{aligned} \qquad (7)$$

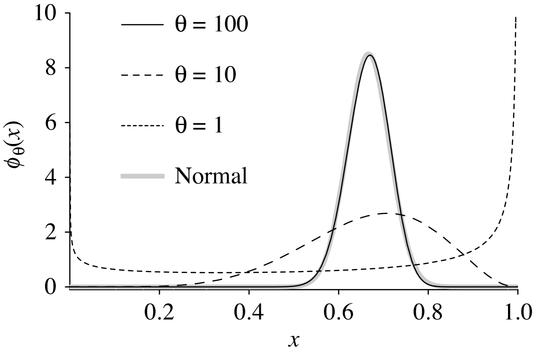

Fig. 1. Plots of the equilibrium distribution of x, the frequency of allele A_1, for three different strengths of mutation under neutrality. A normal distribution with mean $E[X] = 2/3$ and variance $\mathrm{Var}[X] = 2/909$, obtained using the equations in the text, is shown for comparison.

in which we use $r = r_1 + r_2$. Here, we show that the time to a coalescent event does not depend on the allelic states of the sample when $\theta \to \infty$, and is exponential with rate $r(r-1)/2$, as in Kingman's (unconditional) coalescent.

In studying the limit $\theta \to \infty$, let $\mathbf{Q}_\theta$ be the transition rate matrix of the ancestral process back to the first coalescent event. As above, we do not need to distinguish which particular lineages are involved in each event. We consider a process with a total of $r + 3$ states, so that $\mathbf{Q}_\theta$ is an $(r+3) \times (r+3)$ matrix. The first $r + 1$ states contain ordered samples that all have the same total number of lineages, $r = r_1 + r_2$, but which differ in the values of r_1 and r_2. Here we will index the states by 1 plus the number of A_1 lineages, so that state 1 represents $(r_1, r_2) = (0, r)$ and state $r + 1$ represents $(r_1, r_2) = (r, 0)$. States $r + 2$ and $r + 3$ are absorbing states and contain the corresponding ordered samples with one fewer A_1 lineage and one fewer A_2 lineage, respectively. Transitions among states 1 through $r + 1$ are mutation events and transitions to states $r + 2$ and $r + 3$ are coalescent events, between lineages of type A_1 and A_2, respectively.

With the ancestral process defined so, and using the rates (7), we have

$$\mathbf{Q}_\theta = \theta\mathbf{A} + \mathbf{B} + O(1/\theta),$$

where

$$\mathbf{A} = \lim_{\theta\to\infty} \mathbf{Q}_\theta/\theta$$

and

$$\mathbf{B} = \lim_{\theta\to\infty} (\mathbf{Q}_\theta - \theta\mathbf{A})$$

exist and have entries of order 1. The entries of $\mathbf{A}$ are

$$\begin{aligned}
&r_1\alpha_2/2 \quad \text{for transitions } (r_1, r_2) \longrightarrow (r_1 - 1, r_2 + 1),\\
&-(r_1\alpha_2 + r_2\alpha_1)/2 \quad \text{for transitions } (r_1, r_2) \longrightarrow (r_1, r_2),\\
&r_2\alpha_1/2 \quad \text{for transitions } (r_1, r_2) \longrightarrow (r_1 + 1, r_2 - 1),
\end{aligned}$$

with every other entry equal to zero. The entries of $\mathbf{B}$ are

$$\begin{aligned}
&r_1(\alpha_2 - r_1\alpha_2 + r_2\alpha_1)/2\alpha_1 \quad \text{for transitions}\\
&\quad (r_1, r_2) \longrightarrow (r_1 - 1, r_2 + 1),\\
&-r(r-1)/2 \quad \text{for transitions } (r_1, r_2) \longrightarrow (r_1, r_2),\\
&r_2(\alpha_1 + r_1\alpha_2 - r_2\alpha_1)/2\alpha_2 \quad \text{for transitions}\\
&\quad (r_1, r_2) \longrightarrow (r_1 + 1, r_2 - 1),\\
&r_1(r_1 - 1)/2\alpha_1 \quad \text{for transitions } (r_1, r_2) \longrightarrow (r_1 - 1, r_2),\\
&r_2(r_2 - 1)/2\alpha_2 \quad \text{for transitions } (r_1, r_2) \longrightarrow (r_1, r_2 - 1),
\end{aligned}$$

again with every other entry equal to zero. Because states $r + 2$ and $r + 3$ are absorbing states, all entries in rows $r + 2$ and $r + 3$ of both **A** and **B** are equal to zero (as are the entries in columns $r+2$ and $r+3$ of **A**).

To explain, consider the mutation event $(r_1, r_2) \longrightarrow (r_1 - 1, r_2 + 1)$ in the first lines of these equations for **A** and **B**. The rates of this event in **A** and **B** are the coefficients of θ and 1 in a series expansion of the corresponding, third line of (7) for large θ. In contrast, the largest term in the expansion of (7) for the coalescent event $(r_1, r_2) \longrightarrow (r_1 - 1, r_2)$ is of order 1. This appears in column $r+2$ of **B**, and the corresponding entry in **A** is zero. Thus, to leading order in θ, the matrix $\theta\mathbf{A}$ contains the rates of transitions involving only mutation events, while the rates of all coalescent events are confined to **B**. The matrix **B** also contains the $O(1)$ parts of the rates of mutation events. When θ is large, the time scale for mutation is θ times faster than the time scale for coalescence.

We are interested in the t-step transition probability matrix $\exp(t\mathbf{Q}_\theta)$ in the limit $\theta \to \infty$. Since $\mathbf{Q}_\theta = \theta\mathbf{A} + \mathbf{B} + O(1/\theta)$, we have

$$\exp\left\{t\mathbf{Q}_\theta\right\} = \exp\left\{t\theta\mathbf{Q}_\theta/\theta\right\} = \exp\left\{t\theta\left(\mathbf{A} + \frac{\mathbf{B}}{\theta} + O\left(\frac{1}{\theta^2}\right)\right)\right\}.$$

Following a similar application by Lessard and Wakeley (2004)–see Section 3 of that paper – we have

$$\begin{aligned}\exp\left\{\mathbf{A} + \frac{\mathbf{B}}{\theta} + O\left(\frac{1}{\theta^2}\right)\right\} &= \sum_{i\geq 0} \frac{\left(\mathbf{A} + \frac{\mathbf{B}}{\theta} + O\left(\frac{1}{\theta^2}\right)\right)^i}{i!} \\ &= \exp\left\{\mathbf{A}\right\} + \frac{\mathbf{C}}{\theta} + O\left(\frac{1}{\theta^2}\right)\end{aligned}$$

where

$$\mathbf{C} = \sum_{i\geq 1}\left\{\frac{\sum_{k=0}^{i-1}\mathbf{A}^k\mathbf{B}\mathbf{A}^{i-k-1}}{i!}\right\}.$$

Then, Lemma 1 of Möhle (1998) guarantees that the t-step transition probability matrix $\exp(t\mathbf{Q}_\theta)$ converges to

$$\lim_{\theta\to\infty}\exp\left\{t\mathbf{Q}_\theta\right\} = \mathbf{P}\exp\left\{t\mathbf{G}\right\}$$

in which

$$\mathbf{P} = \lim_{r\to\infty}\exp\left\{r\mathbf{A}\right\}$$

is the stationary distribution of the fast process (here mutation), and the infinitesimal generator is given by

$$\mathbf{G} = \mathbf{PCP} = \mathbf{PBP}.$$

The last equality holds because $\mathbf{PA} = \mathbf{AP} = \mathbf{0}$, which follows from the definition of **P** above. Namely, $\mathbf{P} = \lim_{r\to\infty}\mathbf{P}(r)$, where $\mathbf{P}(r)$ is the unique solution to both the forward equation $d\mathbf{P}(r)/dr = \mathbf{P}(r)\mathbf{A}$ and the backward equation $d\mathbf{P}(r)/dr = \mathbf{AP}(r)$, with $\mathbf{P}(0) = \mathbf{I}$; for example, see Theorem 2.1.1 in Norris (1997). At stationarity, $d\mathbf{P}/dt = \mathbf{PA} = \mathbf{AP} = \mathbf{0}$.

The matrix **A** describes a continuous-time Markov process with three non-communicating sets of states. Two of these are the two absorbing states, which are entered upon coalescence between a pair of A_1 alleles or between a pair of A_2 alleles. The entries in the upper left $(r+1)\times(r+1)$ block of **A** are the transition rates of an ergodic Markov process among the $r+1$ states in which there are r uncoalesced lineages. Therefore, this process of mutation between A_1 and A_2 among the r ancestral lineages has a unique stationary distribution, contained in the upper left $(r+1)\times(r+1)$ block of **P**. Thus, **P** has the form

$$\mathbf{P} = \begin{pmatrix} p_0 & \cdots & p_r & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ p_0 & \cdots & p_r & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

in which p_{r_1} is the probability that r_1 of the r lineages have type A_1 and the other $r - r_1 = r_2$ have type A_2. Since **A** is a tri-diagonal matrix, we can solve for the equilibrium by solving

$$p_{r_1}r_1\alpha_2/2 = p_{r_1-1}(r_2+1)\alpha_1/2$$

for $1 \leq r_1 \leq r$, subject to the constraint $\sum_{r_1=0}^{r} p_{r_1} = 1$. We have

$$p_{r_1} = \binom{r}{r_1}\alpha_1^{r_1}\alpha_2^{r_2},$$

which is what we expect given (6) and the fact that there are $\binom{r}{r_1}$ ordered samples that have r_1 lineages of type A_1 and the $r - r_1 = r_2$ of type A_2. Using this formula for p_{r_1}, we can also verify that $\mathbf{PA} = 0$ as required for the equilibrium solution of the backward equation mentioned above.

From analyses of small samples (not shown), we deduce that the first $r+1$ rows of the limiting t-step transition probability matrix, $\mathbf{P}\exp\{t\mathbf{G}\}$, are identical, with

$$\exp\{-tr(r-1)/2\}\binom{r}{r_1}\alpha_1^{r_1}\alpha_2^{r_2}$$

in column $r_1 + 1\,(0 \leq r_1 \leq r)$,

$$(1 - \exp\{-tr(r-1)/2\})\,\alpha_1$$

in column $r+2$, and

$$(1 - \exp\{-tr(r-1)/2\})\,\alpha_2$$

in column $r+3$. The last two rows of $\mathbf{P}\exp\{t\mathbf{G}\}$ have ones on the diagonal and zeros everywhere else. Thus, in the limiting $\theta \to \infty$ process, any sample of size r instantaneously assumes the stationary distribution of allelic states, p_{r_1}, which then decays steadily at rate $r(r-1)/2$ as a result of coalescence. A fraction α_1 of coalescent events are between alleles of type A_1 and a fraction α_2 are between alleles of type A_2.

We can also understand this by looking directly at the infinitesimal generator $\mathbf{G} = \mathbf{PBP}$. Because $(\mathbf{P})_{i,r+2} = (\mathbf{P})_{i,r+3} = 0$ for $1 \leq i \leq r+1$ and $(\mathbf{P})_{r+2,r+2} = (\mathbf{P})_{r+3,r+3} = 1$, the rates of coalescence are simple averages over the stationary distribution of allelic states, p_{r_1}. The rate of coalescence between A_1 alleles, from any uncoalesced state $1 \leq i \leq r+1$, is given by

$$\begin{aligned}(\mathbf{G})_{i,r+2} &= \sum_{j=1}^{r+3}(\mathbf{P})_{i,j}(\mathbf{B})_{j,r+2}(\mathbf{P})_{r+2,r+2} \\ &= \sum_{r_1=0}^{r} p_{r_1}(\mathbf{B})_{r_1+1,r+2} \\ &= \sum_{r_1=0}^{r}\binom{r}{r_1}\alpha_1^{r_1}\alpha_2^{r_2}r_1(r_1-1)/2\alpha_1 \\ &= \alpha_1 r(r-1)/2. \end{aligned}\tag{8}$$

Similarly, the rate of coalescence between A_2 alleles is equal to $\alpha_2 r(r-1)/2$. The total rate of coalescence is the sum of these and is equal to $r(r-1)/2$. Subsequent to a coalescent event, the same logic applies to the $r-1$ lineages that remain, which shows that the entire ancestral process converges to the standard neutral coalescent process, in which lineages are exchangeable (Kingman, 1982a,b,c) in the limit $\theta \to \infty$.

2.3. Convergence of conditional ancestral processes with strong selection

We now turn to the limit $\sigma \to \infty$ for constant θ, α_1, and $\alpha_2 = 1 - \alpha_1$. In contrast to the case of $\sigma = 0$, there are no simple expressions for the sampling probabilities that appear

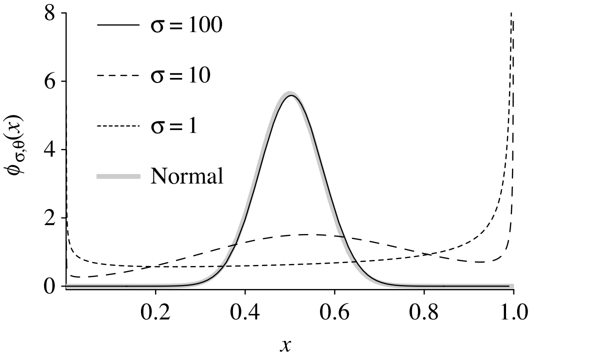

Fig. 2. Plots of the equilibrium distribution of x, the frequency of allele A_1, for three different strengths of selection, with $\theta = 1$ and $\alpha_1 = 2/3$. A normal distribution with mean $E[X] = 1/2$ and variance $\text{Var}[X] = 1/200$ is shown for comparison to the curve for $\sigma = 100$.

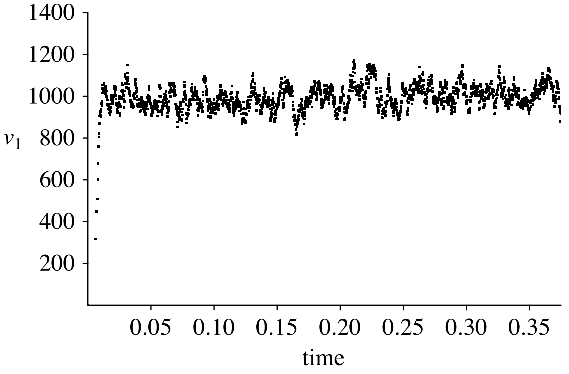

Fig. 3. The number of virtual lineages of type A_1 through time, starting from the sample $(2, 0, 0, 0)$ and with $\sigma = 1000$, $\theta = 1$, and $\alpha_1 = 2/3$. The process described by (3) was simulated, but using the approximations for the ratios of sampling probabilities given in the Appendix. Only every 500th value of v_1 is shown.

in ratios in (3). Instead, we have the integral representation of $p_{\sigma,\theta}(r_1, r_2, v_1, v_2)$ given in (5). As in Wakeley (2008a), we use the behavior of the ratios of sampling probabilities when σ is large to derive the limiting ($\sigma \to \infty$) ancestral process for (3).

Our approach here is heuristic, but the idea is the same as in the previous section. We separate events based on their time scales. The fast events are branching, coalescence, and mutation among virtual lineages. The slow events are coalescence and mutation among real lineages. Simulations, presented in Section 2.4, support our approach and results; for example, the conclusion that v_1 and v_2 approach stationarity quickly when σ is large (see Fig. 3). The direct application of Lemma 1 in Möhle (1998), which is for a finite Markov chain, is not justified because the state space of (v_1, v_2) is infinite. In addition, we do not know the stationary distribution of (v_1, v_2). Still, we invoke the kind of averaging we did in (8) to obtain the rates of the limiting process.

To develop some intuition about the sampling probability $p_{\sigma,\theta}(r_1, r_2, v_1, v_2)$ and hence about the limiting process, consider the equilibrium frequency $\phi_{\sigma,\theta}(x)$ as σ grows. Due to the factor $\exp(-\sigma(x^2+(1-x)^2)/2)$ in $\phi_{\sigma,\theta}(x)$, the density is centered around $x = 1/2$ when σ is large. Further, it is very well approximated by a normal distribution with mean equal to 1/2 and variance equal to $1/(2\sigma)$, consistent with the Gaussian diffusion results of Norman (1975). Fig. 2 plots $\phi_{\sigma,\theta}(x)$ for $\sigma = 1$, 10, and 100, with $\theta = 1$ and $\alpha_1 = 2/3$. The normal distribution fits quite well when $\sigma = 100$. As σ grows, the density becomes more and more concentrated on $x = 1/2$.

Then, recalling (6), we can guess that the sampling probability $p_{\sigma,\theta}(r_1, r_2, v_1, v_2)$ will be very close to $(1/2)^{r_1+r_2+v_1+v_2}$ when σ is large. From the way in which each event affects the lineages, we can see that the ratios of sampling probabilities in (3) should assume values close to 2 for coalescent events, 1 for mutation events, and 1/4 for branching events. Analysis of (5) bears this out, but is complicated by the fact that in general we must allow v_1 and v_2 to take on any values. In the Appendix, we present series expansions for the ratios of sampling probabilities for large σ, and for three different cases of v_1 and v_2. The three cases are (i) v_1 and v_2 finite, meaning $O(1)$ as $\sigma \to \infty$, (ii) v_1 and v_2 both $O(\sqrt{\sigma})$, and (iii) v_1 and v_2 both $O(\sigma)$ but with deviations of order $\sqrt{\sigma}$. In all three cases, we find

$$\frac{p_{\sigma,\theta}(r'_1, r'_2, v'_1, v'_2)}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} = \left(\frac{1}{2}\right)^{r'_1+r'_2+v'_1+v'_2-r_1-r_2-v_1-v_2} (1 + o(1)), \quad (9)$$

in agreement with the intuitive argument just given based on the behavior of $\phi_{\sigma,\theta}(x)$ as σ grows. In case (i), the $o(1)$ term in (9) is $O(1/\sigma)$ and in cases (ii) and (iii) it is $O(1/\sqrt{\sigma})$. However, we emphasize that (9) holds for given v_1 and v_2 in each of these three cases, while in fact v_1 and v_2 will vary randomly.

Substituting (9) into (3), we can see how the events fall into two groups. The rates of events affecting the virtual lineages, given in lines 6–13 of (3), depend either directly on σ or indirectly on σ through v_1 and v_2 (we expect v_1 and v_2 to grow large when σ is large, as in Fig. 3). For the sake of illustration, let us ignore the $o(1)$ parts of the ratios of sampling probabilities (9). Then, the fast process of branching, coalescence, and mutation affecting virtual lineages has transitions from (r_1, r_2, v_1, v_2) to

$$\begin{array}{ll}
(r_1, r_2, v_1 - 1, v_2) & \text{with rate} \quad 2r_1v_1 + v_1(v_1 - 1) \\
(r_1, r_2, v_1, v_2 - 1) & \text{with rate} \quad 2r_2v_2 + v_2(v_2 - 1) \\
(r_1, r_2, v_1 - 1, v_2 + 1) & \text{with rate} \quad v_1\theta\alpha_1/2 \\
(r_1, r_2, v_1 + 1, v_2 - 1) & \text{with rate} \quad v_2\theta\alpha_2/2 \\
(r_1, r_2, v_1, v_2) & \text{with rate} \quad v_1\theta\alpha_1/2 + v_2\theta\alpha_2/2 \\
(r_1, r_2, v_1 + 2, v_2) & \text{with rate} \quad (r_1 + v_1 + 2r_2 + 2v_2)\sigma/8 \\
(r_1, r_2, v_1 + 1, v_2 + 1) & \text{with rate} \quad (r_1 + v_1 + r_2 + v_2)\sigma/8 \\
(r_1, r_2, v_1, v_2 + 2) & \text{with rate} \quad (2r_1 + 2v_1 + r_2 + v_2)\sigma/8
\end{array} \quad (10)$$

and it is this process that we expect to approach stationarity rapidly when σ is large. In contrast, the rates of events among the real lineages, given in lines 1–5 of (3), appear to depend only weakly on σ, v_1, and v_2, through the $o(1)$ terms in (9) which are detailed in the Appendix. Then, among the real lineages we have slow transitions from (r_1, r_2, v_1, v_2) to

$$\begin{array}{ll}
(r_1 - 1, r_2, v_1, v_2) & \text{with rate} \quad r_1(r_1 - 1) \\
(r_1, r_2 - 1, v_1, v_2) & \text{with rate} \quad r_2(r_2 - 1) \\
(r_1 - 1, r_2 + 1, v_1, v_2) & \text{with rate} \quad r_1\theta\alpha_1/2 \\
(r_1 + 1, r_2 - 1, v_1, v_2) & \text{with rate} \quad r_2\theta\alpha_2/2 \\
(r_1, r_2, v_1, v_2) & \text{with rate} \quad r_1\theta\alpha_1/2 + r_2\theta\alpha_2/2
\end{array} \quad (11)$$

and it is these rates (but including the $o(1)$ parts from the Appendix) which we should average over the stationary distribution of v_1 and v_2 if we are to follow the separation-of-time-scales approach given in (8).

Looking at (10) we can see that branching will dominate the very recent ancestral process when σ is large and that the time between these events will be very short due to (4). This will cause (v_1, v_2), to grow quickly from the initial sample value of $(0, 0)$, leading to a rapid increase in the rates of mutation and coalescence affecting virtual lineages. A balance will be achieved in which $(v_1, v_2) \sim (\sigma, \sigma)$. To see this, let v'_1 and v'_2 be the numbers of virtual lines when the next event occurs in the ancestry. Considering only the largest rates (of coalescence and branching) in the fast process, we have

$$
\begin{aligned}
E[(v_1', v_2')|(v_1, v_2)] &\approx (v_1, v_2) + (-1, 0)v_1^2/\lambda_{v_1,v_2} \\
&+ (0, -1)v_2^2/\lambda_{v_1,v_2} + (2, 0)(v_1 + 2v_2)\sigma/8\lambda_{v_1,v_2} \\
&+ (1, 1)(v_1 + v_2)\sigma/8\lambda_{v_1,v_2} + (0, 2)(2v_1 + v_2)\sigma/8\lambda_{v_1,v_2}
\end{aligned}
$$

where $\lambda_{v_1,v_2} = v_1^2 + v_2^2 + (v_1 + v_2)\sigma/2$. Thus we have $E[(v_1', v_2')|(\sigma, \sigma)] \approx (\sigma, \sigma)$. This is consistent with the result from the unconditional ASG, that the number of virtual lineages will at equilibrium follow a "zero-truncated" Poisson(σ) distribution (Mano, 2009), which has mean equal to $\sigma - 1 + \sigma/(e^\sigma - 1)$. Again, the simulations presented in Section 2.4 support these ideas (see Fig. 3).

Of the three cases for which we have obtained approximations to the ratio of sampling probabilities in the Appendix, probably the most important is case (iii), in which we have assumed that v_1 and v_2 are both $O(\sigma)$, because this approximation should be valid once v_1 and v_2 have reached their equilibrium. Specifically, in case (iii) we have assumed that $v_1 = \sigma + c_1\sqrt{\sigma}$ and $v_2 = \sigma + c_2\sqrt{\sigma}$, so the resulting series expansions for the ratios are given in terms of the rescaled virtual parameters, $c_1 = (v_1 - \sigma)/\sqrt{\sigma}$ and $c_2 = (v_2 - \sigma)/\sqrt{\sigma}$. Then, analogously to (8), we should average the rates of events among the real lineages over the equilibrium distribution of c_1 and c_2 (or v_1 and v_2).

For example, averaging the rate of a coalescent event between two real A_1 lineages over the equilibrium distribution of c_1 and c_2, denoted $p(c_1, c_2)$ below, we have, approximately, $r_1(r_1 - 1)$ times

$$
\sum_{c_1,c_2} p(c_1, c_2)\left(1 - \frac{2(c_1 - c_2)}{5\sqrt{\sigma}} + \frac{2(4c_1^2 - 4c_1c_2 + 5(1 - r_1 + r_2 - (\alpha_1 - \alpha_2)\theta))}{25\sigma}\right).
$$

Taking the sum inside the parentheses, we have

$$
1 - \frac{2(E[c_1] - E[c_2])}{5\sqrt{\sigma}} + \frac{2(4\mathrm{Var}[c_1] - 4\mathrm{Cov}[c_1, c_2] + 5(1 - r_1 + r_2 - (\alpha_1 - \alpha_2)\theta))}{25\sigma}, \quad (12)
$$

and as long as these moments of c_1 and c_2 are bounded as $\sigma \to \infty$, we obtain simply $r_1(r_1 - 1)$ as the limiting rate of type-1 coalescence. This is identical to the first line of (11). Similar terms arise in the analysis of the rates of the other possible events among the real lineages. Averaging each of these and taking the limit gives exactly the rates in (11). Therefore, we predict that the limiting ancestral process among the real lineages is the structured coalescent described by those rates.

We note that this is identical to the result that Kaplan et al. (1988) obtained by invoking the Gaussian diffusion of Norman (1975) to validate their assumption that the frequencies of the two alleles are given by the deterministic prediction. For example, compare the rates in (11) to those comprising $h_{ij}(x)$ on page 823 of Kaplan et al. (1988), together with the following equivalence between our notation and theirs: $i = r_1$, $j = r_2$, $\beta_2 = \theta\alpha_1/2$, $\beta_1 = \theta\alpha_2/2$, and $x = 1/2$.

2.4. Comparing simulations to limiting results for $\sigma = 0$ and $\sigma \to \infty$

We used two different simulation approaches to investigate the convergence of the conditional ancestral process, given by (3), to the predictions of the limiting model obtained in the previous section, and to characterize the fast process described above. The first program was written in Mathematica (Wolfram, 1999) and simulated the exact process with the rate in (3) using numerical integration to compute the sampling probabilities. This was only feasible for σ up to about 100, aided by the fact that we stopped the simulation once the first event occurred among the real lineages. The second program was written in the C programming language, and used the expressions in the Appendix to approximate the ratios of sampling probabilities. This allowed somewhat larger values of σ to be investigated, but also becomes exceedingly slow when σ is very large due to the huge numbers of fast events that occur before the first event among the real lineages is observed. Both programs are available from the authors upon request.

We used these programs to ask whether the requirements of the separation-of-time-scales method used above appear correct, and whether the predictions of the limiting ancestral process given by (11) are approached for finite, large σ. In order to illustrate the second points, we assumed a sample with visibly different predictions under $\sigma = 0$ and $\sigma \to \infty$. Specifically, we assumed a sample of just two A_1 alleles, (2, 0, 0, 0), and with $\theta = 1$ and $\alpha_1 = 2/3$. In this case, the expected time back to the first event among the real lineages (excluding empty mutation events) is equal to 0.75 under neutrality and 0.375 in the strong-selection limit. These values are simply ones over the sums of the rates of all relevant events, using (7) and (11) respectively.

Fig. 3 shows the rapid growth of v_1 and subsequent variation around its expected value σ, over one run of the process starting from the sample (2, 0, 0, 0) and for $\sigma = 1000$, plotted back to the expected time to an event among the real lineages. The same kind of behavior is, of course, observed for v_2. Over many simulation runs for several large values of σ, we observed that $E[v_1] = E[v_2] \sim \sigma$, $\mathrm{Var}[v_1] = \mathrm{Var}[v_2] \sim 3\sigma$ and $\mathrm{Cov}[v_1, v_2] \sim -2\sigma$. In the absence of direct knowledge about the equilibrium distribution of (v_1, v_2), we take this together with Fig. 3 as support that the separation-of-time-scales argument above is reasonable in this case.

To translate this into the rescaled parameters $c_1 = (v_1 - \sigma)/\sqrt{\sigma}$ and $c_2 = (v_2 - \sigma)/\sqrt{\sigma}$, in the present case, with $\sigma = 1000$, $\theta = 1$, $\alpha = 2/3$, and sample (2, 0, 0, 0), over one million samples of v_1 and v_2 we found $\overline{c_1} = -0.09$, $\overline{c_2} = 0.06$, $\overline{c_1^2} = 3.1$, $\overline{c_2^2} = 3.2$, $\overline{c_1c_2} = -1.6$. This lends support to our conclusion that (12) converges to 1 as σ tends to infinity.

Fig. 4 shows the average time back to an event involving the real lineages in the sample as a function of σ. We consider only events that change the state of the sample, so we ignore the empty mutation events. For the sample (2, 0, 0, 0) and with $\theta = 1$ and $\alpha_1 = 2/3$, then from (11) the limiting rate of coalescence is equal to 2 and the limiting rate of non-empty mutation is 2/3, so the expected time back to one of these events is $3/8 = 0.375$. Fig. 4 shows a relatively rapid shift of the average time from the neutral expectation of 0.75 to this limiting $\sigma \to \infty$ prediction, which occurs between about $\sigma = 0.1$ and $\sigma = 100$. In addition, there is good agreement between the result of the two different methods of computing the ratios of sampling probabilities in the area where the simulations overlapped.

Fig. 5 shows the fraction of simulation replicates in which the first event involving the real lineages was a coalescent event. The probabilities of this under neutrality and in the $\sigma \to \infty$ limit are 0.9 and 0.75, respectively. Again, the conditional ASG mimics the neutral model when $\sigma = 0.1$, and is very close to the $\sigma \to \infty$ prediction when $\sigma = 100$, with a rapid shift in between. Both here and in Fig. 4 the results of the "exact" program (circles) and the approximate program (crosses) agree well in the range where they overlap.

Fig. 6 compares the cumulative distribution function (CDF) of the time to an event among the real lineages in simulations to the CDF of the exponential distribution. Because the rates in (11) are constant, we expect the time to an event to follow an exponential distribution with mean equal to the sum of the rates (again ignoring empty mutation events). As above, we simulated the ancestry of sample (2, 0, 0, 0), with $\theta = 1$ and $\alpha_1 = 2/3$, and in this case only four different values of σ. First, we computed

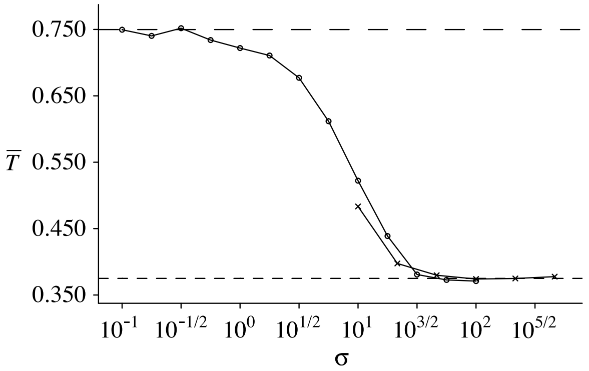

Fig. 4. The average over 10 000 replicates of the time back to the first (non-empty) event among the real lineages, starting from the sample (2, 0, 0, 0) and with $\theta = 1$ and $\alpha_1 = 2/3$, for a range from small to large σ. Neutral and structured-coalescent predictions are given by long-dashed and short-dashed lines, respectively. Circles display the results of the exact simulation, using numerical integration in (3), and crosses display the results of the approximate simulations, using the expressions in the Appendix.

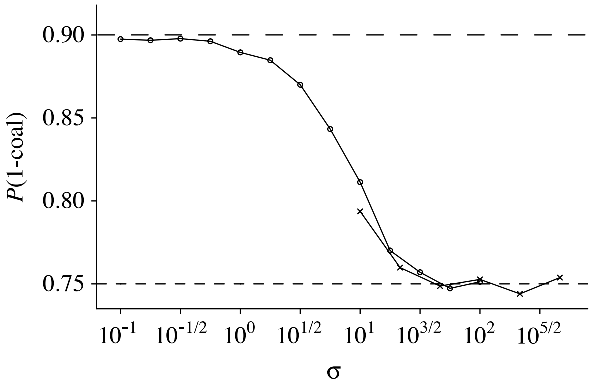

Fig. 5. The fraction of times, out of 10 000 replicates, that the first (non-empty) event among the real lineages was a coalescent event, starting from the sample (2, 0, 0, 0) and with $\theta = 1$ and $\alpha_1 = 2/3$, over a range from small to large σ. Neutral and structured-coalescent predictions are given by long-dashed and short-dashed lines, respectively. Circles display the results of the exact simulation, using numerical integration in (3), and crosses display the results of the approximate simulations, using the expressions in the Appendix.

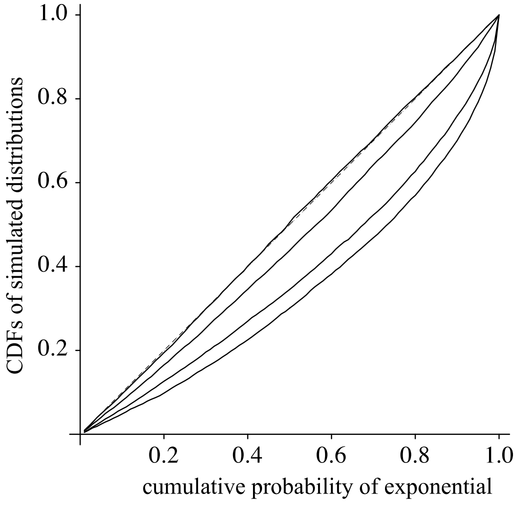

Fig. 6. Comparison of the cumulative distribution functions (CDFs) of the time to an event among the real lines observed in simulations to the CDF of the exponential distribution. The ancestry of a sample (2, 0, 0, 0), with $\theta = 1$ and $\alpha_1 = 2/3$, was simulated 10 000 times each for four different strengths of selection: $\sigma \approx 56$, $\sigma \approx 18$, $\sigma \approx 5.6$, $\sigma \approx 1.8$ (from the diagonal out). A dashed line is shown on the diagonal to indicate a perfect fit to the exponential distribution. For larger σ than 56 it would also be expected that the fit would be good.

cutoffs for the exponential (mean 1) distribution, to create 100 bins of equal probability. Then, for each simulation replicate, we divided the observed time back to an event by the expected time (0.375), compared this to the cutoffs, and added 1 to the number of observations in the appropriate bin. Fig. 6 plots the cumulative probability for the exponential versus that observed in simulations. When $\sigma = 10^{1.75} \approx 56$, the CDFs match very closely (along the diagonal), while when $\sigma = 10^{0.25} \approx 1.8$, the simulated distribution of times is quite different (has a longer tail) than the exponential distribution. These simulations used the "exact" program, integrating numerically to compute sampling probabilities.

3. Discussion

The results we have presented here suggest that, via a separation of time scales between events affecting virtual lineages and events affecting real lineages, the conditional ancestral selection graph with symmetric balancing selection converges to a structured coalescent process in the limit as the selection parameter σ tends to infinity. The form of the structured coalescent is such that the two allelic types comprise two subpopulations, each with one half the size of the total population. Hence, the rates of coalescence in (11) are twice the rates of the Kingman coalescent. In addition, mutation between alleles plays the role of migration between the two subpopulations.

The form of the limiting rates in (11) is identical to those in Kaplan et al. (1988) and Hudson and Kaplan (1988), who derived their results by assuming that the allele frequency is held constant at its equilibrium expected value, with justification provided by Norman (1975). Thus, our results are those we anticipated. What we have shown, both in a heuristic analysis and in simulations, is that this limiting model arises from within the complicated dynamics of the conditional ancestral selection process. This happens despite the confounding fact that the numbers of virtual lineages become enormous when σ is large.

The programs used here allow the rate of convergence to the structured coalescent model of Kaplan et al. (1988) to be studied as σ grows. At least on a log scale, convergence looks rapid, appearing to contradict the statement in Barton and Etheridge (2004) that "balancing selection must be extremely strong for the deterministic limit to be accurate" (see their Figure 10). However, the results plotted in Figs. 4–6 imply that $\sigma = 100$ may be sufficient to be very close the deterministic limit. Although we might legitimately say that $\sigma = 100$ represents extremely strong balancing selection, it is useful to know that convergence appears to be achieved at this point rather than only for truly huge values of σ.

Together with other recent results (Wakeley, 2008a), our findings suggest that other strong-selection limits may exist, for example in the more general models, such as in Donnelly and Kurtz (1999), and that these may be equivalent to some neutral models that have already been described.

Acknowledgments

We thank Tom Kurtz and Anja Sturm for helpful discussions.

Appendix

Here present three different approximations of the ratios of sampling probabilities in (3). As in the main text, we represent the ratio generally as

$$\frac{p_{\sigma,\theta}(r_1', r_2', v_1', v_2')}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)},$$

and recall that the only possible values for the differences $r_1' - r_1$, $r_2' - r_2$, $v_1' - v_1$, and $v_2' - v_2$ are -1, 0, 1, and 2. We also use the notation $k_1 = v_1' - v_1$ and $k_2 = v_2' - v_2$. We obtained all of the approximations below under the assumption that r_1, r_2, θ, and α_1 (and of course $\alpha_2 = 1 - \alpha_1$) are constant. Our analysis of the ratios of sampling probabilities utilizes a change of variable $x = 1/2 + z/\sqrt{\sigma}$ in the sampling-probability integral

$$\int_0^1 x^{\theta\alpha_1+r_1+v_1-1}(1-x)^{\theta\alpha_2+r_2+v_2-1}e^{-\sigma(x^2+(1-x)^2)/2}dx$$

to give

$$\left(\frac{1}{2}\right)^{r_1+r_2+v_1+v_2}\int_{-\sqrt{\sigma}/2}^{\sqrt{\sigma}/2}\left(1+\frac{2z}{\sqrt{\sigma}}\right)^{\theta\alpha_1+r_1+v_1-1} \times \left(1-\frac{2z}{\sqrt{\sigma}}\right)^{\theta\alpha_2+r_2+v_2-1} e^{-z^2}dz \qquad (13)$$

where we have factored out and ignored the term

$$\frac{e^{-\sigma/4}}{2^{\theta-2}\sqrt{\sigma}},$$

which does not depend on r_1, r_2, v_1, and v_2, and which like the constant C in (5) will appear in both the numerator and the denominator of every ratio of sampling probabilities.

Our method was to obtain series expansions of the terms inside the integral for large σ, and under different assumptions about the magnitudes of v_1 and v_2, then evaluate the integral. Substituting the resulting expressions into the numerator and denominator of the ratio of two sampling probabilities then gave large-σ series expansions for the ratios. We considered three cases: (i) v_1 and v_2 both finite, (ii)

$$v_1 = c_1\sqrt{\sigma}$$
$$v_2 = c_2\sqrt{\sigma}$$
$$v_1' = c_1\sqrt{\sigma} + k_1$$
$$v_2' = c_2\sqrt{\sigma} + k_2,$$

and (iii)

$$v_1 = \sigma + c_1\sqrt{\sigma}$$
$$v_2 = \sigma + c_2\sqrt{\sigma}$$
$$v_1' = \sigma + c_1\sqrt{\sigma} + k_1$$
$$v_2' = \sigma + c_2\sqrt{\sigma} + k_2,$$

the derivations of which are lengthy, and were done with the aid of the program Mathematica (Wolfram, 1999). Briefly, for each of the three cases above, we substituted the assumed values of v_1 and v_2 or v_1' and v_2' inside the integral in (13), expanded in terms of σ up to order $1/\sigma$, then evaluated the integral, ignoring terms that approach zero faster than $1/\sigma$. We then took the ratio of two sampling probabilities, $p_{\sigma,\theta}(r_1', r_2', v_1', v_2')$ and $p_{\sigma,\theta}(r_1, r_2, v_1, v_2)$, again expanding in terms of σ up to order $1/\sigma$. The Mathematica notebooks containing these derivations are available from the authors upon request.

In case (i), where v_1 and v_2 are both finite, we have

$$\begin{aligned}\frac{p_{\sigma,\theta}(r_1', r_2', v_1', v_2')}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} &= 2^{r_1+r_2-r_1'-r_2'-k_1-k_2}\\ &\times (1 + \sigma^{-1}((\theta\alpha_1 + r_1' + v_1' - 1)(\theta\alpha_1 + r_1' + v_1' - 2)\\ &- (\theta\alpha_1 + r_1 + v_1 - 1)(\theta\alpha_1 + r_1 + v_1 - 2)\\ &+ 2(\theta\alpha_1 + r_1 + v_1 - 1)(\theta\alpha_2 + r_2 + v_2 - 1)\\ &- 2(\theta\alpha_1 + r_1' + v_1' - 1)(\theta\alpha_2 + r_2' + v_2' - 1)\\ &+ (\theta\alpha_2 + r_2' + v_2' - 1)(\theta\alpha_2 + r_2' + v_2' - 2)\\ &- (\theta\alpha_2 + r_2 + v_2 - 1)(\theta\alpha_2 + r_2 + v_2 - 2))).\end{aligned}$$

In case (ii), where v_1 and v_2 are both of order $\sqrt{\sigma}$, we have

$$\frac{p_{\sigma,\theta}(r_1', r_2', v_1', v_2')}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} = 2^{r_1+r_2-r_1'-r_2'-k_1-k_2}\left(1 + W_1/\sqrt{\sigma} + W_2/\sigma\right)$$

where

$$W_1 = 2(c_1 - c_2)(k_1 - k_2 - r_1 + r_1' + r_2 - r_2')$$

and

$$\begin{aligned}W_2 = {} & 2\theta(\alpha_1 - \alpha_2)(k_1 - k_2 - r_1 + r_1' + r_2 - r_2')\\ &+ (k_1 - k_2 + r_1' - r_2')^2 - (r_1 - r_2)^2 + r_1 + r_2 - k_1 - k_2\\ &- r_1' - r_2' + 2(c_1 - c_2)^2(k_1^2 - k_2 + r_1 - r_1' + r_2 - r_2'\\ &+ (k_2 + r_1 - r_1' - r_2 + r_2')^2)\\ &- 2(c_1 - c_2)^2k_1(1 + 2k_2 + 2r_1 - 2r_1' - 2r_2 + 2r_2')\\ &+ 4(c_1 - c_2)(c_1 + c_2)(k_2 - k_1 + r_1 - r_1' - r_2 + r_2')\end{aligned}$$

and $c_1 = v_1/\sqrt{\sigma}$ and $c_2 = v_2/\sqrt{\sigma}$.

In case (iii), where v_1 and v_2 are both of order σ with deviations of order $\sqrt{\sigma}$, we have

$$\frac{p_{\sigma,\theta}(r_1', r_2', v_1', v_2')}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} = 2^{r_1+r_2-r_1'-r_2'-k_1-k_2}\left(1 + \frac{2}{5}(W_1/\sqrt{\sigma} + W_2/\sigma)\right)$$

where

$$W_1 = (c_1 - c_2)(k_1 - k_2 - r_1 + r_1' + r_2 - r_2')$$

and

$$\begin{aligned}W_2 = {} & \theta(\alpha_1 - \alpha_2)(k_1 - k_2 - r_1 + r_1' + r_2 - r_2')\\ &+ (k_1 - k_2)(r_1' - r_2') - k_1k_2 - r_1'r_2' + r_1r_2 + k_1(k_1 - 1)/2\\ &+ k_2(k_2 - 1)/2 + r_1'(r_1' - 1)/2 + r_2'(r_2' - 1)/2\\ &- r_1(r_1 - 1)/2 - r_2(r_2 - 1)/2 + (c_1 - c_2)^2(k_1 - k_2 - r_1\\ &+ r_1' + r_2 - r_2')(k_1 - k_2 - r_1 + r_1' + r_2 - r_2')\\ &+ (c_1 - c_2)c_1(k_2 + r_2' - r_2 + 3(r_1 - k_1 - r_1'))\\ &- (c_1 - c_2)c_2(k_1 + r_1' - r_1 + 3(r_2 - k_2 - r_2'))\end{aligned}$$

and $c_1 = (v_1 - \sigma)/\sqrt{\sigma}$ and $c_2 = (v_2 - \sigma)/\sqrt{\sigma}$.

By assumption, in both cases (ii) and (iii) above c_1 and c_2 are finite constants. In fact, during the ancestry of the sample, these rescaled numbers of virtual lineage will vary randomly, and we have to account for this in describing the limiting $\sigma \to \infty$ ancestral process. For now, we can say that, given c_1 and c_2, in all three cases we have

$$\frac{p_{\sigma,\theta}(r_1', r_2', v_1', v_2')}{p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} = 2^{r_1+r_2-r_1'-r_2'-k_1-k_2}(1 + o(1))$$

where $k_1 = v_1' - v_1$ and $k_2 = v_2' - v_2$. In case (i) the $o(1)$ term is $O(1/\sigma)$, while in cases (ii) and (iii) it is $O(1/\sqrt{\sigma})$.

References

Asthana, S., Schmidt, S., Sunyaev, S., 2005. A limited role for balancing selection. Trends Genet. 21, 30–32.

Baake, E., Bialowons, R., 2008. Ancestral processes with selection: Branching and Moran models. In: Miekisz, J. (Ed.), Banach Center Publications, vol. 80. Institute of Mathematics. Polish Academy of Sciences, Warsaw, pp. 33–52.

Barton, N.H., Etheridge, A.M., 2004. The effect of selection on gene genealogies. Genetics 166, 1115–1131.

Barton, N.H., Etheridge, A.M., Sturm, A.K., 2004. Coalescence in a random background. Ann. Appl. Prob. 14, 754–785.

Bubb, K.L., Bovee, D., Buckley, D., Haugen, E., Kibukawa, M., Paddock, M., Palmieri, A., Subramanian, S., Zhou, Y., Kaul, R., Green, P., Olsen, M.V., 2006. Scan of the human genome reveals no new loci under ancient balancing selection. Genetics 173, 2165–2177.

Charlesworth, D., Kamau, E., Hagenblad, J., Tang, C., 2006. *Trans*-specificity at loci near the self-incompatibility locus in Arabidopsis. Genetics 172, 2699–2704.

Donnelly, P., Kurtz, T.G., 1999. Genealogical models for Fleming–Viot models with selection and recombination. Ann. Appl. Probab. 9, 1091–1148.

Ewens, W.J., 2004. Mathematical Population Genetics, Volume I: Theoretical Foundations. Springer-Verlag, Berlin.

Fearnhead, P., 2002. The common ancestor at a nonneutral locus. J. Appl. Probab. 39, 38–54.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Clarendon, Oxford.

Haldane, J.B.S., 1932. The Causes of Natural Selection. Longmans Green & Co, London.

Herbots, H.M., 1997. The structured coalescent. In: Progress in Population Genetics and Human Evolution. In: Donnelly, P., Tavaré, S. (Eds.), IMA Volumes in Mathematics and its Applications, vol. 87. Springer-Verlag, New York, pp. 231–255.

Hudson, R.R., Kaplan, N.L., 1988. The coalescent process in models with selection and recombination. Genetics 120, 831–840.

Kamau, E., Charlesworth, B., Charlesworth, D., 2007. Linkage disequilibirum and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. Genetics 176, 2357–2369.

Kaplan, N.L., Darden, T., Hudson, R.R., 1988. Coalescent process in models with selection. Genetics 120, 819–829.

Kaplan, N.L., Hudson, R.R., Langley, C.H., 1989. The hitchhiking effect revisited. Genetics 123, 887–899.

Karlin, S., McGregor, J., 1964. On some stochastic models in genetics. In: Gurland, J. (Ed.), Stochastic Models in Medicine and Biology. The University of Wisconsin Press, Madison, pp. 245–271.

Kingman, J.F.C., 1982a. The coalescent. Stochastic Process. Appl. 13, 235–248.

Kingman, J.F.C., 1982b. On the genealogy of large populations. J. Appl. Prob. 19A, 27–43.

Kingman, J.F.C., 1982c. Exchangeability and the evolution of large populations. In: Koch, G., Spizzichino, F. (Eds.), Exchangeability in Probability and Statistics. North-Holland, Amsterdam, pp. 97–112.

Krone, S.M., Neuhauser, C., 1997. Ancestral processes with selection. Theoret. Pop. Biol. 51, 210–237.

Lessard, S., Wakeley, J., 2004. The two-locus ancestral graph in a subdivided population: Convergence as the number of demes grows in the island model. J. Math. Biol. 48, 275–292.

Mano, S., 2009. Duality, ancestral and diffusion processes in models with selection. Theoret. Pop. Biol. doi:10.1016/j.tpb.2009.01.007.

Möhle, M., 1998. A convergence theorem for Markov chains arising in population genetics and the coalescent with partial selfing. Adv. Appl. Prob. 30, 493–512.

Moran, P.A.P., 1958. Random processes in genetics. Proc. Camb. Phil. Soc. 54, 60–71.

Moran, P.A.P., 1962. Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

Neuhauser, C., Krone, S.M., 1997. The genealogy of samples in models with selection. Genetics 145, 519–534.

Norman, M.F., 1975. Approximation of stochastic processes by Gaussian diffusions, and applications to Wright–Fisher genetic models. SIAM J. Appl. Math. 29, 225–242.

Norris, J.R., 1997. Markov Chains. Cambridge University Press, Cambridge.

Notohara, M., 1990. The coalescent and the genealogical process in geographically structured population. J. Math. Biol. 29, 59–75.

Richman, A.D., Uyenoyama, M.K., Kohn, J.R., 1996. Allelic diversity and gene genealogy at the self-incompatibility locus in Solanaceae. Science 273, 1212–1216.

Slade, P.F., 2000a. Simulation of selected genealogies. Theoret. Pop. Biol. 57, 35–49.

Slade, P.F., 2000b. Most recent common ancestor distributions in genealogies under selection. Theoret. Pop. Biol. 58, 291–305.

Stephens, M., Donnelly, P., 2003. Ancestral inference in population genetics models with selection. Aust. N. Z. J. Stat. 45, 395–430.

Takahata, N., 1988. The coalescent in two partially isolated diffusion populations. Genet. Res., Camb. 53, 213–222.

Takahata, N., 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species polymorphism. Proc. Natl. Acad. Sci, USA 87, 2419–2423.

Vekemans, X., Slatkin, M., 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. Genetics 137, 1157–1165.

Wakeley, J., 2008a. Conditional gene genealogies under strong purifying selection. Mol. Bol. Evol. 25, 2615–2626.

Wakeley, J., 2008b. Coalescent Theory: An Introduction. Roberts & Company Publishers, Greenwood Village, Colorado.

Wolfram, S., 1999. The Mathematica Book, 4th edition. Wolfram Media/Cambridge University Press, Cambridge, UK.

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

Wright, S., 1939. The distribution of self-sterility alleles in populations. Genetics 24, 538–552.

Wright, S., 1949. Adaptation and selection. In: Jepson, G.L., Simpson, G.G., Mayr, E. (Eds.), Genetics, Paleontology and Evolution. Princeton Univ. Press, Princeton.

Wright, S., 1969. Evolution and the genetics of populations. In: Vol. 2: The Theory of Gene Frequencies. University of Chicago Press, Chicago.