

The Structured Ancestral Selection Graph and the Many-Demes Limit

Paul F. Slade*¹ and John Wakeley[†]

*School of Mathematics and Statistics, School of Biological Sciences, SUBIT, University of Sydney, New South Wales 2006, Australia and
[†]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received June 21, 2004

Accepted for publication October 28, 2004

ABSTRACT

We show that the unstructured ancestral selection graph applies to part of the history of a sample from a population structured by restricted migration among subpopulations, or demes. The result holds in the limit as the number of demes tends to infinity with proportionately weak selection, and we have also made the assumptions of island-type migration and that demes are equivalent in size. After an instantaneous sample-size adjustment, this structured ancestral selection graph converges to an unstructured ancestral selection graph with a mutation parameter that depends inversely on the migration rate. In contrast, the selection parameter for the population is independent of the migration rate and is identical to the selection parameter in an unstructured population. We show analytically that estimators of the migration rate, based on pairwise sequence differences, derived under the assumption of neutrality should perform equally well in the presence of weak selection. We also modify an algorithm for simulating genealogies conditional on the frequencies of two selected alleles in a sample. This permits efficient simulation of stronger selection than was previously possible. Using this new algorithm, we simulate gene genealogies under the many-demes ancestral selection graph and identify some situations in which migration has a strong effect on the time to the most recent common ancestor of the sample. We find that a similar effect also increases the sensitivity of the genealogy to selection.

KIMURA (1983) strongly promoted the idea that the abundant genetic variation seen in nearly every species studied must be neutral. OHTA and KIMURA (1971) suggested a less strict version of this in which polymorphisms are explained by the constant input of nearly neutral mutations, *i.e.*, variants with selective advantages or disadvantages smaller than the reciprocal of the population size. After a great deal of debate and many analyses, summarized in GOLDING (1994), there are now few adherents to the strict neutral mutation hypothesis. Molecular techniques currently allow huge numbers of polymorphisms to be assayed with relative ease, and the resulting genomic data can be used to estimate the strength of selection associated with genetic differences (SAWYER and HARTL 1992; BUSTAMANTE *et al.* 2002). Recent statistical inferences made from such data emphasize the importance of selection and show that selective advantages or disadvantages on the order of the reciprocal of the population size may be fairly common (SAWYER *et al.* 2003).

The methods used in these works to estimate selection parameters from intraspecific genomic data rely on predictions of sample allele frequencies derived from forward-time diffusion theory. They do not use the ancestral genetic process known as the coalescent, which was introduced in the early 1980s by KINGMAN (1982a,b),

HUDSON (1983), and TAJIMA (1983). The reason for this is clear: it is much simpler to predict the frequencies of selected alleles using diffusion theory than using the backward-time approach of the coalescent. For neutral loci, the coalescent provides a simple and useful description of the genetic ancestry of a sample of genetic data—the genealogy for short—from a large well-mixed, or panmictic, population of constant size through time. Due to the close connection between genealogies and genetic data, and to the ease with which genealogies can be simulated, a growing set of computational tools use the coalescent to make inferences about population history from DNA sequences; see STEPHENS (2001) and TAVARÉ (2004) for reviews.

The coalescent is robust to many kinds of deviations from its underlying assumptions (KINGMAN 1982b; MÖHLE 1998c), but it does not hold when members of the sample, or the ancestral lineages of the sample, are not exchangeable. In fact, the relative simplicity of the neutral coalescent process flows directly from the exchangeability of the lineages. Exchangeability, which results in the model from the assumptions of neutrality and panmixia, means that the statistical properties of a sample do not depend on how the sampled items are labeled (KINGMAN 1982b; ALDOUS 1985). Possible “labels” include the geographic locations where the samples were collected or the allelic states of the samples if these are known. When samples are not exchangeable, a related but more complicated model called the structured coa-

¹Corresponding author: School of Mathematical Sciences, University of Adelaide, SA 5005, Australia. E-mail: paul.slade@adelaide.edu.au

lescent exists (SLATKIN 1987; STROBECK 1987; NOTOHARA 1990; WILKINSON-HERBOTS 1998), and computational methods of inference are being developed for this model as well (NATH and GRIFFITHS 1996; BEERLI and FELSENSTEIN 1999, 2001; BAHLO and GRIFFITHS 2000; DE IORIO and GRIFFITHS 2004).

Genealogies in the presence of strong selection, *i.e.*, with selective advantages or disadvantages much greater than the reciprocal of the population size, can be modeled using the structured coalescent with the subdivisions being the alternative alleles (KAPLAN *et al.* 1988, 1989). For weak selection, KRONE and NEUHAUSER (1997) and NEUHAUSER and KRONE (1997) showed that the genealogy of a sample taken at random with respect to allelic variation is described by an exchangeable ancestral process. The ancestral selection graph (ASG) is a branching-coalescing random graph within which the simple, bifurcating genealogy of a sample is embedded. The elegant exchangeability of the ASG comes with a cost: it lends itself only to very computationally intensive techniques. In addition, there appears to be very little about genealogies in the ASG that differs greatly from genealogies under neutrality (KRONE and NEUHAUSER 1997).

For these reasons, and also to be able to control for sampling, recent extensions to the ASG have modeled genetic ancestry conditional on the frequencies of selected alleles in the sample (SLADE 2000a,b; STEPHENS and DONNELLY 2003; BARTON *et al.* 2004). This leads to dramatic improvements in computation time while preserving some degree of exchangeability among the sampled lineages (SLADE 2000a). The conditional simulation of selected genealogies by SLADE (2000a,b) enables the timing structure of common ancestry to be quantified. In contrast to the unconditional ASG, genealogies conditional upon allele frequencies in the sample may be either larger or smaller on average than genealogies under neutrality (SLADE 2000a); see also the review of SLADE and NEUHAUSER (2003). The extent of these differences depends strongly on the initial sample configuration of alleles, particularly for small mutation rates. Under neutrality, samples in which there are approximately equal numbers of each allele have longer times to common ancestry than more unbalanced samples. However, the selective effect on genealogies for samples in which there are approximately equal numbers of each allele is greater than that of more unbalanced samples. We show that restricted migration can cause unbalanced samples to be converted to balanced samples, via the instantaneous sample-size adjustment mentioned above, and thus the time to common ancestry under selection is reduced.

Our purpose is to consider the ASG in the presence of population subdivision and migration. Population structure is evident in many genetic data sets but modeling it presents difficulties (SLATKIN 1985; AVISE 2000; HEY and MACHADO 2003), creating impediments both to inference and to understanding. Therefore, it is valu-

able to identify cases in which simple models may be applied even though the underlying dynamics appear complicated. We present a rigorous derivation of the ASG with very many demes (subpopulations) in an island model of migration (WRIGHT 1931) and show that it possesses a simple structure akin to that under neutrality (WAKELEY 1998). Five parameters collapse into just three (one each for migration, selection, and mutation) in the limit as the number of demes tends to infinity. Interestingly, the population rates of selection and mutation scale differently (see below) in the presence of population subdivision. This was found previously in forward-time analyses (WAKELEY 2003; WAKELEY and TAKAHASHI 2004), but the result is more intuitive in the present case. We also describe a newly enhanced process for the simulation of genealogies conditional on the frequencies of alleles in the sample. We show using conditional simulations that population subdivision and migration among very many demes accentuate the action of selection on the genealogy.

THEORY

The structured ancestral selection graph is the ancestral process for a class of forward-time models. Following KRONE and NEUHAUSER (1997) we begin by describing a continuous-time model in which reproduction occurs according to the MORAN (1958, 1962) model, but where the population is subdivided into demes (subpopulations) connected by migration. There are D demes, and in a general model these might have sizes N_i , $i = 1, \dots, D$. We assume that each of the $\sum_{i=1}^D N_i$ lineages in the population can be in one of two possible allelic states, A_1 and A_2 , with relative rates of reproduction λ_1 and λ_2 , respectively. However, as in KRONE and NEUHAUSER (1997), we begin with a description in which the states of lineages are not specified and so each lineage undergoes two kinds of reproduction (birth/death) events. Type 2 birth/death events occur with rate $\lambda_2 - \lambda_1$ per lineage. These represent the action of selection since they are realized only if the lineage that reproduces is of the favored allelic type at the time of the event. Type δ birth/death events occur with rate λ_1 per lineage. The majority of birth/death events will be of this type, and they can be thought of as neutral in the sense that they are realized regardless of the allelic type of the parental lineage. Note that at this point, without specifying the allelic types, lineages are exchangeable within demes but are not exchangeable between demes. Note also that, while this is a model of directional selection, it is possible to include frequency dependence (NEUHAUSER 1999) or general diploid selection (SLADE 2000a).

Subdivision is mediated by a collection of $D \times D$ migration probabilities, m_{ij} , for $i, j = 1, \dots, D$. When a lineage in deme j reproduces, either by a type δ or by a type 2 event, the individual its offspring will replace (*i.e.*, the individual picked to die) is chosen uni-

formly at random from among the N_i lineages in deme i with probability m_{ij} . With probability $1 - m$, where $m = \sum_{j=1}^D m_{ij}$, the individual to die is chosen uniformly at random from within the same deme as the reproducing individual. As usual in Moran-type models, the individual chosen to reproduce might also be the one chosen to die. Although the model does allow “migration” back to the same deme (with probability m_{ii}) the effect of this becomes negligible in the large- D limit we consider below.

The general model described above forms the basis for a structured ancestral selection graph. This would be obtained via the usual large- N limit, with D finite, in the same way that the structured coalescent is obtained in the neutral case (NOTOHARA 1990; WILKINSON-HERBOTS 1998). That is, we would let $N = \sum_{i=1}^D N_i$ and assume that $c_i = N_i/N$, $M_{ij} = N_i m_{ij}$, $\sigma = \lambda_2 - \lambda_1$, and $\theta = Nu$, where u is the neutral mutation rate, are all finite as $N \rightarrow \infty$, and time is measured in units of N generations, where a generation is defined to be λ_1^{-1} time units. Note that a crucial assumption of the structured coalescent is that migration occurs only between the D demes enumerated in the initial sample, and the ancestry is therefore always restricted to those particular demes.

We do not pursue this large- N limit here, but instead consider a limiting ancestral process that approximates the behavior of a population divided into very many demes and in which selective differences are small. Thus, we assume that the sample size is small relative to the number of demes in the population, rather than to the deme size. Historically, migration can take ancestors to any deme within the population. This follows recent coalescent work under the assumption of neutrality (WAKELEY 1998) and work on the forward-time diffusion approximation for allele frequencies in the presence of selection (WAKELEY 2003; WAKELEY and TAKAHASHI 2004). We set $\lambda_2 = (1 + s_D)\lambda_1$, and we use λ_{1s_D} in place of $\lambda_2 - \lambda_1$ for the rate of type 2 birth/death events below. This corresponds to a model in which there are two alleles, A_1 and A_2 , and allele A_2 has fitness equal to $1 + s_D$ while A_1 has fitness equal to 1. We put a subscript on the selection coefficient in recognition of the fact that the limit we seek has Ds (and Du) finite as $D \rightarrow \infty$. Following LESSARD and WAKELEY (2004), the large- N , large- D limit can also be studied. Similar limit results here require more stringent conditions: roughly, that D goes to infinity faster than N (see APPENDIX A).

For simplicity, we assume that migration occurs according to WRIGHT’s (1931) island model, but adapted for Moran-type reproduction as in WAKELEY and TAKAHASHI (2004). That is, we assume that $m_{ij} = m/D$ for all i and j , and that $N_i = N$ for all i . Thus, we have a model in which there are D demes, each of size N , and each of which receives a migrant from the total population with probability m at each birth/death event. As in the unstructured ancestral selection graph, it is helpful to consider a graphical representation, which KRONE and

NEUHAUSER (1997) refer to as the *percolation diagram*. This is shown in Figure 1 and is the structured analog of Figure 1 in KRONE and NEUHAUSER (1997). Following a single lineage, either forward or backward in time, 2-arrows are encountered at rate λ_{1s_D} and δ -arrows are encountered at rate λ_1 . Each arrow has a probability m of connecting to a lineage sampled uniformly at random from the entire population and probability $1 - m$ of connecting to a lineage sampled uniformly at random from the same deme. Thus, there are four kinds of branches and these appear repeatedly in the history of every lineage with rates $\lambda_1(1 - m)$, $\lambda_1 m$, $\lambda_{1s_D}(1 - m)$, and $\lambda_{1s_D} m$. Note that λ_1 is simply a scaling factor that specifies the units in which time will be measured; $\lambda_1 = 1$ corresponds to the usual notion of measuring time in generations.

The *dual* or ancestral process is obtained by following lineages back in time, *i.e.*, up the graph in Figure 1. Because 2-arrows are realized only if the parental allele is of the favored type, both lineages must be followed in the ancestral process. A single genealogy is obtained, in the unconditional ancestral process, by tracing lineages back to the ultimate ancestor, assigning its type from the equilibrium distribution of allele frequencies (see below), and then following lineages forward in time, with mutation, and trimming off branches appropriately (KRONE and NEUHAUSER 1997; NEUHAUSER and KRONE 1997). Rather than this unconditional ancestral process, we consider the genealogy of a sample containing known numbers of different alleles. This leads to more efficient simulation of selected genealogies. Using the conditional approach, it is possible to label ancestral lineages as either *virtual* or *real*, and it is necessary to trace lineages back only to the most recent common ancestor of the sample, *i.e.*, among the *real* lineages, rather than all the way back to the ultimate ancestor (SLADE 2000a). There is also a minimal representation of the ancestral process that reduces the number of possible events required at each ancestral transition (SLADE and NEUHAUSER 2003). We describe how this is done in the present, structured model in METHODOLOGY.

The essence of the large- D result here, as in the neutral case, is that the rates at which events occur that affect the history of a sample from the population are very different when every lineage is in a separate deme than when at least one deme contains more than one lineage. The dependence on D is such that a *separation-of-timescales* result applies, as, for example, in MÖHLE (1998a,b). Because of this, the history of a sample can be divided into two phases (WAKELEY 1999). The first is a short *scattering* phase in which coalescent events occur between samples from the same deme and migration events in which lineages move to unoccupied demes (*i.e.*, demes that do not contain any ancestral lineages). The scattering phase occurs on a fast timescale and becomes an instantaneous sample size adjustment in the limit as D tends to infinity. The scattering phase ends

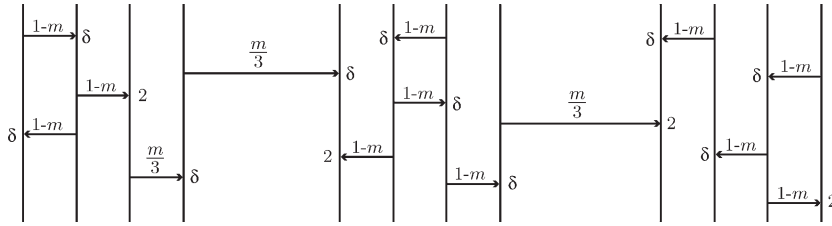


FIGURE 1.—Graphical depiction of the structured selection process for the case of $D = 3$ demes.

when all remaining lineages are in different demes. At this point, the *collecting* phase begins in which pairs of lineages come together into a single deme and coalesce, eventually finding a common ancestor of the entire sample. The migration events that could be ignored in the scattering phase, *i.e.*, those in which lineages move to occupied demes, are now essential since a coalescent event can occur only between a pair of lineages if they are in the same deme. We use the scattering/collecting terminology even though we have incorporated selection.

Forward-time analysis: To simulate genealogies in both the conditional and the unconditional ASG, it is necessary to know the equilibrium distribution of the frequencies of A_1 and A_2 in the population. In fact, the ASG is a model specifically for a sample from such an equilibrium population. There will be no equilibrium without mutation, and here we follow KRONE and NEUHAUSER (1997) in assuming that there is a probability of mutation u_d per birth/death event. We note that asymmetric mutation can also be accommodated (SLADE 2000a). The forward-time dynamics of allele frequencies in a subdivided population may be complicated, but in the case of a large number of demes they are nearly as simple as in an unstructured population (WAKELEY 2003). In a model very similar to the one we consider here, WAKELEY and TAKAHASHI (2004) showed that the frequencies of alleles in the total population change according to a diffusion process that is identical to the diffusion process in an unstructured population, only with a different timescale. Further, as these frequencies change (slowly) the collection of demes closely tracks an equilibrium distribution of allele frequencies.

Let the random variable $X(t)$ be the frequency of allele A_1 at time t , and let x represent a particular value of $X(t)$. Because we consider the limit $D \rightarrow \infty$ while Ds_D and Du_d remain finite, mutation and selection do not appear in the equilibrium

$$v_k = \binom{N}{k} \binom{Nm x}{1-m}_{(k)} \binom{Nm(1-x)}{1-m}_{(N-k)} / \binom{Nm}{1-m}_{(N)} \quad (1)$$

(WAKELEY and TAKAHASHI 2004), which here is the fraction of demes in the population that contain k copies of allele A_1 and in which $a_{(k)} = a(a+1) \cdots (a+k-1)$ is an ascending factorial. Clearly, the equilibrium vector ν does depend on the frequency of allele A_1 in the total population, and this will change over time. However, Theorem 3.3 of ETHIER and NAGYLAKI (1980) can be

used to show that the collection of demes is always sufficiently close to ν that changes in X depend only on the ensemble properties $E_\nu[K] = \sum_{k=0}^N k v_k = Nx$ and $\text{Var}_\nu[K] = x(1-x)N^2/(Nm+1-m)$ and that the diffusion of X is identical to the usual unstructured diffusion except that the timescale is increased by the factor $1 + (1-m)/(Nm)$ (WAKELEY and TAKAHASHI 2004).

By considering overall limiting rates of type δ and type 2 birth/death events—which are $\lambda_1 ND$ and $\lambda_1 s_D ND(1-x)$, respectively—and conditioning on the number of copies of allele A_1 in the deme where the reproduction event occurs, it can be shown that

$$X \rightarrow \begin{cases} X + \frac{1}{ND} & \text{with rate } \lambda_1 ND x(1-x) \frac{Nm}{Nm+1-m} \\ & + \lambda_1 ND(1-x)u_d \left[1 + 2x \frac{Nm}{Nm+1-m} \right] \\ & + O(s_D u_d) \\ X - \frac{1}{ND} & \text{with rate } \lambda_1 ND x(1-x)(1+s_D) \frac{Nm}{Nm+1-m} \\ & + \lambda_1 ND x u_d \left[1 + 2(1-x) \frac{Nm}{Nm+1-m} \right] \\ & + O(s_D u_d). \end{cases} \quad (2)$$

We let

$$\lambda_1 = \frac{ND}{2} \left(1 + \frac{1-m}{Nm} \right) \quad (3)$$

and we further assume that

$$NDs_D \rightarrow \sigma \quad \text{and} \quad NDu_d [1 + (1-m)/(Nm)] \rightarrow \theta \quad \text{as } D \rightarrow \infty. \quad (4)$$

Then, the diffusion process for the frequency of allele A_1 has drift parameter $a(x) = -(\sigma/2)x(1-x) + (\theta/2)(1-2x)$ and diffusion parameter $b(x) = x(1-x)$. This process has a unique stationary distribution

$$h(x) = Bx^{\theta-1}(1-x)^{\theta-1}e^{-\sigma x}, \quad (5)$$

where B is a normalizing constant; that is, $h(x)$ satisfies

$$0 = \frac{1}{2} \frac{d^2}{dx^2} [b(x)h(x)] - \frac{d}{dx} [a(x)h(x)] \quad (6)$$

(WRIGHT 1949; KIMURA 1955). With reference to Figure 1, the two-part equilibrium (1) and (5), which predicts the distribution of A_1 among demes and across the total population, is what would be observed by assigning A_1 's and A_2 's to the lineages and then running the process forward a very long time.

TABLE 1

The nine possible results of type δ and type 2 birth/death events in the history of n lineages currently in n different demes

Rate	Event	Outcome	Magnitude
$U_1 = \lambda_1 n(1 - m)$	δ , no migration	No change	$O(1)$
$U_2 = \lambda_1 nm \frac{D - (n - 1)}{D}$	δ , migration to u-deme	No change	
$U_3 = \lambda_1 nm \frac{n - 1}{D} \frac{1}{N}$	δ , migration to o-deme, coalescent	Coalescent event	$O(1/D)$
$U_4 = \lambda_1 nm \frac{n - 1}{D} \frac{N - 1}{N}$	δ , migration to o-deme, no coalescence	Possible coalescent	
$U_5 = \lambda_1 s_D n(1 - m) \frac{1}{N}$	2, no migration, coalescent	No change (self-collision)	
$U_6 = \lambda_1 s_D n(1 - m) \frac{N - 1}{N}$	2, no migration, no coalescence	Possible branch	
$U_7 = \lambda_1 s_D nm \frac{D - n}{D}$	2, migration to u-deme	Branch event	
$U_8 = \lambda_1 s_D nm \frac{n}{D} \frac{1}{N}$	2, migration to o-deme, coalescent	Collision	$O(1/D^2)$
$U_9 = \lambda_1 s_D nm \frac{n}{D} \frac{N - 1}{N}$	2, migration to o-deme, no coalescence	Possible branch	

U-deme and o-deme mean unoccupied deme and occupied deme, respectively.

The collecting phase: Consider the case in which n lineages are in, or are sampled from, n different demes. By tracing the lineages back in time without knowing their allelic types, we find a simple ancestral process in the limit $D \rightarrow \infty$ and generate an intuitive understanding for the different scaling of mutation *vs.* selection in (4) and in WAKELEY (2003) and WAKELEY and TAKAHASHI (2004). The four kinds of arrows, or reproduction events, discussed above will be encountered by these n lineages and will sometimes move them to occupied demes so that they might coalesce. We can recognize nine possible events for such a sample, and these are listed in Table 1. For example, the fourth kind of event is that a δ -arrow takes one of the n lineages to one of the other $n - 1$ occupied demes but does not connect to the resident ancestral lineage. The result is that now the n lineages, while still distinct, reside in $n - 1$ demes. Thus, at the next event the two lineages that reside together in one deme can coalesce. The interpretations of the other events in Table 1 are equally straightforward. Note that the events are categorized according to their effect on the ancestral lineages and also by their probabilities of occurring with reference to the limit $D \rightarrow \infty$ (while Ds_D remains finite). As in the unstructured ancestral selection graph, when a 2-arrow is encountered, both paths are followed and the lineage splits into two lineages. We are interested in the rates of branching and

coalescence as we follow the history of the sample back in time.

It is appropriate that we have ignored mutations above and in Table 1 because here we are dealing with the unconditional ancestral process. Mutations will occur with probability u_D , $0 \leq u_D \leq 1$, at both types of arrows or reproduction events. Along a single lineage, they occur with rate $\lambda_2 u_D = ND(1 + s_D)u_D[1 + (1 - m)/(Nm)]/2$, which becomes $\theta/2$ in the limit $D \rightarrow \infty$. As in KRONE and NEUHAUSER (1997) and NEUHAUSER and KRONE (1997), we superimpose this mutation process on the unconditional ancestral process. In the conditional ancestral process we consider below, it is necessary to treat mutation, coalescence, and branching simultaneously.

When D is large, the first two events in Table 1 will account for the vast majority of events. These events have rates of $O(1)$, which here means that they have a finite and nonzero limit as $D \rightarrow \infty$ for a given λ_1 . It is easy to see that this is true because the rates of these events depend on D only through λ_1 . These events are lineage switches within and between demes, but ones that preserve the basic sample structure of n lineages in n different demes. They do not change the rates at which events occur and Table 1 continues to apply. Note that with other kinds of (non-island) population structure, these events would change the state of the sample by moving lineages among different types of

demes (WAKELEY and ALIACAR 2001). Here, they can be ignored due to the symmetry of the island model. The next most numerous events will be events 3–7, and these have rates of $O(1/D)$, which means that these rates will approach zero as $D \rightarrow \infty$ for a given λ_1 , but that their rate of approach to zero will be inversely proportional to D . Again, the limiting process we seek is for Ds_D finite as $D \rightarrow \infty$, so that s_D is of $O(1/D)$. Four of these events—3, 4, 6, and 7—do fundamentally alter the state of the sample. These are migration events to occupied demes, both with and without coalescence, and branching events, both with and without a migration event to an unoccupied deme.

KRONE and NEUHAUSER (1997) and NEUHAUSER and KRONE (1997) use the term *collision* to denote the event that an ancestral lineage splits and immediately coalesces with another ancestral lineage. In the unstructured ancestral selection graph, all collisions become negligible in the limit ($N \rightarrow \infty$ and $D = 1$), and it is not necessary to distinguish between different kinds of collisions. Here, because N is finite, some collisions will occur with rates comparable to regular coalescent events. Event 5 in Table 1 is of this sort and has a rate of $O(1/D)$. However, event 5 is a collision in which the lineage splits and then immediately coalesces with itself. Both forward and backward in time, this has no effect. The descendant lineage has the same parent regardless of whether the parental allele is A_1 and the 2-arrow is not followed or the parental allele is A_2 and the 2-arrow is followed. We refer to this event as a self-collision. Only self-collisions have the potential to occur with rates comparable to regular coalescent events. Other kinds of collisions, for example, event 8 in Table 1, which includes some self-collisions, have rates of $O(1/D^2)$. The number of occurrences of all events with rates of this magnitude is shown later, in APPENDIX A, to be negligible in the limiting ($D \rightarrow \infty$) process.

Now, consider what happens in the limiting process. By substituting the value of λ_1 from Equation 3 into the entries of Table 1, the limiting rates become

$$U_1^* = \lim_{D \rightarrow \infty} U_1 = \infty \quad (7)$$

$$U_2^* = \lim_{D \rightarrow \infty} U_2 = \infty \quad (8)$$

$$U_3^* = \lim_{D \rightarrow \infty} U_3 = \binom{n}{2} \frac{Nm + 1 - m}{N} \quad (9)$$

$$U_4^* = \lim_{D \rightarrow \infty} U_4 = \binom{n}{2} \frac{(N-1)(Nm + 1 - m)}{N} \quad (10)$$

$$U_6^* = \lim_{D \rightarrow \infty} U_6 = n \frac{\sigma}{2} \frac{(1-m)(N-1)(Nm + 1 - m)}{N^2 m} \quad (11)$$

$$U_7^* = \lim_{D \rightarrow \infty} U_7 = n \frac{\sigma}{2} \frac{Nm + 1 - m}{N} \quad (12)$$

$$U_8^* = \lim_{D \rightarrow \infty} U_8 = 0 \quad (13)$$

$$U_9^* = \lim_{D \rightarrow \infty} U_9 = 0, \quad (14)$$

where $\binom{n}{2} = n(n-1)/2$ and $\sigma = ND s_D$. Again, the first two types of events do not change the state of the sample, so it is not problematic that their rates become infinite. It simply means that the ancestral lineages will encounter many birth/death events before anything of relevance happens to the sample. Event 3 is a coalescent event in which the number of lineages decreases from n to $n-1$ and these $n-1$ lineages are all in different demes. Event 7 is a branching event, in which case the sample goes from n to $n+1$ lineages, and a migration event leaves all $n+1$ lineages in different demes. In event 4 the number of lineages remains n , but now two of them are in the same deme, while in event 6 the number of lineages increases to $n+1$, and again two are in the same deme. To summarize, the first step that matters in the limiting ($D \rightarrow \infty$) process for n lineages in n different demes can be a coalescent event or a branching event, but in either case the remaining lineages are all in different demes; alternatively it can be a migration event or a branching event that results in two lineages residing together in the same deme while the rest are in separate demes.

We note that Equations 13 and 14, correspond to events that would disrupt the dual process or greatly complicate the branching structure, respectively. The zero rates with which we have described them above refer to their instantaneous rates in the limiting process, but do not fully reveal that even over the entire graph, until the ultimate ancestor is reached, these events will occur with probability zero in the limit as $D \rightarrow \infty$. Proof of this that allows us to omit these events without loss of generality is deferred to APPENDIX A.

With the same level of detail as in Table 1, 27 different events can be distinguished for a sample of n lineages in $n-1$ demes, *i.e.*, where a single pair resides in one deme. However, it is unnecessary to distinguish all of these, and Table 2 shows them grouped into just four kinds of events. The important difference from Table 1 is that events with rates of $O(1)$ now have the potential to affect the state of the sample. The $O(1)$ events in Table 2 are a coalescent event between the two lineages that share a deme and a migration event that sends one of these two lineages to an unoccupied deme. The other events that can change the sample are of $O(1/D)$ or smaller, and these are branching events and migration events to occupied demes. These fast events would be problematic if they were to actually occur in the limiting process, but whenever a pair of lineages resides in the same deme events with rates of $O(1)$ will dominate. These events, 1 and 2 in Table 2, end with all remaining lineages in different demes. This guarantees that there will never be more than one multiply occupied deme and, that when there is, it will contain just two lineages (APPENDIX A). That is, in the limiting process, either a V_1 or a V_2 event always preempts a V_4 event during the state where precisely one deme contains two lineages. Once in the resulting state, Table 1 applies and the highest

TABLE 2

The 27 distinguishable type δ and type 2 birth/death events in the history of n lineages currently in $n - 1$ different demes, grouped into just four kinds of events

Rate	Outcome
$V_1 = \lambda_1 2m + O(1/D)$	Migration (n lineages in n demes)
$V_2 = \lambda_1 2(1 - m)(1/N) + O(1/D)$	Coalescence ($n - 1$ lineages in $n - 1$ demes)
$V_3 = \lambda_1 [(n - 2) + 2(1 - m)(N - 1)/N] + O(1/D)$	No change (n lineages in $n - 1$ demes)
$V_4 = \lambda_1 2s_D + \lambda_1 (n - 2)m(n - 2)/D + O(s_D/D)$	Two or more lineages in at least 1 deme

rate events that affect the sample configuration are of $O(1/D)$.

Thus, in a relatively short time, a sample of n lineages in $n - 1$ demes will be converted to a sample of k lineages in k demes. The value of k depends on whether a coalescent event or a migration event occurs. If it is a coalescent event, then $k = n - 1$, and if it is a migration event, then $k = n$. On the timescale above, with $\lambda_1 = ND[1 + (1 - m)/(Nm)]/2$, these events become instantaneous because their rates approach infinity in the limit $D \rightarrow \infty$. The result is an instantaneous adjustment that has two possible outcomes, with probabilities

$$P_2(\text{mig}) = \lim_{D \rightarrow \infty} \frac{V_1}{V_1 + V_2} = \frac{Nm}{Nm + 1 - m} \quad (15)$$

and

$$P_2(\text{coal}) = \lim_{D \rightarrow \infty} \frac{V_2}{V_1 + V_2} = \frac{1 - m}{Nm + 1 - m}. \quad (16)$$

This will act as a filter on the four-rate Poisson process described by Equations 9–12. Starting with the sample of n lineages in n different demes, whenever a migration event to an occupied deme occurs without immediate coalescence (event 4), there is a chance $P_2(\text{mig})$ that the sample will be returned to its original state. The rest of the time, *i.e.*, with probability $P_2(\text{coal})$, event 4 is converted to event 3, which is a coalescent event. Whenever a branching event occurs where both parents stay in the same deme (event 6), there is a chance $P_2(\text{coal})$ that it is converted into self-collision, so that the sample returns to its original state, and a chance $P_2(\text{mig})$ that it is converted into an observable branching event (event 7).

It is straightforward to show that a Poisson process filtered in this way is equivalent to a Poisson process with an adjusted rate; for example, see WAKELEY (1999). Thus, the probabilities $P_2(\text{mig})$ and $P_2(\text{coal})$ narrow the relevant number of $O(1/D)$ events to just two. Using Equations 9–12, and simplifying, the limiting rates of coalescence and branching become

$$U_3^* + P_2(\text{coal})U_4^* = \binom{n}{2} \quad (17)$$

and

$$U_7^* + P_2(\text{mig})U_6^* = n \frac{\sigma}{2}, \quad (18)$$

respectively. Therefore, the rates of coalescence and branching in the limiting ($D \rightarrow \infty$) ancestral process become identical to the rates of coalescence and branching in a panmictic ancestral selection graph, but where time is measured in units of λ_1 generations. From Equations 3 and 15 we can write $\lambda_1 = ND/(2P_2(\text{mig}))$. Since we assume that $0 < m \leq 1$ and $N \geq 1$, we have $0 < P_2(\text{mig}) \leq 1$ and $\lambda_1 \geq ND/2$. With a migration rate of $m = 1$, the present model reduces to a panmictic model with timescale $\lambda_1 = ND/2$. Otherwise, the effect of island-model subdivision is to lengthen the timescale of the ancestral (and the forward-time) process.

Note that the processes of selection and mutation respond differently to subdivision. Mutation scales with λ_1 , while the selection parameter σ scales simply with $ND/2$. In other words, there are different effective population sizes for selection and for mutation. The effective size for selection is smaller and is equal to $P_2(\text{mig}) = Nm/(Nm + 1 - m)$ times the effective size for mutation. The reason for this is that, when a branching event creates a new lineage in the ancestral process, it has a $P_2(\text{coal}) = (1 - m)/(Nm + 1 - m)$ chance of being erased by a coalescent event, so only $P_2(\text{mig})$ of them are observable. In the present model, this cancels, exactly, the factor $1/P_2(\text{mig})$ by which the number of branching events that occur in the limiting process is increased relative to the number under panmixia. For the same reason, there is also a difference between the scalings for recombination, which splits lineages, and mutation in similarly structured populations (NORDBORG 2000; LESSARD and WAKELEY 2004).

The scattering phase: The results obtained above apply to the history of a sample of n lineages in n different demes. They follow from the fact that coalescent events within multiply occupied demes and migration events from multiply occupied demes to unoccupied demes occur with rates that are D times larger than the rates of branching events and migration events to occupied demes. Because of this, whenever multiple samples are taken from one or more demes, there will be a brief scattering phase and this will leave the remaining lineages in different demes (*cf.* Table 2) so that the collect-

ing-phase results above apply. The number of lineages that remain at the end of the scattering phase, and that then enter the collecting phase, will depend on the number of coalescent events that occur within the multiply sampled deme or demes. In the limit as $D \rightarrow \infty$, the duration of the scattering phase becomes negligible so that it can be treated as instantaneous when time is measured in proportion to D generations.

Because we have assumed that s_D is $O(1/D)$, whereas the rates of within-deme coalescence and migration to occupied demes depend only on N , m , and the sample size(s), the scattering phase here is the same as that in a population in which all genetic variation is selectively equivalent. As in WAKELEY (1998), the scattering phase for a sample of size n all from a single deme is equivalent to a series of n Bernoulli trials with probabilities of success

$$P_k(\text{mig}) = \frac{\mathcal{M}}{\mathcal{M} + k - 1}, \quad (19)$$

for $k = 1, 2, \dots, n$, and in which $\mathcal{M} = Nm/(1 - m)$. Note that $k = 2$ yields $P_2(\text{mig})$ given in Equation 15. Success is defined to be the migration of one of the lineages to an unoccupied deme, and each of these increases by one the number of lineages that will enter the collecting phase. If we use n' to denote the number of lineages that will enter the collecting phase, then

$$P[n'|n] = \frac{|S_n^{n'}| \mathcal{M}^{n'}}{\mathcal{M}_{(n)}} \quad (20)$$

is its probability function (WAKELEY 1998), in which the $|S_n^{n'}|$ are unsigned Stirling numbers of the first kind and $\mathcal{M}_{(n)} = \mathcal{M}(\mathcal{M} + 1) \cdots (\mathcal{M} + n - 1)$. Finally, the scattering phase occurs independently among demes, so that we can multiply the probabilities (20) over samples from different demes.

METHODOLOGY

We have shown that the ancestry of a sample from a subdivided population, with equal deme sizes and island-type migration, and in which two alleles are subject to selection and mutation, includes an instantaneous scattering phase and then enters a collecting phase, which is equivalent to the unstructured ASG with mutation and selection parameters scaled appropriately. This means that if Y is any measurement on the sample, such as the time to common ancestry or the total length of the genealogy, we can compute properties of Y by conditioning on the outcome of the scattering phase. For example, if $E[Y|\mathbf{n}]$ is the expected value of Y for the sample $\mathbf{n} = (n_1, \dots, n_d)$ of size $|\mathbf{n}| = \sum_{i=1}^d n_i$ from d different demes, then

$$E[Y|\mathbf{n}] = \sum_{\mathbf{n}'} E[Y|\mathbf{n}'] P[\mathbf{n}'|\mathbf{n}], \quad (21)$$

where $P[\mathbf{n}'|\mathbf{n}]$ is the product of probabilities like (20) over demes. Because the scattering phase always ends

with each remaining lineage in a separate deme, $E[Y|\mathbf{n}']$ is the expected value of Y for a sample of size $|\mathbf{n}'|$ in our rescaled, unstructured, collecting-phase ASG. Below, we use the framework of Equation 21 both analytically and in simulations, where our results are also consistent with keeping track of allelic configurations during the scattering phase.

A compression of the Markov process that describes the conditional evolution of the unstructured ASG is found. This continuous-time Markov jump chain is an extension of the conditional ancestral selection graph of SLADE (2000a) and is derived with new clarity. The result is a maximal compression of the ancestral selection graph in which the timing properties of the graph are retained. This continues, and improves upon, a similar enhancement of the conditional graph by SLADE (2000b) that reduces its branching rate. This is an alternative version of the minimal representation of the conditional graph as discussed in SLADE and NEUHAUSER (2003). Without the timing structure of the graph in place, results that describe some probabilistic properties of a single ancestral lineage within the ASG are obtained by FEARNHEAD (2002). The details of our general derivation are deferred to APPENDIX B.

The simulations presented in RESULTS are performed using the corresponding adaptation of a computational Monte Carlo method as in SLADE (2000a,b). The purpose of such a simulation scheme is the calculation of an approximate probability distribution over the realizations of the genealogy. For a particular genealogy then, a weight is attributed to its associated time to the most recent common ancestor (T_{MRCA}), and the weighted average is evaluated upon completion of a large number of repetitions. The improved efficiency in having attained the maximal compression of the genealogical Markov chain is substantial; running time until convergence that was previously required is now at least fully halved. The feasibility of at least doubling the branching and mutation rates is also achieved. The performance gets better still when calculating statistics of the T_{MRCA} distribution, as it would also do for properties of subtrees of the genealogy. Our compression reduces the number of possible transitions required to describe the conditional ASG at any event, from 10 to only 6. Thus, the size of the state space of a realization of the conditional graph decreases exponentially. Hence, variance of any corresponding simulation technique is reduced. Depending on the parameter combinations, to produce the composite PDFs in RESULTS with a Pentium 3.06GHz Xeon processor, computing time ranged from a few hours to several days.

RESULTS

From Equation 21 we can see that simple estimators of the migration parameter \mathcal{M} are as useful when weak selection is present as they are when all variation is neutral. For example, let q_0 and q_1 be the probabilities

of identity in state for samples of size two taken from the same deme and from two different demes, respectively. Taking the scattering phase into account, we have

$$1 - q_0 = \frac{\mathcal{M}}{\mathcal{M} + 1} (1 - q_1), \quad (22)$$

and if \hat{q}_0 and \hat{q}_1 are estimates of q_0 and q_1 from some data, then

$$\hat{\mathcal{M}} = \left[\frac{1 - \hat{q}_1}{1 - \hat{q}_0} - 1 \right]^{-1} \quad (23)$$

is an estimate of \mathcal{M} . Note that (23) has the same form as the various estimators of gene flow based on F_{ST} or pairwise sequence differences under the assumption of selective neutrality; see, for example, HUDSON *et al.* (1992).

Neither the unstructured ASG nor our slightly more complicated structured ASG lends itself to analytical computations much more involved than the above. From Equation 21 it is clear how analysis of our structured ASG is related to analysis of the unstructured graph. So, for example, it is possible to average such expressions as KRONE and NEUHAUSER'S (1997) Equation 3.5 for the expected time to the ultimate ancestor of the sample over $n = |\mathbf{n}'|$ at the start of the collecting phase. Instead of this, we have used simulations that include the scattering phase and the maximal compression of the ASG described in APPENDIX B to compute the probability density function of the time to the most recent common ancestor of a sample (T_{MRCA}).

In unstructured nonneutral genealogies with small mutation rates certain sample configurations, among a sample of size n , have reduced mean coalescent times when compared to their neutral counterparts, and other sample configurations have enlarged coalescent times (SLADE 2000a; see also SLADE and NEUHAUSER 2003). The mixtures of reduced sample configurations at the start of the collecting phase in the structured model result in novel distributions of the T_{MRCA} under selection. This mixture is the same under neutrality as it is with selection. The effect of the scattering phase under neutrality is governed by the level of mutation in the collecting phase. Although the sample size is reduced, it can easily correspond to an increased T_{MRCA} because the adjusted sample configuration requires longer to assimilate its ancestors (particularly for low mutation rates).

In the unstructured ASG as the mutation rate decreases the conditional genealogy becomes more sensitive to the presence of selection (SLADE 2000a). However, the sample configurations that yield the greatest selective effect are also the samples least likely to be obtained, having extremely small sampling probabilities. The sampling distribution is increasingly U-shaped as θ decreases, and under selection, samples dominated by the favored allele are more likely than those dominated by the unfavored allele. Introducing population subdivision makes the most sensitive samples accessible because the scattering phase delivers adjusted samples to the

collecting phase. Low migration rates yield sample configurations at the start of the collecting phase that consist of close to $d A_1$ alleles and $d A_2$ alleles, provided all of the d demes initially sampled contained at least one of each allele. The highest sensitivity to selection therefore is achieved when the population is structured between very many demes connected by an exceedingly low migration rate. Note that this effect is independent of whether the initial sample contains mostly the favored or the unfavored allele.

To convert a coalescent time unit into years, it is necessary to calculate the product of the (estimated) effective population size and number of years per reproductive generation. When the migration rate $\mathcal{M} \ll 1$ we compare the structured model only since according to the factor calculated in THEORY, $1 + \mathcal{M}^{-1}$, a coalescent time unit in the unstructured ASG represents incomparably fewer numbers of generations. On the other hand, when $\mathcal{M} \gg 1$ there is a meaningful comparison between results with and without very many demes. When $\mathcal{M} = 0.01$ in the structured genealogies (both neutral and selected) every generation that elapses according to coalescent time translates as 100 additional generations in an unstructured genealogy. Population subdivision magnifies the extent of changes to coalescence times and thus immediately accentuates the effect of selection on genealogical time depth.

The interaction between migration and selection as it affects the distribution of the conditional T_{MRCA} is shown in Figures 2 and 3. The mutation rate is set at three levels, $\theta = 0.01, 0.1,$ and 1 . Low mutation rates are appropriate since we consider alleles corresponding to amino acid replacements at a hypothetical nucleotide site within a single locus. (We note that $\theta = 0.001$ yields a distribution of the T_{MRCA} under neutrality that is extremely similar to that obtained for $\theta = 0.01$.) The selection rate in our three cases is chosen in accordance with the weighted average predicted to be operating at nucleotide sites by SAWYER *et al.* (2003), namely $\sigma = 5$ and 7.5 . The migration rate is set to two levels, $\mathcal{M} = 0.01$ in Figure 2 and $\mathcal{M} = 10$ in Figure 3.

The probability density function (PDF) for the low migration rate and low mutation rate case is presented in Figure 2A. A substantial change in the shape of the PDF is clear under selection. Increasing the mutation rate, the next lowest migration rate case is shown Figure 2B. The extent of the effect of selection appears to diminish at our highest level of the mutation rate in Figure 2C. The parameters (mean and standard deviation) of these distributions are compared in Table 3. In the last case just mentioned, we have a particularly novel result as the mean starts to increase again, *i.e.*, toward neutrality, with the selection strength above some threshold. In both Figure 2 and Table 3 it is assumed that samples of size 12 were taken from each of four demes, and that each of the four samples included one copy of allele A_1 and 11 copies of allele A_2 , or (1, 11) for short.

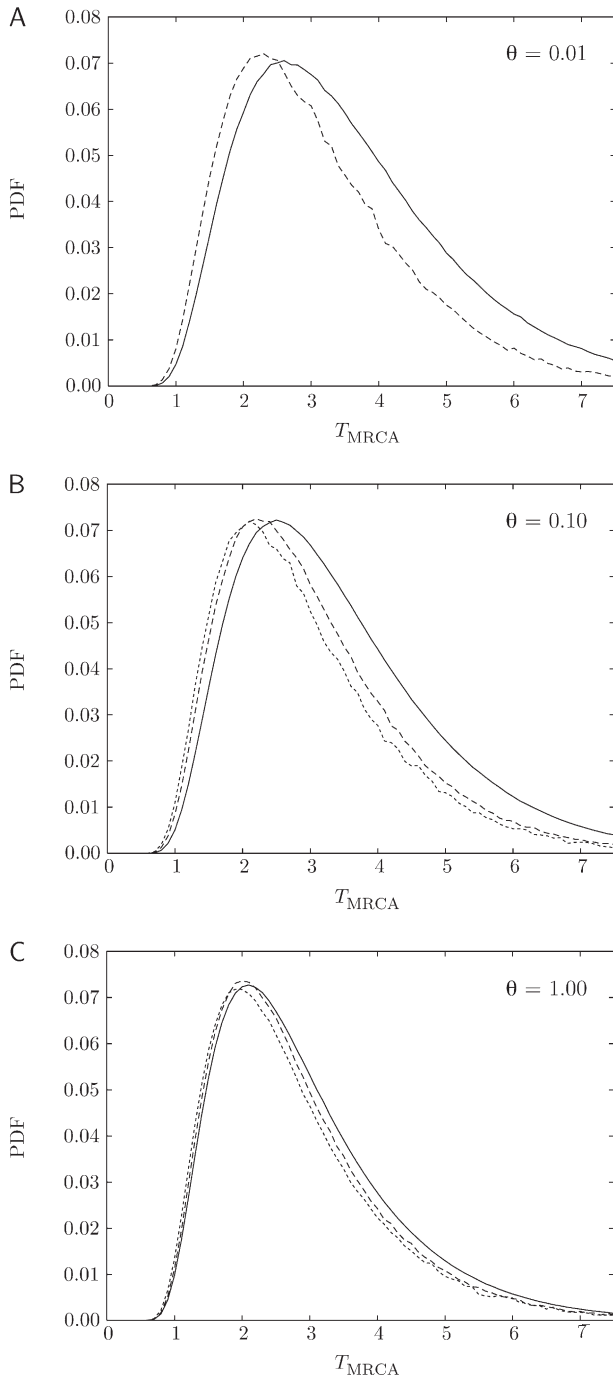


FIGURE 2.—PDFs for the T_{MRCA} of a sample of size 48 spread evenly among four demes and with sample allele counts (1 A_1 , 11 A_2). A–C depict results for three different mutation rates as indicated. The migration parameter was $M = 0.01$, and the solid curve shows the PDF for $\sigma = 0$; the long-dashed curves, for $\sigma = 5.0$; and the short-dashed curves, for $\sigma = 7.5$. The case of $\theta = 0.01$ and $\sigma = 7.5$ could not be completed due to prohibitive computational requirements.

In contrast, when the migration rate is high the sample size adjustment is moderate, and the initial sample configurations show very similar responses to selection. The PDFs shown in Figure 3, A and B, correspond to different initial samples within each deme of being either all (1, 11) of A_1 and A_2 alleles or vice versa all

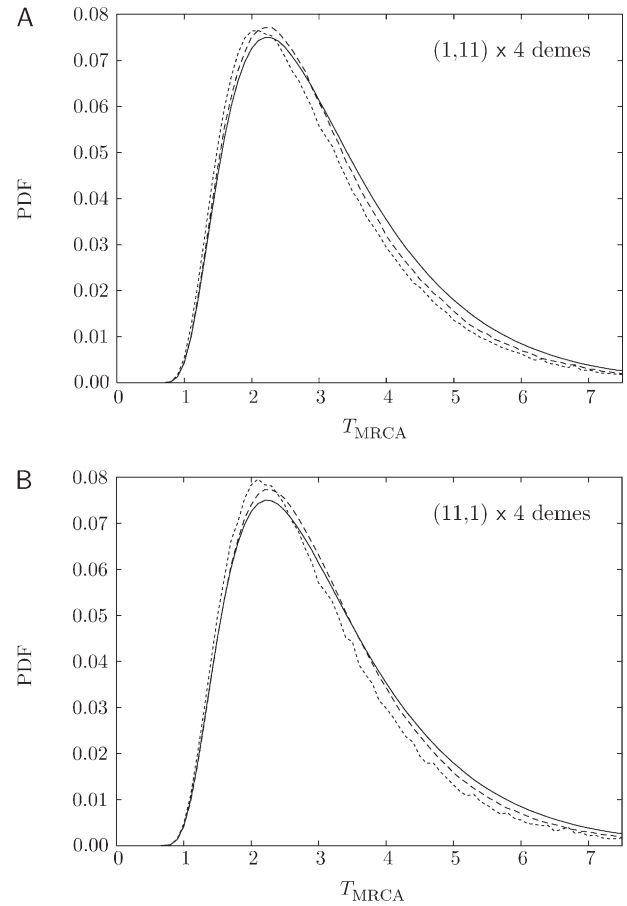


FIGURE 3.—PDFs for the T_{MRCA} of a sample of size 48 spread evenly among four demes, but comparing two sample distributions of alleles A_1 and A_2 : (1, 11) as in Figure 2 and (11, 1). In all cases, $M = 10$ and $\theta = 0.1$. As in Figure 2, the solid curve shows the PDF for $\sigma = 0$; the long-dashed curves, for $\sigma = 5.0$; and the short-dashed curves, for $\sigma = 7.5$.

(11, 1), respectively. The effect on the shape of the PDFs can be discerned; however, in Table 4 comparisons of the parameters of the distributions under panmixia are also shown. With a migration rate as high as $M = 10$, the effects of many-demes population structure in Table 4, for both neutral and selected genealogies, can still be distinguished. In the neutral case (Table 4, left column), even with the sample size adjustment the mean T_{MRCA} is considerably higher than it is for panmixia. Comparing the two rows of Table 4, we can see that this also reflects the 10% lengthening of the genealogies expected when $M = 10$. The difference in the selective effect between the two completely different initial samples is only minor.

DISCUSSION

We have obtained a simple structured ancestral selection graph in the limit as the number of demes tends to infinity and with island-type migration among demes of equal size. The result is understood as an approximation to the ancestral process for samples in the presence

TABLE 3

The means and the standard deviations (in parentheses) of the PDFs depicted in Figure 2

θ	σ		
	0.0	5.0	7.5
0.01	2.425 (0.361)	2.023 (0.494)	—
0.10	2.291 (0.376)	1.998 (0.517)	1.878 (0.414)
1.00	1.888 (0.722)	1.618 (0.804)	1.744 (0.730)

of weak selection from populations composed of a large number of demes. Convergence to the simpler process occurs due to the difference in timescales between migration and drift within demes on the one hand and selection and mutation across the entire population on the other. Under neutrality, results of this sort can be proven using the technique developed by MÖHLE (1998a,b) for calculating the transition rates of (ancestral) Markov processes with a separation of timescales. For ancestries under weak selection, the framework of birth-death processes is more appropriate because the addition of new lineages at type 2 branching events increases the dimensionality that would be required in Möhle's analysis. We have shown that one can truncate the system by ignoring the transitions of very small rates to obtain a closed approximated system; we then found the correct transition rates in the limiting process were provided by the corresponding calculation.

Our analysis does not provide an estimate of the error incurred in applying the limiting model to real (*i.e.*, finite) populations. One way to address this is using simulations, such as those reported in LESSARD and WAKELEY (2004). In that article, simulations demonstrated the rapid convergence of the distribution of the total length of the gene genealogy of a sample of size two to that predicted by the limiting ($D \rightarrow \infty$) ancestral recombination graph with island model migration among demes. Populations composed of 100 or more demes were closely approximated by the limiting model.

It is clear that any quantity of interest can be computed using the framework of Equation 21, that is, by conditioning on the outcome of the scattering phase. In fact, in the case of weak selection and mutation considered here, there is a lot of parity with the strictly neutral

model (WAKELEY 1998). In other words, migration structures selected variation within and between demes in a way that is identical to the way it structures neutral variation. This can be seen as a consequence of the fact that the model applies to $0 < m \leq 1$ while $s_D \rightarrow 0$ as $D \rightarrow \infty$, so that migration is stronger than selection within demes. However, using the improved algorithm for simulation of selected genealogies conditional on the frequencies of A_1 and A_2 in the sample, we have also shown a strong effect of migration on the T_{MRCA} of some samples. Namely, limited migration can convert even very unbalanced samples into balanced ones, *e.g.*, $(1, 11) \rightarrow (1, 1)$, and place the collecting-phase ancestral sample into the range of samples for which the T_{MRCA} is reduced by selection. This happens because there are no mutations during the scattering phase, while coalescent events between like alleles can occur. However, we note that although unbalanced samples, like $(1, 11)$, within demes are more likely to be obtained than balanced samples, like $(6, 6)$, when the migration rate is small it will be even more typical for single-deme samples to be fixed for one allele or the other. This can be seen in Equation 1, which converges to $\nu_0 = 1 - x$ and $\nu_N = x$ in the limit $m \rightarrow 0$.

We note also that if we had not assumed the mutation rate to be small ($u_D \rightarrow 0$ as $D \rightarrow \infty$), we would have obtained a neutral limiting graph. When the mutation rate in the ancestral selection graph is infinite the branching-coalescing structure collapses to a neutral coalescent process (NEUHAUSER and KRONE 1997). Allowing mutation in the scattering phase of the structured ancestral selection graph also yields a neutral (coalescent) ancestral process, since this would render an infinite mutation rate in the collecting phase. That is, if the mutation probability μ is $O(1)$ instead of $O(1/D)$ then mutation is a fast timescale event and will occur in the scattering phase amid the rapid migration and coalescent events also of $O(1)$. Whenever the process switches over to the slower transitions, such as those in Table 1 that are $O(1/D)$, mutation still occurs at a fast rate and $\theta \rightarrow \infty$, as $D \rightarrow \infty$.

We would like to clarify the case of zero mutation rate that leads to a neutral coalescent in the ASG. When $\theta = 0$, ancestors are always of one type only, but in the dual process 2-arrows occur at a faster rate than δ -arrows. If all ancestors are A_1 only the slower-rate δ -arrows are

TABLE 4

Comparison of the means and the standard deviations (in parentheses) of the T_{MRCA} in the unstructured ASG to those of the PDFs depicted in Figure 3 rescaled by the factor $1 + M^{-1}$ so that they are measured in the same units

	$(1, 11) \times 4$ demes			$(11, 1) \times 4$ demes	
	$\sigma = 0$	$\sigma = 5$	$\sigma = 7.5$	$\sigma = 5$	$\sigma = 7.5$
$M = 10$	2.318 (0.432)	2.195 (0.408)	2.173 (0.377)	2.271 (0.889)	2.165 (0.752)
$M \rightarrow \infty$	2.060 (0.368)	1.986 (0.350)	1.972 (0.330)	2.008 (0.837)	1.918 (0.738)

needed to describe the ancestral process since 2-arrows are never followed, and this yields the neutral coalescent. On the other hand, having all A_2 ancestors both the δ -arrows and the faster-rate 2-arrows participate in the ancestral process. However, the outcome of 2-arrows is determined in this case; they will always be followed and no uncertainty arises in the path taken by the dual process of a sample. No virtual ancestors are needed and when a branching event occurs in the Markov process corresponding to the conditional ASG, it represents a so-called null transition, as discussed in APPENDIX B. Recalculating the Markov process without these transitions yields the same neutral coalescent as required.

We thank both referees for their reports. P.F.S. thanks the Wakeley Lab and the Department of Organismic and Evolutionary Biology at Harvard for their hospitality during a visit. J.W. was supported by a Presidential Early Career Award for Scientists and Engineers from the National Science Foundation (DEB-0133760). P.F.S. was also supported by a travel fund from the Faculty of Science at the University of Sydney. This is research publication no. 0009 from Sydney University Bioinformatics and Technology Centre.

LITERATURE CITED

- ALDOUS, D. J., 1985 Exchangeability and related topics, pp. 1–198 in *École d'Été de Probabilités de Saint-Flour XII–1983* (Lecture Notes in Mathematics, Vol. 1117), edited by A. DOLD and B. ECKMANN. Springer-Verlag, Berlin.
- AVISE, J. C., 2000 *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.
- BAHLO, M., and R. C. GRIFFITHS, 2000 Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* **57**: 79–95.
- BARTON, N. H., A. M. ETHERIDGE and A. K. STURM, 2004 Coalescence in a random background. *Ann. Appl. Probab.* **14**: 754–785.
- BEERLI, P., and J. FELSENSTEIN, 1999 Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**: 763–773.
- BEERLI, P., and J. FELSENSTEIN, 2001 Maximum-likelihood estimation of a migration matrix and effective population sizes in n populations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA* **98**: 4563–4568.
- BUSTAMANTE, C. D., R. NIELSEN, S. A. SAWYER, K. M. OLSEN, M. D. PURUGGANAN *et al.*, 2002 The cost of inbreeding in *Arabidopsis*. *Nature* **416**: 531–534.
- DE IORIO, M., and R. C. GRIFFITHS, 2004 Importance sampling on coalescent histories. ii. subdivided population models. *Adv. Appl. Probab.* **36**: 434–454.
- ETHIER, S. N., and T. NAGYLAKI, 1980 Diffusion approximations of Markov chains with two timescales and applications to population genetics. *Adv. Appl. Probab.* **12**: 14–49.
- FEARNHEAD, P., 2002 The common ancestor at a nonneutral locus. *J. Appl. Probab.* **39**: 38–54.
- GOLDING, B., 1994 *Non-Neutral Evolution*. Chapman & Hall, New York.
- GRIFFITHS, R. C., 1991 The two-locus ancestral graph, pp. 100–117 in *Selected Proceedings of the Symposium on Applied Probability*, edited by I. V. BASAWA and R. L. TAYLOR. Institute of Mathematical Statistics, Hayward, CA.
- HEY, J., and C. A. MACHADO, 2003 The study of structured populations—new hope for a difficult and divided science. *Nat. Rev. Genet.* **4**: 535–543.
- HUDSON, R. R., 1983 Testing the constant-rate neutral allele model with protein sequence data. *Evolution* **37**: 203–217.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992 Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.
- KAPLAN, N. L., T. DARDEN and R. R. HUDSON, 1988 Coalescent process in models with selection. *Genetics* **120**: 819–829.
- KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The “hitchhiking effect” revisited. *Genetics* **123**: 887–899.
- KARLIN, S., and H. M. TAYLOR, 1975 *A First Course in Stochastic Processes*. Academic Press, New York.
- KIMURA, M., 1955 Stochastic processes and the distribution of gene frequencies under natural selection. *Cold Spring Harbor Symp. Quant. Biol.* **20**: 33–53.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA.
- KINGMAN, J. F. C., 1982a The coalescent. *Stoch. Proc. Appl.* **13**: 235–248.
- KINGMAN, J. F. C., 1982b Exchangeability and the evolution of large populations, pp. 97–112 in *Exchangeability in Probability and Statistics*, edited by G. KOCH and F. SPIZZICHINO. North-Holland, Amsterdam.
- KRONE, S. M., and C. NEUHAUSER, 1997 Ancestral processes with selection. *Theor. Popul. Biol.* **51**: 210–237.
- LESSARD, S., and J. WAKELEY, 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.
- MÖHLE, M., 1998a Coalescent results for two-sex population models. *Adv. Appl. Probab.* **30**: 513–520.
- MÖHLE, M., 1998b A convergence theorem for Markov chains arising in population genetics and the coalescent with partial selfing. *Adv. Appl. Probab.* **30**: 493–512.
- MÖHLE, M., 1998c Robustness results for the coalescent. *J. Appl. Probab.* **35**: 438–447.
- MORAN, P. A. P., 1958 Random processes in genetics. *Proc. Camb. Philos. Soc.* **54**: 60–71.
- MORAN, P. A. P., 1962 *Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford.
- NATH, H. B., and R. C. GRIFFITHS, 1996 Estimation in an island model using simulation. *Theor. Popul. Biol.* **50**: 227–253.
- NEUHAUSER, C., 1999 The ancestral graph and gene genealogy under frequency-dependent selection. *Theor. Popul. Biol.* **56**: 203–214.
- NEUHAUSER, C., and S. M. KRONE, 1997 The genealogy of samples in models with selection. *Genetics* **145**: 519–534.
- NORDBORG, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial selfing. *Genetics* **154**: 923–929.
- NOTOHARA, M., 1990 The coalescent and the genealogical process in geographically structured population. *J. Math. Biol.* **29**: 59–75.
- OHTA, T., and M. KIMURA, 1971 On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**: 18–25.
- SAWYER, S. A., and D. L. HARTL, 1992 Population genetics of polymorphism and divergence. *Genetics* **132**: 1161–1176.
- SAWYER, S. A., R. J. KULATHINAL, C. D. BUSTAMANTE and D. L. HARTL, 2003 Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J. Mol. Evol.* **57**: S154–S164.
- SLADE, P. F., 2000a Simulation of selected genealogies. *Theor. Popul. Biol.* **57**: 35–49.
- SLADE, P. F., 2000b Most recent common ancestor probability distributions in gene genealogies under selection. *Theor. Popul. Biol.* **58**: 291–305.
- SLADE, P. F., and C. NEUHAUSER, 2003 Nonneutral genealogical structure and algorithmic enhancements of the ancestral selection graph. *Theor. Biol.* **8**: 255–278 (erratum: *Theor. Biol.* **8**: 539).
- SLATKIN, M., 1985 Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* **16**: 393–430.
- SLATKIN, M., 1987 The average number of sites separating DNA sequences drawn from a subdivided population. *Theor. Popul. Biol.* **32**: 42–49.
- STEPHENS, M., 2001 Inferences under the coalescent, pp. 213–238 in *Handbook of Statistical Genetics*, edited by D. J. BALDING, M. J. BISHOP and C. CANNINGS. John Wiley & Sons, Chichester, England.
- STEPHENS, M., and P. DONNELLY, 2003 Ancestral inference in population genetics models with selection. *Aust. N. Z. J. Stat.* **45**: 395–430.
- STROBECK, C., 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* **117**: 149–153.
- TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- TAVARÉ, S., 2004 Ancestral inference in population genetics, pp. 1–188 in *École d'Été de Probabilités de Saint-Flour XXXI–2001* (Lec-

ture Notes in Mathematics, Vol. 1837), edited by O. CANTONI, S. TAVARÉ and O. ZEITOUNI. Springer-Verlag, Berlin.
 WAKELEY, J., 1998 Segregating sites in Wright's island model. *Theor. Popul. Biol.* **53**: 166–175.
 WAKELEY, J., 1999 Non-equilibrium migration in human history. *Genetics* **153**: 1863–1871.
 WAKELEY, J., 2003 Polymorphism and divergence for island model species. *Genetics* **163**: 411–420.
 WAKELEY, J., and N. ALLIACAR, 2001 Gene genealogies in a meta-population. *Genetics* **159**: 893–905 (corrigendum: *Genetics* **160**: 1263–1264).
 WAKELEY, J., and T. TAKAHASHI, 2004 The many-demes limit for

selection and drift in a subdivided population. *Theor. Popul. Biol.* **66**: 83–91.
 WILKINSON-HERBOTS, H. M., 1998 Genealogy and subpopulation differentiation under various models of population structure. *J. Math. Biol.* **37**: 535–585.
 WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.
 WRIGHT, S., 1949 Adaptation and selection, pp. 365–389 in *Genetics, Paleontology and Evolution*, edited by G. L. JEPSON, G. G. SIMPSON and E. MAYR. Princeton University Press, Princeton, NJ.

Communicating editor: D. CHARLESWORTH

APPENDIX A

We first consider a process that omits U_8 , U_9 , and V_4 ; a branching-coalescing graph results similar to the one described in THEORY but without any problematic events. We show that in our original process, each of these three events will occur with probability zero, by the T_{UA} , in the limit as $D \rightarrow \infty$.

Analogously to that of KRONE and NEUHAUSER (1997), for the structured ancestral selection graph with many demes collisions and such can be ignored. Due to the similarity of the techniques employed to this end we present an outline of the proof only.

Couple the original limiting ancestral process with another limiting ancestral process that has a higher branching rate $\sigma > \sigma/2$. This produces an exact copy of all events in the original process, except for the addition of branching events that result from the higher rate.

Let M_n denote the maximum number of ancestors that can be present in the graph. Equation 2.4 in Theorem 3 of GRIFFITHS (1991) allows the following calculation, the details of which are apparently new,

$$P(M_n > \lfloor \varepsilon n \rfloor) = \frac{\sum_{j=0}^{n-2} \sigma^{-j} j!}{\sum_{j=0}^{\lfloor \varepsilon n \rfloor - 1} \sigma^{-j} j!} \leq \begin{cases} \frac{(n-1)!}{(\lfloor \varepsilon n \rfloor - 1)!} \sigma^{\lfloor \varepsilon n \rfloor - n + 1} & \text{if } 0 < \sigma \leq 1 \\ \frac{(n-1)!}{(\lfloor \varepsilon n \rfloor - 1)!} \sigma^{\lfloor \varepsilon n \rfloor - 1} & \text{if } \sigma > 1, \end{cases} \tag{A1}$$

where $\lfloor x \rfloor$ is the integer part of x and $\varepsilon > 1$ is any constant. The probability above is largest when $\sigma > 1$, and we need only consider that case in what follows, without loss of generality. Applying Stirling's formula, we then have the following upper bound on the probability in question,

$$C \sigma^{\lfloor n \varepsilon \rfloor} \frac{(n-1)^{n-1/2}}{(\lfloor \varepsilon n \rfloor - 1)^{\lfloor \varepsilon n \rfloor - 1/2}} e^{\lfloor \varepsilon n \rfloor - n} \leq \tilde{C} \sigma^{\lfloor \varepsilon n \rfloor} e^{\lfloor \varepsilon n \rfloor - n} n^{-\lfloor \varepsilon n \rfloor + n} \rightarrow 0, \tag{A2}$$

as $n \rightarrow \infty$. One may then deduce the expression given by GRIFFITHS (1991) that $M_n/n \rightarrow 1$, in probability, as $n \rightarrow \infty$. Furthermore, we know that throughout the graph M_n is bounded above by some constant K_n dependent on n , but independent of ND and time.

Theorem 4.7.1 in KARLIN and TAYLOR (1975) yields an upper bound on the expected total time until the ultimate ancestor (T_{UA}). That is, some constant K_n , independent of time and ND , exists such that $E(T_{UA}) < K_n$. The combination of events $\{U_8, U_9\}$ occurs with rate $\lambda_1 s_D n m (n/D)$ in the original ancestral process. Fix a constant $0 < \alpha < 1/5$. At most $(ND)^\alpha$ units of time elapse before the ultimate ancestor is found. The number of ancestors in the graph is perpetually bounded above by $(ND)^\alpha$ prior to finding the ultimate ancestor. Therefore, in the dominating coupled limiting process, the probability of $\{U_8, U_9\}$ occurring prior to the ultimate ancestor being reached cannot be greater than

$$(ND)^\alpha \sigma \left[1 + \frac{1-m}{Nm} \right] (ND)^\alpha \frac{m}{D} (ND)^\alpha = m \sigma N^{3\alpha} D^{-(1-3\alpha)} \left[1 + \frac{1-m}{Nm} \right], \tag{A3}$$

the right side of which tends to zero as $D \rightarrow \infty$, since $\alpha < 1/5$.

The other event that could spoil the integrity of the ancestral process with an instantaneous proliferation of ancestors is V_4 . Events U_4 and U_6 in Table 1 lead to the collection of events in Table 2, $\{V_1, V_2, V_3, V_4\}$. Note that since the former occur with a rate $O(1/D)$, instead of $O(1)$, there is a bound on the total number of entries into the set of V events that does not depend on D . The event $\{(U_4 \cup U_6) \cap V_4\}$ occurs with rate

$$\lambda_1 \left(mn \frac{n-1}{D} \frac{N-1}{N} + (1-m) s_D n \frac{N-1}{N} \right) \times \frac{V_4}{V_1 + V_2 + V_4}. \tag{A4}$$

As in the previous paragraph, in the limiting process up until the ultimate ancestor is found, the probability that an event such as this ever occurs cannot be greater than the following:

$$(ND)^\alpha (m(ND)^\alpha ((ND)^\alpha - 1)(N-1)/2 + (1-m)\sigma(ND)^\alpha (N-1)/N) \times \left[1 + \frac{1-m}{Nm} \right] \frac{2s_D + ((ND)^\alpha - 2)^2 (m/D)}{2m + 2(1-m)(1/N) + 2s_D + ((ND)^\alpha - 2)^2 (m/D)}. \quad (\text{A5})$$

One may verify that the expression above does indeed tend to zero as $D \rightarrow \infty$, since $\alpha < 1/5$.

For a finite N there is no need to restrict the migration probability m for our zero limit results to hold. For the large- N limits, we assume a very small m such that $Nm \rightarrow M$ as $N \rightarrow \infty$. Let us also replace s_D by $s_{ND} = \sigma/(ND)$ in (A5). Three limits are separated into two types for us to describe. A limit with respect to *both* N and D that converges to zero can be obtained by setting $N = D^\beta$, where $\beta < (1 - 5\alpha)/(5\alpha)$. When passing both to the limit, we could instead allow $N = D$ and fix $\alpha < 1/10$. After the substitutions just referred to, the limit is then taken with respect to the single variable represented in the equation, D . There are two *iterated limits* where we first pass to the limit with respect to D and then with respect to N , and vice versa. In the former we easily have convergence to zero, as required. In the latter, D is held constant while N passes to infinity and we find that (A5) would diverge. This is a consequence of our notation and we address this problem below.

We have already proven a bound on the maximum number of ancestors and on the $E(T_{UA})$ and that this bound depends on sample size but is independent of population size ND and time. [See (A2) and Karlin and Taylor's theorem.] We have expressed this bound relative to N and D by writing $(ND)^\alpha$. Thus when D is passed to the limit first (or together with N) this all works out according to (A3) and (A5), and the limits do indeed converge to zero. That one of the iterated limits corresponding to (A5) diverges, namely as $N \rightarrow \infty$ before D , is technically consistent but violates the earlier proven bounds. This failure of notation indicates that the technique cannot be applied to the iterated limit of N before D . In that case an expression of the form $(K_n)^5/D$ describes its growth as N and then D are passed to the limit. Thus, we have now ensured eventual convergence to zero of all limits. We have shown that, over the entire graph until the T_{UA} , the collision probabilities are bounded above by a term $O(1/D)$. These probabilities then converge to zero in all cases as $D \rightarrow \infty$.

APPENDIX B

A realization of the conditional graph can include the exact ancestral relationships that describe the ancestors involved at each event and the waiting times between events. An arbitrary but fixed order is required among the ancestors to identify which particular ancestors participate in any event. Combinatorial factors used subsequently to remove this order among the ancestors are valid provided that the transition probabilities and boundary conditions are translated accordingly (SLADE 2000a). Our formulation here uses an unordered allelic configuration and an ordered decomposition into real and virtual ancestors.

As the conditional graph evolves backward in time, given an initial sample configuration, branching events add ancestors to the graph, but these are not actual ancestors of the sample. These *virtual* ancestors are contained within the graph for the probabilistic purpose of representing selective evolution and do not belong to the genealogy; the remaining actual ancestors are *real* ancestors of the genealogy. Define the pair $\mathbf{r} = (r_1, r_2)$ as the number of real ancestors of types A_1 and A_2 , respectively. Define also the pair $\mathbf{v} = (v_1, v_2)$ as the number of virtual ancestors of each allelic type. Then, $r_1 + v_1 = n_1$ and $r_2 + v_2 = n_2$. Let \mathbf{e}_i be the i th unit vector.

The continuous-time Markov jump chain identified in SLADE (2000b) makes transitions as

$$(\mathbf{r}, \mathbf{v}) \rightarrow \begin{cases} (\mathbf{r} - \mathbf{e}_i, \mathbf{v}) & \text{with rate } r_i(r_i - 1)/n_i, \\ (\mathbf{r}, \mathbf{v} - \mathbf{e}_i) & \text{with rate } v_i(v_i - 1 + 2r_i)/n_i, \\ (\mathbf{r} + \mathbf{e}_i - \mathbf{e}_j, \mathbf{v}) & \text{with rate } \theta \frac{n_i + 1}{n} r_j/n_j, \\ (\mathbf{r}, \mathbf{v} + \mathbf{e}_i - \mathbf{e}_j) & \text{with rate } \theta \frac{n_i + 1}{n} v_j/n_j, \\ (\mathbf{r}, \mathbf{v} + \mathbf{e}_i) & \text{with rate } \sigma(n_1 + 1)/(n + 1), \\ (\mathbf{r}, \mathbf{v}) & \text{with rate } \sigma n_2/n, \end{cases} \quad (\text{B1})$$

where $n = n_1 + n_2$, we have factored out $n/2$ from the above rates, and $i, j = 1, 2$ for $i \neq j$. Now the transitions just described occur with probabilities given by their relative rates, directly obtained as quotients with the sum of

the rates in the denominator. The waiting time between transitions is distributed exponentially with a rate parameter $\binom{n}{2} + n\theta/2 + n\sigma/2$, when there are n ancestors (in total) in the graph.

Note a *null* transition that leaves the ancestral configuration unchanged can occur. Events such as these can be removed with the use of generating functions. Removing an event in the (embedded) Markov chain above does not, however, filter the entire waiting time by the corresponding factor. The Markov chain contains dependencies among the set of possible transitions that can occur at any particular time. These dependencies correspond to the ancestral process and in particular to the percolation diagram as in Figure 1. Each member of the population produces two types of arrows, as events of independent Poisson processes. That is, these arrows occur independently of other members of the population and between both types of arrow. The null transition referred to above shows that certain 2-arrows are ignored and can be omitted from our consideration of the ancestral process. In fact, when the ancestor in the dual process from which the 2-arrow emanates is known to possess the A_2 allelic type, it is in effect ignored. (Note that looking forward in time this is then the ancestor hit by the 2-arrow.) That is, no branching event corresponding to that 2-arrow need occur. This omission is of no consequence for coalescence or mutation rates, since coalescence corresponds to δ -arrows, and mutation occurs on an arrow with probability zero. The adjusted rate of branching for the new waiting-time distribution is $n\sigma/2 \times n_1/n$, the rate of branching in the original graph multiplied by the proportion of A_1 ancestors in the graph at the time the branching event occurs. The (unconditional) rate of a generic branching event in the new Markov chain is unchanged since we have simply omitted the null transition from the original Markov chain.

Consider also a mutation event from a virtual A_1 to a virtual A_2 ancestor; a recursive term is given in SLADE and NEUHAUSER (2003). However, without the unordering factor, the expression is

$$n \frac{\theta}{2} \frac{n_2 + 1}{n} \frac{v_1}{n_1} q(\mathbf{r}, \mathbf{v} - \mathbf{e}_1 + \mathbf{e}_2) = n \frac{\theta}{2} \left[\frac{v_1}{n_1} q(\mathbf{r}, \mathbf{v} - \mathbf{e}_1) - \frac{v_1}{n} q(\mathbf{r}, \mathbf{v}) \right], \tag{B2}$$

where $q(\mathbf{r}, \mathbf{v})$ is the stationary sampling probability of the marginal allelic configuration $(r_1 + v_1, r_2 + v_2)$. This equation shows that the event just described can be separated into two distinct probabilistic components, a coalescent-type event and another null transition type of event. Operating in reverse to that which removed the null branching event, add a term $v_1\theta/2$ to the mutation rate in the waiting-time distribution. This introduces another null mutation transition to the Markov chain that would occur with this same rate. Thus, in the Markov chain that results, a null transition occurs with a sum total rate of zero.

Therefore, the minimal continuous-time Markov jump chain that describes the unstructured ASG, and the collecting phase here, makes transitions as

$$(\mathbf{r}, v_1) \rightarrow \begin{cases} (\mathbf{r} - \mathbf{e}_i, v_1) & \text{with rate } r_i(r_i - 1)/n_i, \\ (\mathbf{r}, v_1 - 1) & \text{with rate } v_1(v_1 - 1 + 2r_1)/n_1 + \theta v_1/n_1, \\ (\mathbf{r} + \mathbf{e}_i - \mathbf{e}_j, v_1) & \text{with rate } \theta \frac{n_i + 1}{n} r_j/n_j, \\ (\mathbf{r}, v_1 + 1) & \text{with rate } \sigma(n_1 + 1)/(n + 1), \end{cases} \tag{B3}$$

where again we have factored out $n/2$ from the above rates, and $i, j = 1, 2$ for $i \neq j$. These transitions occur with probabilities given by their relative rates. Note that the only virtual ancestors that ever exist in this formulation of the graph are A_1 . The waiting time between transitions in (B3) is now exponential $\left(\binom{n}{2} + n\theta/2(1 + v_1/n) + n_1\sigma/2\right)$.

