



# 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.

© 2009 Elsevier Inc. All rights reserved.

## 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 ( $A_1, A_2, \dots, A_K$ ) 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 –  $\theta$ ,  $\alpha_i$ , and  $\sigma(A_i, A_j)$ , for  $i, j = 1, \dots, K$  – 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](mailto:wakeley@fas.harvard.edu) (J. Wakeley).

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  $\theta$  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

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

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

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

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

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

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

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

such ordered samples and each one has the same probability, given by (2). The subscripts  $\theta$  and  $\sigma$  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  $\theta \rightarrow \infty$  (with  $\sigma = 0$ ) and  $\sigma \rightarrow \infty$  (with  $\theta$  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 exist 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  $\sigma \rightarrow \infty$  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  $\sigma$ .

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  $\sigma \rightarrow \infty$  (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 ( $\sigma \rightarrow \infty$ ) 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

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

$$\begin{aligned} (r_1, r_2, v_1, v_2) & \text{ with rate } v_1 \frac{\theta\alpha_1}{2} + v_2 \frac{\theta\alpha_2}{2} \\ (r_1, r_2, v_1 + 2, v_2) & \text{ with rate } (r_1 + v_1 + 2r_2 + 2v_2) \frac{\sigma p_{\sigma,\theta}(r_1, r_2, v_1 + 2, v_2)}{2 p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} \\ (r_1, r_2, v_1 + 1, v_2 + 1) & \text{ with rate } (r_1 + v_1 + r_2 + v_2) \frac{\sigma p_{\sigma,\theta}(r_1, r_2, v_1 + 1, v_2 + 1)}{2 p_{\sigma,\theta}(r_1, r_2, v_1, v_2)} \\ (r_1, r_2, v_1, v_2 + 2) & \text{ with rate } (2r_1 + 2v_1 + r_2 + v_2) \frac{\sigma p_{\sigma,\theta}(r_1, r_2, v_1, v_2 + 2)}{2 p_{\sigma,\theta}(r_1, r_2, v_1, v_2)}. \end{aligned}$$

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

$$\begin{aligned} \left( \frac{r_1 + v_1 + r_2 + v_2}{2} \right) + (r_1 + v_1 + r_2 + v_2) \frac{\theta}{2} \\ + (r_1 + v_1 + r_2 + v_2) \frac{\sigma}{2}. \end{aligned} \tag{4}$$

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)  $\times P\{\text{Data}|\text{Event}\}/P\{\text{Data}\}$ , 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  $\theta \rightarrow \infty$  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  $\sigma \rightarrow \infty$ . Since  $\sigma = 0$ ,  $v_1 = 0$ , and  $v_2 = 0$ , 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  $\theta$  is large. The limit of the sampling probability is straightforward to obtain, and is

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

Thus, as  $\theta$  grows, each sample independently has probability  $\alpha_1$  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  $\theta$  grows. The shape of this distribution when  $\theta$  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  $\theta$  for mutation or  $\sigma$  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  $\theta$  grows. From  $\phi_{\theta}(x)$ , we obtain

$$E[X] = \alpha_1$$

and

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

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  $\theta$  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  $\theta$  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 ( $\sigma = 0$ ,  $v_1 = 0$ , and  $v_2 = 0$ ) 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

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

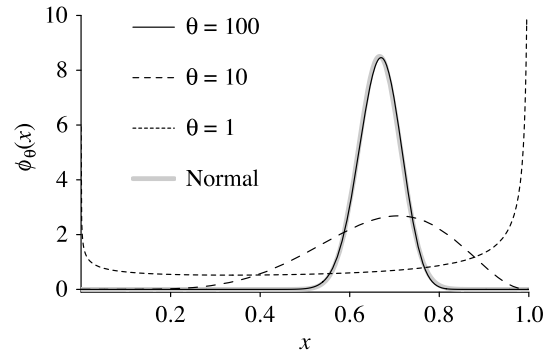


Fig. 1. Plots of the equilibrium distribution of  $x$ , the frequency of allele  $A_1$ , for three different strengths of mutation under neutrality. A normal distribution with mean  $E[X] = 2/3$  and variance  $\text{Var}[X] = 2/909$ , obtained using the equations in the text, is shown for comparison.

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  $\theta \rightarrow \infty$ , and is exponential with rate  $r(r - 1)/2$ , as in Kingman’s (unconditional) coalescent.

In studying the limit  $\theta \rightarrow \infty$ , 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

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

and

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

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

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

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

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

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











