

# Metapopulation models for historical inference

JOHN WAKELEY

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

## Abstract

The genealogical process for a sample from a metapopulation, in which local populations are connected by migration and can undergo extinction and subsequent recolonization, is shown to have a relatively simple structure in the limit as the number of populations in the metapopulation approaches infinity. The result, which is an approximation to the ancestral behaviour of samples from a metapopulation with a large number of populations, is the same as that previously described for other metapopulation models, namely that the genealogical process is closely related to Kingman's unstructured coalescent. The present work considers a more general class of models that includes two kinds of extinction and recolonization, and the possibility that gamete production precedes extinction. In addition, following other recent work, this result for a metapopulation divided into many populations is shown to hold both for finite population sizes and in the usual diffusion limit, which assumes that population sizes are large. Examples illustrate when the usual diffusion limit is appropriate and when it is not. Some shortcomings and extensions of the model are considered, and the relevance of such models to understanding human history is discussed.

*Keywords:* coalescent, extinction, gene genealogies, metapopulation, migration

*Received 25 August 2003; revision received 5 November 2003; accepted 5 November 2003*

## Introduction

The rise of bio-molecular technologies over the last few decades has changed the field of biology quite dramatically. Within those subfields of biology that seek historical explanations for current patterns of biodiversity, as large sections of population genetics and molecular ecology do, the availability of molecular data has been a boon. Molecular data are the closest thing to a transcript of history that biologists are likely to obtain. They are currently being gathered at an unprecedented pace, with the promise of a future filled with powerful and unambiguous inferences. These data necessarily provide indirect evidence of history. They are like the results of a laboratory experiment but one in which the experimental protocols are only sketchily known. Thus, inferences about history made from molecular genetic data, such as DNA sequence data, depend on a framework of probabilistic models and statistical methods. The shift that this represents within the field of population genetics, from a forward-looking, classical view to a backward-looking,

genealogical approach is discussed in an excellent review by Ewens (1990).

The present work concerns one small part of this transformation of biology, namely the search for metapopulation models that include a sufficient amount of biological realism and, through a connection to the well-characterized coalescent process (Kingman 1982a,b; Hudson 1983; Tajima 1983), are amenable to popular computational methods of statistical inference. As any issue of *Molecular Ecology* or the recent book by Hanski & Gilpin (1997) will attest, many species are subdivided into locally breeding populations that exhibit metapopulation dynamics. Populations within a metapopulation may be connected by migration, they may be subject to extinction and recolonization, and they may grow or shrink over time. In addition, there may be changes in the number of populations, and in the rates of migration and of extinction/recolonization across the metapopulation. Finally, selection might be acting on genetic variation. In order to take advantage of the opportunities offered by burgeoning molecular data sets. All of these factors must be included in the developing structure for historical inference.

The specific goals here are rather more modest than this. In particular, it is shown that a result based on a separation

Correspondence: John Wakeley. Present address: 2102 Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA. E-mail: wakeley@fas.harvard.edu

of time-scale for subdivided populations — see Wakeley & Aliacar (2001) and references therein — holds for metapopulations with two kinds of recolonization and within which gamete production can either precede or follow extinction. The recent work of Lessard & Wakeley (2003) has shown that this result holds both when the population sizes are finite and under the usual diffusion approximation for a subdivided population. The usual diffusion approximation for a subdivided population, which dates back to Wright (1931), posits that population sizes approach infinity while scaled parameters, e.g.  $Nm$ , remain finite. Instead, the limiting ancestral process studied here exists in the limit as the number of populations in the metapopulation approaches infinity. The additional assumption of large population size is straightforward to include but it is not necessary to the result. Other assumptions of the model detailed below include selective neutrality of variation and no explicit spatial structure, although the latter could be adjusted (Wakeley & Aliacar 2001). The result is an approximation to the genealogical process for metapopulations that are divided into a large number of populations. It is hoped that the simple structure of the limiting ancestral process will aid in the development of efficient computational methods of inference and will facilitate understanding of the ways in which metapopulation dynamics shape genetic variation.

A focus on genealogies as depictions of history is one of the hallmarks of the new inferential approach to population genetics. From that starting point, however, two quite different camps are apparent. Working from the realization that substantial information about history may be present in the structure of genealogies, the methods in the field of intraspecific phylogeography (Avice *et al.* 1987), or simply phylogeography (Avice 2000), typically begin with a single tree reconstructed from the data. On the other hand, methods based on the coalescent assign little or no significance to single genealogies, instead averaging over them in order to make inferences (Griffiths & Tavaré 1994; Kuhner *et al.* 1995). Phylogeography has its roots in the field of phylogenetic systematics, which historically has been rather anti-statistical, while coalescent theory is firmly grounded in probability and statistics. The potential advantages and shortcomings of both approaches have been reviewed recently (Hey & Machado 2003; Knowles 2003; Wakeley 2003). One of the primary methods of phylogeography — nested clade analysis (Templeton *et al.* 1995) — has finally been tested using simulations, and the results are not encouraging (Knowles & Maddison 2002). This motivated Knowles & Maddison (2002) to propose the term statistical phylogeography to describe the emerging field that infuses phylogeography with coalescent theory and rigorous statistical techniques. There are, of course, strong connections between these modern approaches to the interpretation of DNA sequence data and the pioneer-

ing work of Malécot (e.g. Malécot 1948, 1975) and Wright (1951), who developed similar models and techniques but focused on allelic data.

The job of historical inference is at the intersection of biology, mathematics, statistics and computer science. It is made difficult, in part, because population genetic history is the result of the joint action of the many factors mentioned above. Thus, one of the major goals of theoretical work should be to identify cases in which simplified models, or approximations, are justified in spite of the fact that the actual processes are complicated. The remaining situations will require complicated models or will have to await the development of better theoretical approaches. The original coalescent model (Kingman 1982a, b; Hudson 1983; Tajima 1983) admits none of these complications, and the last two decades have seen it extended in many different ways (Hudson & Kaplan 1988, 1995; Kaplan *et al.* 1988, 1991; Krone & Neuhauser 1997; Neuhauser & Krone 1997; Nordborg 1999). Most of the resulting models are special cases of what has become known as the structured coalescent (Notohara 1990; Nordborg 1997, 2001; Wilkinson-Herbots 1998). The structured coalescent allows for different rates of coalescence within and between different classes of lineages, which can be defined by such things as allelic state or geographical location. It is a much needed and well-described model, but it is complicated because the history of a sample depends on many different parameters.

Recently, a number of related results have shown that the standard, unstructured coalescent arises in a variety of models with structure (Nordborg & Donnelly 1997; Möhle 1998a,b; Wakeley 1998, 1999). These are known as robustness results for the coalescent (Möhle 1998c), and are obtained when the ancestral process involves forces that occur on very different time scales. For example, Kingman's coalescent (Kingman 1982a,b) is a haploid model, but it has been shown to hold for large, two-sex diploid populations because individual genetic lineages will switch back and forth between males and females many times before they coalesce (Möhle 1998a). The only difference is that the rate of coalescence depends on the effective size of the population, which is a function of the numbers of males and females (Möhle 1998a). The results presented below are of this sort, and follow some recent similar work on related models. Genealogical history is a two-phase process: (i) a brief 'scattering' phase that amounts to a stochastic, structured sample size adjustment, and (ii) a much longer 'collecting' phase which is an unstructured coalescent process with an effective size that depends on the many parameters of the model (Wakeley 1999). This allows the large and growing body of knowledge about the analytical, computational, and inferential framework of Kingman's coalescent (Kingman 1982a,b) to be applied to structured populations simply by including the scattering phase.

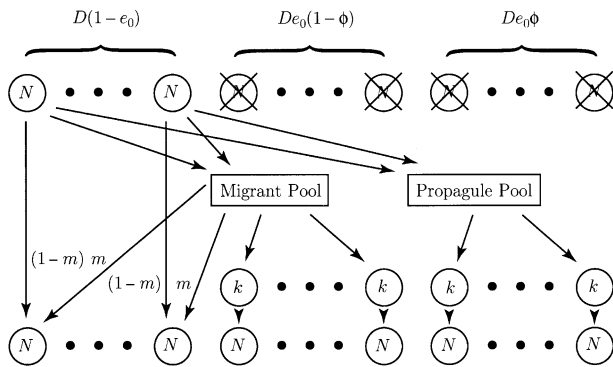


Fig. 1 A graphical depiction of the metapopulation model described in the text.

### A metapopulation model

The model used here is from Whitlock & McCauley (1990) and is a generalization of the model of Slatkin (1977). The metapopulation is divided into  $D$  populations each containing  $N$  haploid individuals. Note that in the population genetic literature these are often referred to as demes (Gilmour & Gregor 1939), and the metapopulation is often simply called the population. Generations are nonoverlapping, and follow the life cycle depicted in Fig. 1. At the beginning of each generation, a fixed number,  $D e_0$ , of populations is chosen at random from the metapopulation. These populations become extinct and do not contribute to the next generation (but see below). The  $D(1 - e_0)$  populations that do not become extinct each contribute an effectively infinite number of gametes, or newborns, to three different pools: their own gamete pool, a migrant gamete pool and a propagule pool. Gametes are contributed to the propagule pool in packets of size  $k$ , while the migrant pool and the gamete pools of the individual populations are unstructured. The  $D e_0$  populations that do become extinct are of two kinds:  $D e_0 \phi$  are recolonized from the propagule pool, and  $D e_0(1 - \phi)$  are recolonized from the migrant pool. In this formulation, in which gametes are produced after extinction, the parameters  $e_0$  and  $\phi$  can vary between zero and one, but  $e_0$  cannot be equal to one or the entire population would become extinct.

Every adult individual dies after contributing its gametes. The next generation of adults is formed, in the usual Wright–Fisher reproductive scheme (Fisher 1930; Wright 1931), by random sampling from the gamete pool(s), but with the structure imposed by Fig. 1. For each descendant of the populations which did not become extinct, a fraction  $1 - m$  of the next generation's adults comes from that population's gamete pool and a fraction  $m$  comes from the migrant pool. Of the populations that did become extinct,  $D e_0 \phi$  received  $k$  colonists from the propagule pool and  $D e_0(1 - \phi)$  receive  $k$  colonists from the migrant pool. Colonists from the propagule pool are certain to have come

from the same parental population. Colonists from the migrant pool may have come from any contributing population. The recolonized populations regain their original size  $N$  immediately by another round of Wright–Fisher sampling from the  $k$  colonists. This is a restricted version of the model of Whitlock & McCauley (1990) because of the assumption of a haploid organism and because Whitlock & McCauley (1990) had a broader vision of the parameter  $\phi$ . The model is clearly abstract and lacking realism in many respects, yet it is hoped that it captures some of the important features of metapopulation dynamics. A few shortcomings and possible extensions of the model are taken up in the Discussion. Here, it should be noted that the results can be applied to diploid organisms by a simple rescaling of the effective population size; see Möhle (1998a,b).

### The large- $D$ approximation

The model described above and depicted in Fig. 1 is complicated, but sample genealogies from such a metapopulation have a relatively simple structure. Even more complicated models, such as some that include explicit geography (Wakeley & Aliacar 2001), have this same simple structure. To obtain the result, it is necessary to make two further assumptions: that genetic variation is selectively neutral and that the metapopulation is comprised of a large number of populations. The first of these is not to be taken lightly, and there is great deal of historical and current debate about the role of selection in shaping genetic variation within and among populations and species (Kimura 1983; Golding 1994; Fay *et al.* 2001, 2002; Bustamante *et al.* 2002; Smith & Eyre-Walker 2002). The second is a safe assumption for many metapopulations, and has been a standard starting point since Wright (1931, 1940) and Levins (1968a,b), although its consequences have rarely been investigated explicitly as a mathematical limit. The derivation of the present result is given in the Appendix for a sample of size two. In the limit as  $D$  approaches infinity, it is shown that the history of the sample from the discrete time model converges to a continuous time process that has the scattering phase/collecting phase structure described above.

The derivation can be understood intuitively by referring to the matrix that describes the discrete time Markov process of coalescence. As shown in Table 1, this matrix can be written as the sum of two different matrices,  $\Pi_D = \mathbf{A} + \mathbf{B}/D$ , one which does not depend on  $D$  at all and one which is proportional to  $1/D$  when  $D$  is large. The sample or the ancestral lineages of the sample can be in one of three states: state 1 is both lineages in the same population, state 2 is the two lineages in different populations, and state 3 is the two lineages have coalesced. State 3 is an absorbing state, which means that the process is followed back in time only to the most recent common ancestor of the sample. The entries of  $\Pi_D$  are the probabilities of moving

between states in a single generation looking back. For example  $(\Pi_D)_{11}$  is the probability that the sample is in state 1 one generation back, given that it is in state 1 now. Part of this is in the matrix **A** and is the probability that the population is not extinct/recolonized and neither of the lineages migrates and they do not coalesce plus the probability that the population is propagule-extinct/recolonized and the lineages do not coalesce either in the extra recolonization sampling step or in the source population (see Fig. 1). The other part is in the matrix **B/D** and is the probability that one or other, or both, lineages move either by migration or extinction/recolonization (without coalescing in the extra sampling step) and they have the same source population but do not coalesce there. The probability of having the same source population is what creates the inverse dependence on  $D$ .

As  $D$  grows, the entries in **B/D**, which are the ones that bring separated lineages together into a single population, become proportionately smaller. The entries in **A** do not depend on  $D$ . Thus, the dynamics from state 1 will depend only on the top row of **A** in the limit as  $D$  tends to infinity, whereas transitions from state 2 depend on the rare events whose probabilities are in the second row of **B/D**. If the sample starts in state 2, it will stay there for an approximately exponentially distributed number of generations,  $D/b_2$  on average, then it will jump to either state 1 or state 3. That is,  $b_2/D$  is the probability that one or the other, or both, lineages move, by migration or extinction/recolonization, and they end up in the same population. While in state 2, the lineages are likely to move many times among populations. When the sample leaves state 2, it will either coalesce or go to state 1. Since the entries in **A** are much greater than the entries in **B/D** when  $D$  is large, the amount of time the sample stays in state 1 is much shorter than the amount of time it had spent in state 2. So, in a short time, the sample will either coalesce or it will move back to state 2 and this whole process will restart. As shown in the Appendix, in the limit  $D \rightarrow \infty$  and if time is measured in proportion to  $D$  generations, a continuous time approximation is valid in which the jumps from state 1 take an infinitesimal amount of time.

The result is a Kingman-type coalescent process (the collecting phase) that only needs to be adjusted for first jump (the scattering phase) from state 1 for a sample from a single population. The adjustment, for this sample of size two, is by a factor  $1 - F$ , which is the probability that the two lineages do not coalesce during the scattering phase. The quantity  $F$  is equivalent to one of the ways in which Wright's  $F_{ST}$  (Wright 1951) has been defined (Slatkin 1991; Charlesworth 1998). Its value here is given by

$$F = \frac{a_1 \frac{1}{N} + e_0 \frac{1}{k}}{1 - a_1 \left(1 - \frac{1}{N}\right)}$$

**Table 1** Backward, single-generation transition matrix  $\Pi_D$  for a sample of size two, split into two parts:  $\Pi_D = \mathbf{A} + \mathbf{B}/D$ . Specifically  $(\Pi_D)_{ij}$  is the probability that the sample moves from state  $i$  (row) to state  $j$  (column) in a single generation

$$\mathbf{A} = \begin{pmatrix} a_1 \left(1 - \frac{1}{N}\right) & 1 - a_1 - e_0 \frac{1}{k} & a_1 \frac{1}{N} + e_0 \frac{1}{k} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\frac{\mathbf{B}}{D} = \begin{pmatrix} \frac{1}{D} b_1 \left(1 - \frac{1}{N}\right) & -\frac{1}{D} b_1 & \frac{1}{D} b_1 \frac{1}{N} \\ \frac{1}{D} b_2 \left(1 - \frac{1}{N}\right) & -\frac{1}{D} b_2 & \frac{1}{D} b_2 \frac{1}{N} \\ 0 & 0 & 0 \end{pmatrix}$$

$$a_1 = (1 - e_0)(1 - m)^2 + e_0 \phi \left(1 - \frac{1}{k}\right)$$

$$b_1 = 1 - (1 - m)^2 + \frac{e_0}{1 - e_0} (1 - \phi) \left(1 - \frac{1}{k}\right)$$

$$b_2 = \frac{1}{1 - e_0} - \frac{D(1 - e_0) - 1}{D - 1} (1 - m)^2$$

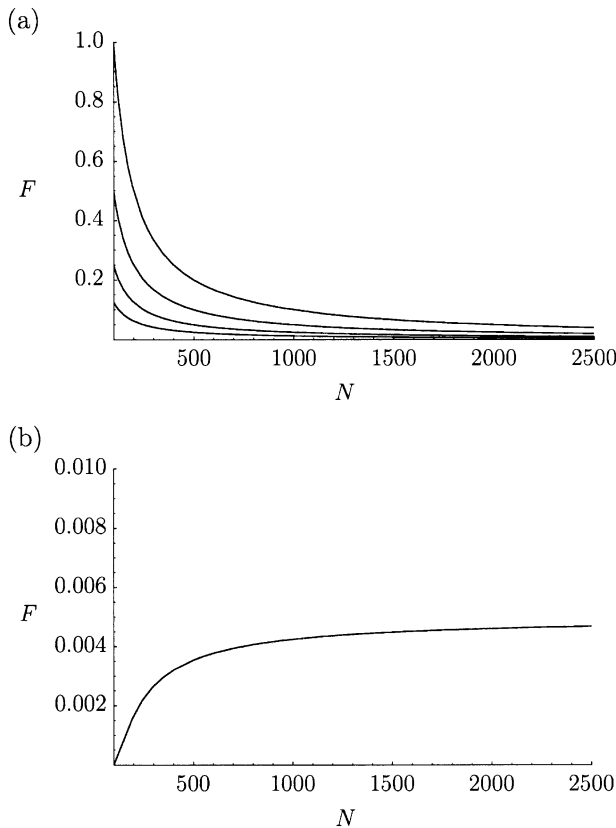
where  $a_1$  is given in Table 1, and is consistent with the expressions for  $F_{ST}$  studied by many authors (Slatkin 1977, 1991; Whitlock & McCauley 1990; Whitlock & Barton 1997; Pannell & Charlesworth 1999, 2000). In addition, the time scale of the collecting-phase coalescent process is determined by an effective population size

$$N_e = \frac{D(1 - e_0)}{\left[1 - (1 - e_0)^2(1 - m)^2\right] \left[F \left(1 - \frac{1}{N}\right) + \frac{1}{N}\right]}$$

which again is consistent with the work of others, but is also identical to the expression recently obtained by Rousset (2003). The difference here is that  $F$  and  $N_e$  are parameters in a well-defined stochastic process (Möhle 1998b), closely related to the unstructured coalescent (Kingman 1982a,b; Hudson 1983; Tajima 1983), rather than descriptors of the average behaviour, defined in a variety of different ways (Ewens 1982; Charlesworth 1998), of genetic drift within and among populations.

### Finite $N$ vs. large- $N$ approximations

The effects of  $D$ ,  $N$ ,  $m$ ,  $e_0$ ,  $\phi$ , and  $k$  on  $F$  and  $N_e$ , and thus on levels and patterns of genetic variation, have been described by several authors; see Pannell & Charlesworth (2000) for a review. The present results, and those of Rousset (2003), allow comparisons to be made between results for finite  $N$



**Fig. 2** The dependence of  $F$  on  $N$  for constant scaled parameters. In (a),  $E_0 = Ne_0$  is held constant at  $E_0 = 100$ , with  $m = 1$  and  $\phi = 0$ , and four different values of  $k$  are used: 1, 2, 4 and 8, from top curve to bottom curve. In (b),  $M = Nm$  is held constant at  $M = 100$ , with  $e_0 = 0$ . Both plots show values of  $F$  as  $N$  varies from 100 to 2500 individuals.

and those obtained under the typical diffusion approximation for a metapopulation which assumes that  $N$  is very large and  $m$  and  $e_0$  are very small. In this large- $N$  limit, patterns of genetic variation depend on the scaled parameters  $M = Nm$  and  $E_0 = Ne_0$ , and these are surprisingly good predictors even when  $N$  is not particularly big and  $m$  and  $e_0$  are not particularly small. However, when  $M$  or  $E_0$  is large,  $N$  has to be larger still for the assumption of small  $m$  and  $e_0$  to be met. Conversely, if  $N$  is not very large, then  $m$  and  $e_0$  must be very small for the usual diffusion approximations to be accurate. Figures 2 and 3 illustrate this for  $F$  and  $N_e$ , respectively.

Figure 2 shows the effect on  $F$  of varying  $N$  for constant values of  $M$  and  $E_0$ . In Fig. 2(a),  $F$  is shown to depend strongly on  $N$  when  $E_0 = 100$  in a model without restricted migration ( $m = 1$ ) and only migrant-pool colonization ( $\phi = 0$ ). On the far left of the graph,  $F$  is large because a high rate of extinction (large  $e_0$ ) is required to keep  $E_0 = 100$ . The dependence on  $N$  is strong even when  $N$  is as large as 500. In fact, with  $m = 1$  and  $\phi = 0$ , as assumed here, the expression for  $F$  reduces to  $e_0/k$ , so Fig. 2(a) simply plots  $F = 100/$

$(Nk)$ . Thus, the effect of  $N$  is stronger when  $k$  is smaller. On the far right of the graph, the effect of subdivision is small, i.e.  $F$  is close to zero, because  $e_0$  is small while  $m = 1$ . A similar strength of dependence on  $N$  is shown in Fig. 2(b), in which  $M = 100$  for a migration-only metapopulation ( $e_0 = 0$ ). In this case,  $F$  increases from zero when  $N = 100$  (and  $m = 1$ ) to a value greater than zero when  $N$  is large. However, although there is a strong dependence of  $F$  on  $N$ , the overall effect of subdivision in this case is small, i.e.  $F$  is small, because  $M = 100$  represents a significant amount of migration regardless of the value of  $m$  (Wright 1931).

As a second illustration that the large- $N$  approximation is sometimes undesirable, consider a model in which gamete production occurs before extinction, rather than after extinction as has been assumed thus far. If gametes are produced before extinction, then populations that become extinct can contribute to the next generation. The only effect of implementing this assumption is to multiply both  $b_1$  and  $b_2$  in Table 1 by the factor  $1 - e_0$ . This does not change  $F$ , but the effective population size is  $1 - e_0$  times smaller if gametes are produced after extinction than if gametes are produced before extinction. However, as  $N$  approaches infinity, and  $\lim_{N \rightarrow \infty} Ne_0 = E_0$  and  $\lim_{N \rightarrow \infty} Nm = M$ , both models give

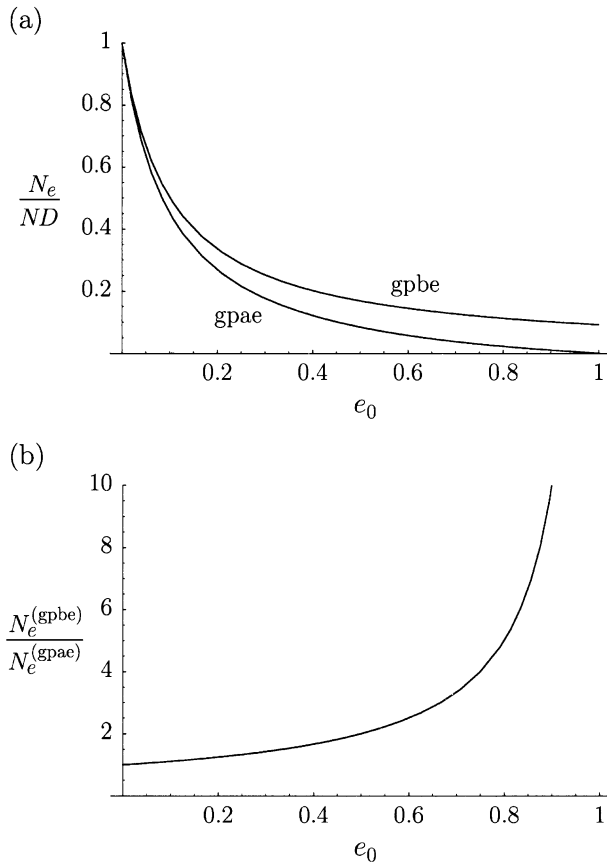
$$N_e = \frac{ND}{2(E_0 + M)F}$$

where

$$F = \frac{1 + \frac{E_0}{k}}{1 + 2M + E_0 \left( 1 - \phi + \phi \frac{1}{k} \right)}$$

These equations for  $N_e$  and  $F$  (or  $F_{ST}$ ) have been found previously by Whitlock & Barton (1997) and Whitlock & McCauley (1990), and are of course consistent with earlier work (Wright 1931, 1940; Slatkin 1977; Maruyama & Kimura 1980). What is interesting is that they do not depend on whether gamete production occurs before or after extinction.

Figure 3 compares the effective population size of two metapopulations in which the local population size is  $N = 100$  and in which the source of subdivision is extinction and migrant-pool recolonization ( $m = 1$  and  $\phi = 0$ ). Because  $N$  is not particularly large, this is a case in which the product  $Ne_0$ , or  $E_0$ , is not expected to be a good predictor unless  $e_0$  is quite small. The two metapopulations compared in Fig. 3 differ in the timing of gamete production: either before extinction (gpbe) or after extinction (gpae). Figure 3(a) plots the effective size of the metapopulations relative to the effective size of an unstructured population of the same total size. On the left, when the rate of extinction,  $e_0$ , is small, the effective sizes are very similar and



**Fig. 3** Comparison of the dependence of  $N_e$  on  $e_0$  when gamete production occurs before extinction (gpbe) and when gamete production occurs after extinction (gpae). In both (a) and (b),  $N = 100$ ,  $m = 1$ ,  $\phi = 0$  and  $k = 10$ . (a) shows  $N_e$  in both models compared to the effective size of a single, unstructured population of size  $ND$ . (b) shows the ratio of effective sizes under the two models.

both are close to the actual size of the metapopulation. On the right, when the rate of extinction  $e_0$  is large, both effective population sizes become much smaller than the total population size. However, they also diverge from each other, and the effective size of the metapopulation in which gamete production follows extinction is the much smaller of the two. In fact, as mentioned above, the ratio of the effective sizes of these two metapopulation is equal to  $1/(1 - e_0)$ , and this is plotted in Fig. 3(b).

Figures 2 and 3 might, alternatively, be used to support the use of the large- $N$  limit, in which  $\lim_{N \rightarrow \infty} N e_0 = E_0$  and  $\lim_{N \rightarrow \infty} Nm = M$ , because the resulting approximations for  $F$  and  $N_e$  are accurate even for moderate values of  $N$ ,  $m$  and  $e_0$ . However, the figures also show that the approximations worsen quickly when  $N$  decreases below some critical value or when  $e_0$  increases above some critical value. Large values of  $m$  appear less problematic since the population becomes panmictic as  $m$  approaches one. As the large- $D$  model holds for finite  $N$  and arbitrary  $m$  and  $e_0$ , as well as in the usual diffusion limit, it can be used to empirically

address the question of whether the sizes of  $N$ ,  $m$  and  $e_0$  are consistent with a large- $N$  approximation.

## Discussion

For metapopulations that consist of a large number of populations, genealogies are characterized by an initial scattering phase followed by an unstructured collecting phase coalescent. During the scattering phase, coalescent events can occur between samples from the same population but not between samples from different populations, and migration events and extinction/recolonization events move lineages to populations that do not already contain lineages ancestral to the sample. When each remaining ancestral lineage is in a separate population, the collecting phase coalescent process begins, and continues until the most recent common ancestor of the entire sample is reached. The collecting phase depends on those rare events that bring lineages together into the same population. Thus it is much longer than the scattering phase, the duration of which becomes negligible in the limit as  $D$  tends to infinity. The effect of this brief scattering phase can be profound, as this is what generates differential patterns of relationship within vs. between populations. Despite the complicated nature of the model, the collecting phase depends only an effective size,  $N_e$ , a composite parameter which is proportional to  $D$  and  $N$  but also depends on  $m$ ,  $e_0$ ,  $\phi$  and  $k$ . In contrast, the scattering phase is determined by the properties ( $N$ ,  $m$ ,  $e_0$ ,  $\phi$ ,  $k$ ) of the sampled populations. Simulations show that this approximation to the ancestral process for a sample from a metapopulation appears to hold for moderate  $D$  (Wakeley 1998; Lessard & Wakeley 2003; Pannell 2003).

Following the two-locus, migration-only case in which both finite and infinite  $N$  were treated recently (Lessard & Wakeley 2003), the Appendix shows that the usual diffusion limit for a metapopulation, which assumes that  $N \rightarrow \infty$  while  $Nm$  and  $N e_0$  remain finite, can be included in the model and the basic result is unchanged. In addition, the scattering/collecting structure is robust to some forms of explicit geography (Wakeley & Aliacar 2001), can include changes in demography over time (Wakeley 1999), and holds for samples larger than two both for finite  $N$  (Lessard & Wakeley 2003) and in the usual diffusion limit (Wakeley 1998). The large- $N$  version of the scattering phase has been described for metapopulations without propagule pool recolonization (Wakeley & Aliacar 2001), in which case tractable analytical descriptions are possible for samples larger than two. Here, for arbitrary  $N$ ,  $m$ ,  $e_0$  and  $k$ , the possibility of multiple coalescent events in a single generation during the scattering phase makes analysis more difficult, but could be modelled using simulations. An odd feature of the present model is the inclusion of an extra sampling or reproduction step for populations which become extinct and are recolonized. Slatkin (1977) originally proposed this

assumption explicitly to simplify the analysis and not to model any specific biological phenomenon. The result is that the populations that undergo extinction/recolonization have two generations over the same period of time as the other populations experience a single generation. One way to achieve parity among populations would be to include another reproduction step in Fig. 1 for the populations that do not become extinct. However, this would enforce an artificial two-generation structure on the model, with extinction/recolonization possible only every other generation. A better solution would be to replace the present model with one in which generations are overlapping and populations can be in a number of states, with regard to when they were last recolonized, and migration could occur at any time. Ingvarsson (1997) studied an aspect of this problem, allowing recolonized populations to grow from size  $k$  to  $N$  over a number of generations and for migration to occur during this growth phase.

#### *All variation is ancestral*

Because mutations along the branches of the genealogy are the source of polymorphism among the members of a sample, the large- $D$  result has dramatic implications for interpreting genetic variation. In particular, since the duration of the scattering phase is negligible in comparison to that of the collecting phase, all genetic variation in a large- $D$  metapopulation results from mutations that occurred in the ancestral (collecting phase) part of the history. Samples from the metapopulation tap into this ancestral variation via the scattering phase, which again is a stochastic sample size adjustment that determines patterns of identity among samples from the same population. This is particularly evident for a sample of size two from the same population, in which there is a probability  $F$  that the samples coalesce immediately so that there is not even a chance of a mutation between them. With probability  $1 - F$  the samples enter the collecting phase, so that their coalescence time is exponentially distributed and some mutations can occur. This also has consequences for population mutation rates and patterns of polymorphism. In the usual diffusion approximation for a haploid subdivided population, the mutation parameter is taken to be  $2Nu$ , or  $4Nu$  if the organisms are diploid and monoecious. Here, the natural way to include mutations is to set  $\theta = 2N_e u$ , where  $N_e$  is the effective size of the collecting phase coalescent. In other words, levels of polymorphism will depend on  $m$ ,  $e$ ,  $\phi$  and  $k$  in addition to  $ND$ . Let  $\pi_w$  and  $\pi_b$  be the number of nucleotide differences between two sequences sampled from the same population ('within') or sampled from two different populations ('between'), and assume that every mutation creates a new polymorphic site (Kimura 1969; Watterson 1975). Then, with  $\theta = 2N_e u$ , the expected numbers of pairwise differences within and between populations are:

$$E[\pi_w] = (1 - F)\theta$$

$$E[\pi_b] = \theta$$

In the case of the island migration model, without extinction/recolonization, these equations reduce to the familiar result that  $E[\pi_w] = 2NDu$  and  $E[\pi_b] = 2Ndu\{1 + [1/(2Nm)]\}$  (Slatkin 1987; Strobeck 1987). In general, however, e.g. with extinction/recolonization, the expected number of pairwise differences within populations is not the same as in a single, panmictic population of the same total size. Although these results for pairwise differences do not offer the hope to distinguish the effects of migration from those of extinction/recolonization, results for the frequencies of polymorphisms in larger samples indicate that this will in fact be possible (Wakeley & Aliacar 2001). In the case of finite  $N$ , the equations for  $E[\pi_w]$  and  $E[\pi_b]$  above assume that  $\lim_{D \rightarrow \infty} Du$  is finite, and if the usual diffusion limit is included they assume that  $\lim_{D \rightarrow \infty} \lim_{N \rightarrow \infty} DNu$  is finite. They hold for a sample of low mutation rate data, such as DNA sequence data, from a large metapopulation. One consequence of this is that, in order for  $\theta = 2N_e u \propto Du$  to be finite, the mutation parameter for an individual population,  $2Nu$ , must be so small as to be negligible. On the other hand, if  $2Nu$  is not small in a metapopulation containing a large number of populations, then the metapopulation mutation rate will be very large. In fact,  $\theta$  would have to be infinite in limit  $D \rightarrow \infty$ . A model of this sort could be useful for analysing some types of genetic data, such as allozyme data or perhaps microsatellite data. In the former case, by assuming an infinite alleles mutation model, Slatkin (1982) showed that migration becomes equivalent to mutation since every migrant lineage will be of a unique allelic type. [See also Vitalis & Couvet (2001a,b) who use this same idea in a study of two-locus identity probabilities.] Considered from the standpoint of sequence data, a model with  $2Nu$  non-negligible would predict an infinite number of mutations/polymorphisms, and this would be contrary to observations of DNA data from metapopulations. For microsatellite or other allelic data in which there was a finite number of possible allelic types, the infinite number of collecting-phase mutations would mean that every collecting-phase lineage would be a random sample from the equilibrium distribution of allelic types.

#### *Humans as a metapopulation*

Excoffier (2003) proposes a hypothetical model for human history that permits mutations during the scattering phase, yet predicts a finite total number of mutations/polymorphisms. The model of Excoffier (2003) assumes extreme metapopulation growth, in which one population produces an effectively infinite number of populations over a short period of time. Mutations occur during the scattering

phase at the usual rate, which here would be with probability  $u$  per lineage per generation. Migration events during the scattering phase still send lineages off to populations which do not contain other ancestral lineages, but the collecting phase is cut short well before the first coalescent event occurs and, on average, after only a finite number of mutations have occurred. Prior to this, all lineages are assumed to have come from one population which gave rise to the entire metapopulation. This model is related to one proposed by Takahata (1995), except that it does not include extinction and recolonization and that the number of populations is assumed to be large. The assumption of a large number of populations does appear justified for humans, although delineating human populations is by no means a simple task (Cavalli-Sforza *et al.* 1994). Using one measure, the number of different human languages and dialects (Grimes 2000), there could be about 6800 different human populations. Excoffier (2003) proposed the extreme recent growth model after Ray *et al.* (2003) discovered a scattering and collecting phase structure to genealogies under a model of range expansion on a two-dimensional grid of populations. The simulations of Ray *et al.* (2003) showed that the extent to which genetic variation in a sample from a single population reflects range expansion depends on  $Nm$  for the population, and thus on the scattering phase. Ray *et al.* (2003) suggested this as a potential explanation of why different human populations do or do not show evidence of ancient growth. This illustrates, within the context of a spatially explicit model, the importance of accounting for population structure in making inferences about human history.

### Acknowledgements

I thank Laurent Excoffier for the invitation to contribute to this special volume. I also thank Laurent Excoffier and three anonymous reviewers for comments on the manuscript. Special thanks go to Sabin Lessard for ongoing discussions of these issues and for detailed comments on the manuscript. This work was supported by a Career Award (DEB-0133760) from the National Science Foundation.

### References

- Avise JC (2000) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge, MA.
- Avise JC, Arnold J, Ball RM, *et al.* (1987) Intraspecific phylogeography: the mitochondrial bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.
- Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL (2002) The cost of inbreeding in *Arabidopsis*. *Nature*, **416**, 531–534.
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes*. Princeton University Press, New Jersey.
- Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.
- Ewens WJ (1982) On the concept of effective size. *Theoretical Population Biology*, **21**, 373–378.
- Ewens WJ (1990) Population genetics theory – the past and the future. In: *Mathematical and Statistical Developments of Evolutionary Theory* (ed. Lessard S), pp. 177–227. Kluwer Academic Publishers, Amsterdam.
- Excoffier L (2003) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, (in press).
- Fay JC, Wyckoff GJ, Wu C-I (2001) Positive and negative selection in the human genome. *Genetics*, **158**, 1227–1234.
- Fay JC, Wyckoff GJ, Wu C-I (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*, **415**, 1024–1026.
- Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon, Oxford.
- Gilmour JSL, Gregor JW (1939) Demes: a suggested new terminology. *Nature*, **144**, 333.
- Golding B (1994) *Non-Neutral Evolution*. Chapman & Hall, New York.
- Griffiths RC, Tavaré S (1994) Ancestral inference in population genetics. *Statistical Science*, **9**, 307–319.
- Grimes BF (2000) *Ethnologue*, 14th edn, Vol. 1. SIL International, Dallas.
- Hanski I, Gilpin ME (1997) *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic Press, San Diego.
- Hey J, Machado CA (2003) The study of structured populations – new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 535–543.
- Hudson RR (1983) Testing the constant-rate neutral allele model with protein sequence data. *Evolution*, **37**, 203–217.
- Hudson RR, Kaplan NL (1988) The coalescent process in models with selection and recombination. *Genetics*, **120**, 831–840.
- Hudson RR, Kaplan NL (1995) Deleterious background selection with recombination. *Genetics*, **141**, 1605–1617.
- Ingvansson PK (1997) The effect of delayed population growth on the genetic differentiation of local populations subject to frequent extinctions and recolonizations. *Evolution*, **51**, 29–35.
- Kaplan NL, Darden T, Hudson RR (1988) Coalescent process in models with selection. *Genetics*, **120**, 819–829.
- Kaplan NL, Hudson RR, Iizuka M (1991) Coalescent processes in models with selection, recombination and geographic subdivision. *Genetic Research Cambridge*, **57**, 83–91.
- Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to the steady flux of mutations. *Genetics*, **61**, 893–903.
- Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Kingman JFC (1982a) The coalescent. *Stochastic Processes Applications*, **13**, 235–248.
- Kingman JFC (1982b) On the genealogy of large populations. *Journal of Applied Probability*, **19A**, 27–43.
- Knowles LL (2003) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- Krone SM, Neuhauser C (1997) Ancestral processes with selection. *Theoretical Population Biology*, **51**, 210–237.
- Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis–Hastings sampling. *Genetics*, **140**, 1421–1430.



- Lessard S, Wakeley J (2003) The two-locus ancestral graph in a subdivided population: convergence as the number of demes grows in the island model. *Journal of Mathematical Biology*, in press.
- Levins R (1968a) *Evolution in Changing Environments*. Princeton University Press, New Jersey.
- Levins R (1968b) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**, 237–240.
- Malécot G (1948) *Les Mathématiques de L'hérédité*. Masson, Paris. [extended translation (1969) *The Mathematics of Heredity*. W. H. Freeman, San Francisco].
- Malécot G (1975) Heterozygosity and relationship in regularly subdivided populations. *Theoretical Population Biology*, **8**, 212–241.
- Maruyama T, Kimura M (1980) Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences USA*, **77**, 6710–6714.
- Möhle M (1998a) Coalescent results for two-sex population models. *Advances in Applied Probability*, **30**, 513–520.
- Möhle M (1998b) A convergence theorem for Markov chains arising in population genetics and the coalescent with partial selfing. *Advances in Applied Probability*, **30**, 493–512.
- Möhle M (1998c) Robustness results for the coalescent. *Journal of Applied Probability*, **35**, 438–447.
- Neuhauser C, Krone SM (1997) The genealogy of samples in models with selection. *Genetics*, **145**, 519–534.
- Nordborg M (1997) Structured coalescent processes on different time scales. *Genetics*, **146**, 1501–1514.
- Nordborg M (1999) The coalescent with partial selfing and balancing selection: an application of structured coalescent processes. In: *Statistics in Molecular Biology and Genetics. IMS Lecture Notes-Monograph Series* (ed. Seillier-Moisewitsch F), Vol. 33, pp. 56–76. Institute of Mathematical Statistics, Hayward, CA.
- Nordborg M (2001) Coalescent theory. In: *Handbook of Statistical Genetics* (eds Balding DJ, Bishop MJ, Cannings C), pp. 179–212. John Wiley & Sons, Chichester.
- Nordborg M, Donnelly P (1997) The coalescent process with selfing. *Genetics*, **146**, 1185–1195.
- Notohara M (1990) The coalescent and the genealogical process in a geographically structured population. *Journal of Mathematical Biology*, **29**, 59–75.
- Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **57**, 949–961.
- Pannell JR, Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **53**, 664–676.
- Pannell JR, Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London, Series B*, **355**, 1851–1864.
- Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.
- Rousset F (2003) Effective size in simple metapopulation models. *Heredity*, **91**, 107–111.
- Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, **12**, 253–262.
- Slatkin M (1982) Testing neutrality in a subdivided population. *Genetics*, **100**, 533–545.
- Slatkin M (1987) The average number of sites separating DNA sequences drawn from a subdivided population. *Theoretical Population Biology*, **32**, 42–49.
- Slatkin M (1991) Inbreeding coefficients and coalescence times. *Genetics Research, Cambridge*, **58**, 167–175.
- Smith NG, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature*, **415**, 1022–1024.
- 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.
- Takahata N (1995) A genetic perspective on the origin and history of humans. *Annual Review of Ecological Systematic*, **26**, 343–372.
- Templeton AR, Routman E, Phillips C (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.
- Vitalis R, Couvet D (2001a) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics*, **157**, 911–925.
- Vitalis R, Couvet D (2001b) Two-locus identity probabilities and identity disequilibrium in a partially selfing subdivided population. *Genetics Research Cambridge*, **77**, 67–81.
- Wakeley J (1998) Segregating sites in Wright's island model. *Theoretical Population Biology*, **53**, 166–175.
- Wakeley J (1999) Non-equilibrium migration in human history. *Genetics*, **153**, 1863–1871.
- Wakeley J (2003) Inferences about the structure and history of populations: coalescents and intraspecific phylogeography. In: *The Evolution of Population Biology—Modern Synthesis* (eds Singh R, Uyenoyama M, Jain S), Cambridge University Press, Cambridge, in press.
- Wakeley J, Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics*, **159**, 893–905 [Corrigendum (Figure 2). *Genetics*, **160**, 1263–1264].
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.
- Whitlock MC, Barton NH (1997) The effective size of a subdivided population. *Genetics*, **146**, 427–441.
- Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.
- Wilkinson-Herbots HM (1998) Genealogy and subpopulation differentiation under various models of population structure. *Journal of Mathematical Biology*, **37**, 535–585.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Wright S (1940) Breeding structure of populations in relation to speciation. *American Naturalist*, **74**, 232–248.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

---

The author is in the Department of Organismic and Evolutionary Biology at Harvard University. His main research interests are in the genetical theory of structured populations.

---

**Appendix**

The large- $D$  approximation follows from a straightforward application of Möhle is (1998b) Theorem 1. For the metapopulation model considered here, the theorem states that the discrete time Markov process with transition matrix  $\Pi_D = \mathbf{A} + \mathbf{B}/D$  converges in distribution to a continuous time process with transition matrix

$$\Pi(t) = \lim_{D \rightarrow \infty} (\mathbf{A} + \mathbf{B}/D)^{[Dt]} = \mathbf{P}e^{t\mathbf{G}}$$

for all  $t > 0$ , and infinitesimal generator  $\mathbf{G} = \mathbf{PBP}$ , where

$$\mathbf{P} = \lim_{t \rightarrow \infty} \mathbf{A}^t$$

The matrices  $\mathbf{A}$  and  $\mathbf{B}/D$  are given in Table 1. Again, state 1 is when the two lineages are in the sample population, state 2 is when the two lineages are in different populations, and state 3 is when they have coalesced into a single ancestral lineage. These matrices,  $\mathbf{A}$  and  $\mathbf{B}$ , have the same structure as the corresponding matrices for the case of a partially selfing population considered by Nordborg & Donnelly (1997) and by Möhle (1998b), with the populations here corresponding to individuals in the partial-selfing model. In fact, the diploid, partial selfing model is a special case of the model discussed in the text in which gametes are produced before extinction, and if  $N = 2, m = 1, e_0 = s, \phi = 1$ , and  $k \rightarrow \infty$ .

The matrix  $\mathbf{P}$  represents jumps that are instantaneous on the time scale of the large- $D$  continuous time approximation. This includes the scattering phase for the sample. It is readily obtained from the matrix  $\mathbf{A}$  as

$$\mathbf{P} = \begin{pmatrix} 0 & 1 - F & F \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where

$$F = \frac{a_1 \frac{1}{N} + e_0 \frac{1}{k}}{1 - a_1 \left(1 - \frac{1}{N}\right)}$$

is the probability that two lineages currently in the same population coalesce before they are separated, either by migration or extinction/recolonization, into different populations. The matrix  $\mathbf{G} = \mathbf{PBP}$  is given by

$$\mathbf{G} = \begin{pmatrix} 0 & -c(1 - F) & c(1 - F) \\ 0 & -c & c \\ 0 & 0 & 0 \end{pmatrix}$$

in which

$$c = b_2 \left[ F \left( 1 - \frac{1}{N} \right) + \frac{1}{N} \right].$$

The exponential form of  $\mathbf{G}$  is given by  $e^{t\mathbf{G}} = \sum_{i=1}^3 r_i l_i e^{\lambda_i t}$ , where  $\lambda_i, r_i$  and  $l_i$  are the eigenvalues and the right and left eigenvectors, respectively, of the matrix above, with the vectors normalized so that  $r_i l_i = 1$  for  $1 \leq i \leq 3$ . These are  $\lambda_1 = \lambda_2 = 0, \lambda_3 = -c, r_1 = (0, 1, 1), r_2 = (1, 0, 0), r_3 = (1 - F, 1, 0), l_1 = (0, 0, 1), l_2 = (1, F - 1, 1),$  and  $l_3 = (0, 1, -1)$ . Finally

$$\Pi(t) = \mathbf{P}e^{t\mathbf{G}} = \begin{pmatrix} 0 & (1 - F)e^{-ct} & 1 - (1 - F)e^{-ct} \\ 0 & e^{-ct} & 1 - e^{-ct} \\ 0 & 0 & 1 \end{pmatrix}$$

describes the ancestral process for a sample of two lineages when time is measured in units of  $D$  generations and  $D$  is large. Specifically  $(\Pi(t))_{ij}$  is the probability that the sample is in state  $j$  at time  $t$  in the past, given that it was sampled in state  $i$  in the present. Thus, the time to common ancestry (state 3) for a sample of two sequences from two different populations (state 2) is exponentially distributed with rate  $c$ . For a sample of two sequences from the same population (state 1), the rate of coalescence is also equal to  $c$  but there is an additional factor  $(1 - F)$  which represents the probability that the sequences do not initially coalesce. The other factor  $F$  can be thought of as the probability density of the coalescent time exactly at  $t = 0$ .

Rescaling time again, now by the factor  $c$ , makes the rate of coalescence for the pair lineages equal to one. Thus, the ancestry of a sample of two sequences from two different populations is given by Kingman 1982a,b) coalescent process when time is measured in units of

$$N_e = \frac{D}{c} = \frac{D(1 - e_0)}{\left[ 1 - (1 - e_0)^2(1 - m)^2 \right] \left[ F \left( 1 - \frac{1}{N} \right) + \frac{1}{N} \right]}$$

generations. As before this rescaling, the ancestry of a sample of two sequences from a single population follows this same coalescent process, but only if the sample does not coalesce during the scattering phase. As  $N$  approaches infinity, and  $\lim_{N \rightarrow \infty} Ne_0 = E_0$  and  $\lim_{N \rightarrow \infty} Nm = M$ , the quantities  $F$  and  $N_e$  converge on the expressions give in the text, which shows that the usual diffusion assumption can be added to the large- $D$  model. Following Lessard & Wakeley (2003), note that the large- $N$  version of the large- $D$  approximation does not depend on the order in which the limits are taken:  $D \rightarrow \infty$  first, as above, or  $N \rightarrow \infty$  first. In the latter case, after applying the definition of the exponential matrix (i.e. Möhle's theorem but with  $\mathbf{A}$  as the identity matrix  $\mathbf{I}$ ), so that time is measured in units of  $N$  generations, the matrix  $\mathbf{G}$  contains both  $O(1)$  and  $O(1/D)$  terms. It is necessary

to reapply a continuous-time analogue of the theorem of Möhle (1998b) as in Lessard & Wakeley (2003) to  $\mathbf{G} = \mathbf{A}^* + \mathbf{B}^*/D$ , in which

$\mathbf{A}^* =$

$$\begin{pmatrix} -1 - 2M - E_0 \left[ (1 - \phi) + \phi \frac{1}{k} \right] & 2M + E_0(1 - \phi) \left( 1 - \frac{1}{k} \right) & 1 + E_0 \frac{1}{k} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{B}^* = \begin{pmatrix} 2M + E_0(1 - \phi) \left( 1 - \frac{1}{k} \right) & -2M - E_0(1 - \phi) \left( 1 - \frac{1}{k} \right) & 0 \\ 2(E_0 + M) & -2(E_0 + M) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Then,  $\mathbf{P}^* = \lim_{t \rightarrow \infty} e^{t\mathbf{A}^*}$  and  $\Pi(t) = \mathbf{P}^* e^{t\mathbf{G}^*}$  where  $\mathbf{G}^* = \mathbf{P}^* \mathbf{B}^* \mathbf{P}^*$ . The result  $\Pi(t)$  is identical to the one found above. This lack of dependence on the order of the limits follows from the fact that the backwards transition matrix can be written  $\Pi = \mathbf{I} + \mathbf{A}^*/N + \mathbf{B}^*/(ND) + \mathbf{O}(N, D)$  as in Lessard & Wakeley (2003).

