Gene Genealogies in a Metapopulation

John Wakeley and Nicolas Aliacar

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received May 18, 2001
Accepted for publication July 18, 2001

ABSTRACT

A simple genealogical process is found for samples from a metapopulation, which is a population that is subdivided into a large number of demes, each of which is subject to extinction and recolonization and receives migrants from other demes. As in the migration-only models studied previously, the genealogy of any sample includes two phases: a brief sample-size adjustment followed by a coalescent process that dominates the history. This result will hold for metapopulations that are composed of a large number of demes. It is robust to the details of population structure, as long as the number of possible source demes of migrants and colonists for each deme is large. Analytic predictions about levels of genetic variation are possible, and results for average numbers of pairwise differences within and between demes are given. Further analysis of the expected number of segregating sites in a sample from a single deme illustrates some previously known differences between migration and extinction/recolonization. The ancestral process is also amenable to computer simulation. Simulation results show that migration and extinction/recolonization have very different effects on the site-frequency distribution in a sample from a single deme. Migration can cause a U-shaped site-frequency distribution, which is qualitatively similar to the pattern reported recently for positive selection. Extinction and recolonization, in contrast, can produce a mode in the site-frequency distribution at intermediate frequencies, even in a sample from a single deme.

THE standard neutral coalescent model (Kingman 1982a,c; Hudson 1983; Tajima 1983) assumes a panmictic species with a constant, large, effective population size over time within which no selective differences exist. While this model is often applied in the analysis of genetic data, its assumptions are probably inappropriate for most organisms. That is, many species are subdivided and/or have changed in size over time and/or are subject to natural selection. The study of genealogical processes in populations subject to these forces has been a major part of the recent effort in population genetics. The key parameter of the coalescent is the effective population size, N_e, because this is what determines the time scale of the process. It is important to note that some biologically interesting characteristics of species are manifest only through this effective size. These include the distribution of offspring number among individuals in the population and the details of the age structure of the population. The robustness of the coalescent to these features is generally thought to be a positive aspect of the model. However, it might also be considered disadvantageous, or at least unfortunate, that genetic data from a large population will not contain information about these important biological characteristics of organisms.

The mathematical simplicity of Kingman's coalescent follows from one important assumption about the population: that lineages or alleles are exchangeable. Roughly speaking, exchangeable lineages are ones whose predicted properties are unchanged if they are relabeled or permuted. More precise definitions can be found in Kingman (1982b) and Aldous (1985). In the case of population subdivision, which is the focus of the present work, the genealogy of a sample depends on the locations of the lineages, so the lineages are not exchangeable. For example, very different genealogies result if all n members of a sample are from the same subpopulation or deme than if one is from one deme and the other $n - 1$ are from another. Even when the entire sample is from a single deme, so that the present-day lineages appear to be exchangeable, the lineages ancestral to the sample will not be exchangeable if they are in different demes. A similar situation holds for natural selection, but in this case the labels are the allelic states of the lineages rather than the geographic locations.

The facility with which extensions to the coalescent can be made depends on this problem of exchangeability. For instance, when the effective size of the population changes over time, the analysis is relatively straightforward because lineages remain exhangeable (Donnelly and Tavaré 1995; Slatkin 1996). In the case of natural selection, Neuhauser and Krone (1997) recently constructed a "dual process" (Donnelly 1984) for an ancestral selection graph in which lineages are exchangeable. Other approaches to selection assumed that it is strong enough to be treated in a similar manner to the problem of population subdivision (Kaplan *et al.* 1988, 1989; Hudson

Corresponding author: John Wakeley, 2102 Biological Laboratories, 16 Divinity Ave., Cambridge, MA 02138.
E-mail: wakeley@fas.harvard.edu

and Kaplan 1995; Nordborg *et al.* 1996; Nordborg 1997). Formally, population subdivision is modeled using the structured coalescent (Notohara 1990, 1997; Wilkinson-Herbots 1998), but analytical results are difficult to obtain because lineages are not exchangeable. The structured coalescent applies to populations in which the per-generation migration rates are on the order of the reciprocal of the deme sizes. However, making this assumption does not simplify the analysis much, as exact expressions for samples of size two (Nagylaki 1998) are about as complicated as those coming out of the structured coalescent. If we are confident that the demography of a population is such that the general structured coalescent process is applicable, then simulation-based maximum-likelihood methods, like those of Beerli and Felsenstein (1999) and Bahlo and Griffiths (2000), may provide the best framework for historical inference. However, there are practical issues in the implementation of these methods. Chief among these in relation to the present work, it is unclear how to account for possibly numerous unsampled demes.

When the demography of a subdivided population is such that lineages are exchangeable, genealogical models that are robust to some of the details of demography can be found and applied. For example, when migration rates among demes are high relative to the sizes of demes, the strong-migration limit of Nagylaki (1980, 2000) and Notohara (1993) approximates the behavior of the population. A Kingman-type coalescent describes the genealogy of the sample but with an effective size that depends on the pattern of migration among demes. This result follows from a separation of timescales between fast migration and slow coalescence (Nagylaki 1980; Nordborg 1997; Möhle 1998b). Because lineages are exchangeable in the strong-migration limit, the structure of the population will only be manifest in the effective size of the coalescent process. In particular, if a sample is taken from such a population, levels of polymorphism within and between demes will be the same. The existence of the strong-migration limit explains why geographic structure is sometimes not observed in samples, even in widely dispersed species that are obviously not panmictic across their entire range.

In fact, it is not uncommon to find evidence of geographic structure in genetic data (Slatkin 1985). With this motivation, the present model provides a framework for historical inference using samples from a metapopulation. A metapopulation is a population subdivided into many different demes among which there is some pattern of migration, extinction, and recolonization. As with the migration-only cases studied previously (Wakeley 1998, 2001), a simple genealogical process exists for samples from a population in which there are a large number of demes. The result holds for a fairly broad class of population structures and has similarities both to the structured coalescent and to the strong-migration limit. In common with the structured coalescent, the scaled rates of demic migration and extinction/recolonization ($M = 2Nm$ and $E = 2Ne$ below) are assumed to be finite. However, as the number of demes in the population becomes large, there is also a strong-migration limit for movement among types of demes (defined below). Lineages become exchangeable in this large-number-of-demes model only after a short burst of within-deme coalescent events. Thus the model predicts higher genetic variation among than within demes. In contrast to the small-number-of-demes structured coalescent (Wilkinson-Herbots 1998) and exact approaches (Nagylaki 1998), in which analytic results are typically confined to samples of size two, here results for arbitrary samples can be obtained. The large-number-of-demes model does not predict isolation by distance in the sense of Wright (1943), but such a pattern could result if the sizes of demes, in numbers of individuals, are positively correlated with their geographic extent; see the discussion.

The term metapopulation was introduced by Levins (1968, 1969) to describe a population that is subdivided into a large number of discrete demes, each of which is subject to random extinction and recolonization. Originally, the concern was for the numbers or fractions of empty and full demes in the population. Later metapopulation models focused on within-deme dynamics and included other processes, such as migration among demes. Over the last 30 years, metapopulation biology has grown into an active subfield of biology as a whole. Most of the emphasis has been on empirical and theoretical ecology. The recent book by Hanski and Gilpin (1997) gives a good overview of the subject. Of course, the study of subdivided populations has, from the beginning, been an important part of population genetics (Wright 1931). The Levins-type metapopulation was promoted by Wright (1940) as a demography that could lead to rapid evolution or speciation. Nevertheless, the rise of metapopulation ecology beginning in the 1970s caused a coincident increase in research on the genetics of metapopulations. Hanski (1998) and Pannell and Charlesworth (2000) provide thorough histories of these developments from the ecological and genetic perspectives, respectively.

Slatkin (1977) described the two fundamental conflicting consequences of extinction and recolonization in a subdivided population. One is the added genetic drift within demes that can occur when extinct demes are recolonized by a small number of individuals. This will tend to increase the level of differentiation among demes. Founder effects, or bottlenecks, such as this can also substantially decrease effective size of the population. The second consequence of extinction and recolonization emphasized by Slatkin (1977) is the increased amount of genetic exchange among demes that results from the movement of colonists across the population. This, like regular migration, will tend to decrease the level of differentiation among demes. These same two

forces were envisioned by Wright (1940) to facilitate the fixation of chromosomal rearrangements or other strongly underdominant mutations. Slatkin (1977) illustrated these effects with quantitative predictions about genetic variation within and between demes for two different models of extinction and recolonization: the propagule-pool model and the migrant-pool model.

In the propagule-pool model, the k individuals that recolonize extinct demes come from a single source deme, and each deme in the population has an equal chance of providing these founders. In the migrant-pool model, the k founders come from the migrant pool, which each deme in the population contributes to equally, so all k may have different source demes. Thus, the structure of movement among demes is similar to that in the island model of migration (Wright 1931; Maruyama 1970; Latter 1973). Under the above assumptions, Slatkin (1977) studied recurrence relations for the probabilities of identity-by-descent for two gene copies sampled either from the same deme or from different demes. He also found expressions for the effective number of alleles in the metapopulation. Maruyama and Kimura (1980) studied probabilities of identity under a propagule-pool model and also obtained expressions for the effective size of the metapopulation. Wade and McCauley (1988) reformulated the model in terms of F_{ST}. Whitlock and Barton (1997) derived formulas for the effective size of a more general metapopulation in which demes may vary in size and within which there can be selection. Pannell and Charlesworth (1999) recast Slatkin's (1977) model in terms of genetic diversity within and between demes and emphasize the important point that no single measure (*e.g.*, F_{ST}) is sufficient to characterize a subdivided population. A full account of research on Slatkin's (1977) model can be found in a recent review by Pannell and Charlesworth (2000).

The model considered here includes variation in the characteristics of demes and allows for structure in the pattern of movement of lineages, by migration or by recolonization, across the population. It is not, therefore, an island model in the strict sense of Wright (1931) or in the general sense of Wakeley (2001). Similar to genealogies in a large migration-only population (Wakeley 1998, 1999), it is shown here that, when the number of demes in the population is large, the genealogy of a sample includes two phases. There is a short recent part of the history, which I have elsewhere called the "scattering" phase (Wakeley 1999), and a more ancient "collecting" phase that dominates the history. Coalescent events during the scattering phase are the source of the within- *vs.* between-deme structure of genetic variation in a sample. The collecting phase is a Kingman-type coalescent process, with an effective size determined by the rate and pattern of movement across the population and by the distribution of deme sizes and propagule sizes. In addition to this effective size, the parameters that determine the pattern of genetic variation in a sample are the rates of migration and extinction/recolonization and the founding-propagule sizes for each sampled deme. Thus, in a large metapopulation, on the one hand, many interesting aspects of the biology of the species are tied together in the effective size of the collecting phase. On the other hand, we have a robust framework for investigating other phenomena, such as changes in effective size over time, within the context of a metapopulation. While many analytic results are possible, and some are given below, the model is also easy to program. Simulations are used here to show the contrasting effects of migration and extinction/recolonization on site-frequency distribution at polymorphic sites in a sample from a single deme.

THEORY

Large metapopulation model: Consider a population that is subdivided into a large number of local populations or demes. The total number of demes is D, and these are arbitrarily labeled 1 through D. Deme i has diploid size N_i, or, equivalently, haploid size $2N_i$. In either case $2N_i$ copies of each genetic locus reside within deme i. The results presented below apply in a straightforward way to haploid organisms and to diploid monoecious organisms with the additional assumption that migration and recolonization are gametic rather than zygotic. This leads to an apparent factor of two difference of terms involving k below, relative to results of previous authors, but this is not a meaningful difference. Nagylaki (1998) has shown that results for zygotic migration will be equivalent to those for gametic migration as long as the effective number of migrants each deme accepts each generation is not too small.

Each deme receives migrants from other demes in the population and is also subject to extinction/recolonization. If a deme goes extinct, it is recolonized immediately. Thus, there are no empty habitat patches in this model as there often are in ecological models of metapopulations; *e.g.*, see Hanski (1997). This assumption is unnecessary as long as the total number of extant demes remains constant from one generation to the next (Pannell and Charlesworth 1999). Deme i receives M_i (haploid) migrants each generation. That is, a fraction $M_i/(2N_i)$ of deme i is replaced by migrants every generation. The other portion, $1 - M_i/(2N_i)$, is derived from the previous generation of deme-i individuals. Reproduction within each deme occurs according to the Wright-Fisher model (Fisher 1930; Wright 1931). The parameter M_i is the scaled backward migration rate for deme i. Correspondingly, E_i is the scaled extinction/recolonization rate, so $E_i/(2N_i)$ is the per-generation probability that deme i goes extinct. If deme i goes extinct, it is recolonized by k_i individuals, which immediately restore the deme to its original size of $2N_i$ gene copies. This step also occurs according to the

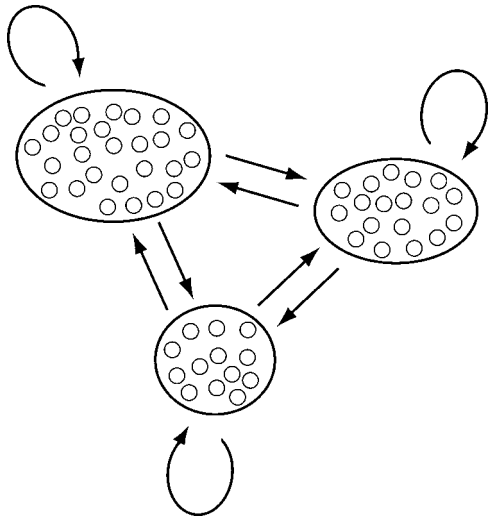


FIGURE 1.—An example, with $K = 3$ classes of demes, of the population structure assumed throughout this work. Within each of the three classes, or regions, there are many demes. Arrows depict the movement of lineages, by migration and/or extinction/recolonization, both within and among regions.

Wright-Fisher model. That is, the $2N_i$ descendants are obtained by sampling with replacement from the k_i colonists. It is important to note that the subscripts of N, M, E, and k refer to individual demes, not to the classes of demes introduced below.

The population is assumed to comprise K different types of demes, which may represent different geographic regions. Demes of type i make up a fraction β_i of all demes; thus, $\Sigma_{i=1}^{K}\beta_i = 1$. In addition, demes of type i receive a portion m_{ij} of their gametes via migration from demes of type j, where $1 \le j \le K$. In total, each deme of type i is a fraction $m_j = \Sigma_{j=1}^{K} m_{ij}$ of its gametes replaced by migrants every generation. Thus, $M_j/(2N_j) = m_i$ for every type i deme, j. Migrants into a type i deme might have come from another deme of type i. They may also originate in the same deme they migrate to, although the effect of this is negligible when the number of demes is large. Demes of type i go extinct with probability e_i each generation and are recolonized by a mixture of gametes from the different classes of demes in proportions e_{i1}/e_i, e_{i2}/e_i, . . . e_{iK}/e_i ($\Sigma_{j=1}^{K} e_{ij} = e_i$). When a lineage is a migrant or a colonist from a deme of type i, it is equally likely to have come from each of the $\beta_i D$ type i demes. The structure of this model is depicted in Figure 1.

Separation of timescales and the structure of genealogies: As in WAKELEY (1998, 1999, 2000, 2001), the results presented here will hold for metapopulations that are composed of a large number of demes. In particular, the sample size must be much smaller than the number of demes in the population ($n \ll D$). This does not appear to be an unrealistic assumption for some metapopulations in nature and is one that is commonly made in theoretical studies of metapopulations (HANKSI 1997). It leads to a separation of timescales in the ancestral process of a sample, which is similar to that found in studies of partial selfing (NORDBORG 1997, 1999; NORDBORG and DONNELLY 1997). A useful convergence theorem, derived in context of partial selfing, was found by MÖHLE (1998a). Consider the genealogy of a sample from such a population. The separation of timescales is a consequence of the fact that at any given time in the past, the overwhelming majority of demes in the population will not contain any lineages ancestral to the sample. Demes that do contain ancestral lineages are called occupied demes (WAKELEY 1999), and the fraction of these in the population is never $> n/D$. Two kinds of events differ vastly in rate. The first is migration and extinction/recolonization events in which the source deme is occupied. The second is coalescent events within demes and migration or extinction/recolonization events in which the source deme is unoccupied. Events of the second type dominate the history of the sample because they are approximately D times more likely than events of the first type.

Given this, it is necessary to distinguish sample configurations in which every lineage is in a separate deme from those in which at least one deme contains multiple lineages. When at least one deme contains multiple lineages, migration events and extinction/recolonization events will send lineages to unoccupied demes and coalescent events will join together lineages within demes until each remaining lineage is in a separate deme. This scattering phase takes a negligible amount of time compared to the waiting time to the next relevant event, which is a migration or extinction/recolonization event to an occupied deme. At least one event of this type must occur before another coalescent event can happen. In fact, if $n \ll D$, so many will occur that the movement of the lineages among unoccupied demes by migration and extinction/recolonization will reach a statistical equilibrium before two lineages will have the chance to coalesce. This is the essence of MÖHLE's (1998a) result and of the strong-migration limit (NAGYLAKI 1980). As shown below, the collecting phase is a Kingman-type coalescent process with a characteristic effective size. Thus, the structure of genealogies is twofold. First there is a one-time stochastic sample size adjustment, the scattering phase, which results in greater relatedness within than between demes; then the bulk of the history is spent in a collecting phase coalescent process. Figure 2 illustrates the structure of genealogies under this approximation for a sample from one deme.

Scattering phase: Consider the recent history of a sample $\boldsymbol{n} = (n_1, \ldots, n_d)$ taken from d different demes. The total sample size is $n = \Sigma_{j=1}^{d} n_i$. Following the genealogy of the sample from deme i, it will take on the order of $2N_i$ generations for the scattering phase to be complete, fewer if M_i and/or E_i are large. Let n_i' represent the number of lineages remaining of the sample n_i from deme i at the end of the scattering phase. When there

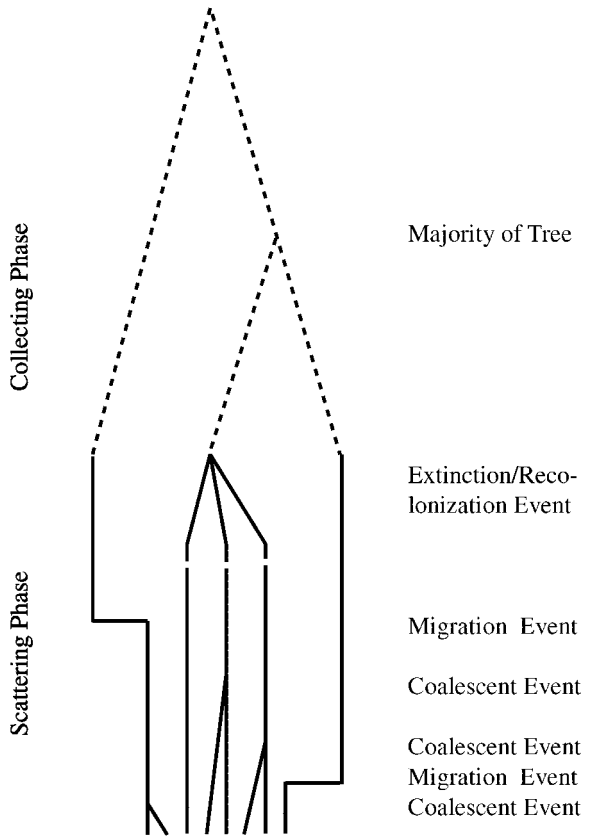


FIGURE 2.—An example of a genealogy of sample size eight from a single deme. In this case, during the scattering phase, there are two migration events (to some unoccupied demes that are not pictured) and then an extinction/recolonization event with $k = 2$ in which all of the lineages remaining in the deme are descended from a single common ancestor. The coalescent collecting phase of the three resulting lineages is shown above. The relative duration of the scattering phase is greatly exaggerated for purposes of illustration.

are j lineages in the deme, the scaled rates of migration, coalescence, and extinction/recolonization are jM_i, $j(j-1)/2$, and E_i, respectively. The value of n_i' will depend upon how many migration events occurred before the deme experiences an extinction/recolonization event and on the number of colonist-parents there are of the lineages that exist at the time of this event. The probability that the extinction event occurs when there are j lineages, and not before, is given by

$$P_{\mathrm{E}}(j|n_i) = \begin{cases} \dfrac{2E_i/j}{2M_i + (j-1) + 2E_i/j} \displaystyle\prod_{l=j+1}^{n_i} \dfrac{2M_i + (l-1)}{2M_i + (l-1) + 2E_i/l}, \\ \quad 2 \le j \le n_i \\ \displaystyle\prod_{l=2}^{n_i} \dfrac{2M_i + (l-1)}{2M_i + (l-1) + 2E_i/l}, \\ \quad j = 1. \end{cases} \tag{1}$$

The first case specifies that $n_i - j$ migration or coalescent events occur, which leaves j lineages in deme i, and then an extinction/recolonization event occurs. In the second case, the scattering phase for deme i ends, with each remaining lineage in a separate deme, before an extinction/recolonization event has occurred.

To know how many lineages remain at the end of the scattering phase, we must first distinguish histories that involve different numbers of migration events. The probability that x migration events occur and the extinction event occurs when there are j lineages is given by

$$P_{\mathrm{ME}}(x, j|n_i) = \begin{cases} \dfrac{2E_i/j}{2M_i + (j-1) + 2E_i/j} \dfrac{S_{n_i j}^{(x+1)} (2M_i)^x}{\prod_{l=j+1}^{n_i} 2M_i + (l-1) + 2E_i/l}, \\ \quad 2 \le j \le n_i \\ \dfrac{S_{n_i 1}^{(x+1)} (2M_i)^x}{\prod_{l=2}^{n_i} 2M_i + (l-1) + 2E_i/l}, \\ \quad j = 1 \end{cases} \tag{2}$$

in which

$$S_{n_i j}^{(x+1)} = \text{coefficient of } (2M_i)^x \quad \text{in} \prod_{l=j+1}^{n_i} [2M_i + (l-1)]. \tag{3}$$

These coefficients can be generated recursively,

$$S_{j,i}^{(l)} = (j-1) S_{j-1,i}^{(l)} + S_{j-1,i}^{(l-1)}, \tag{4}$$

starting with $S_{i,i}^{(1)} = 1$, and $S_{j,1}^{(l)}$ are unsigned Stirling numbers of the first kind. The source deme of each migrant is determined by the stochastic migration process described above.

Given that an extinction event occurs when there are j lineages remaining in deme i, the probability that these have y colonist-parents in the propagule of size k_i is given by

$$G_i[y|j] = \frac{\mathcal{S}_j^{(y)} \prod_{l=0}^{y-1} (k_i - l)}{k_i^j}. \tag{5}$$

Equation 5 is the usual backward Wright-Fisher process; see, for example, WATTERSON (1975). That is, the number of parents of the j lineages has the same distribution as the number of nonempty cells when j balls are thrown randomly into k_i boxes. The coefficients, $\mathcal{S}_j^{(y)}$, are Stirling numbers of the second kind. The source deme of each colonist-parent is determined by the stochastic extinction/recolonization process described above.

When an extinction/recolonization event occurs, as long as D is large, each lineage will have a different source deme and the scattering phase will end for deme i. Thus, this model is a general version of SLATKIN's (1977) migrant pool model. At the other extreme is SLATKIN's (1977) propagule-pool model in which all y lineages in (5) would have the same source deme, chosen randomly according to some probability function. If this were the case, the scattering phase would continue, but with the scaled coalescent, migration, and extinction/recolonization rates of the source deme. If there was no migration, the propagule-pool model would always give $n_i' = 1$. WADE and MCCAULEY (1988) proposed an intermediate model, in which a fraction,

ϕ, of the lineages would follow the propagule-pool model and the other $1 - \phi$ would follow the migrant pool model. These and more complicated schemes could be modeled within the present framework but are not pursued here. When an extinction/recolonization event occurs the scattering phase is over for the sample from that deme. If k_i is equal to one or if the rate of extinction/recolonization is low, the present results will be identical to those of a propagule-pool model.

If there are x migration events and y colonist-parent lineages in the sample from deme i, then the scattering phase for deme i ends with $n_i' = x + y$ lineages each in separate demes. The probability function for n_i' is

$$P[\boldsymbol{n}_i'|\boldsymbol{n}_i] = \sum_{j=1}^{n} \sum_{x=0}^{n-j} P_{\mathrm{ME}}(x, j|n_i)\, G_i[n_i' - x|j], \qquad (6)$$

where we define $G_i[y|j]$ to be equal to zero if y is <1 or $>j$. Because events occur independently in different demes, the joint probability function of all the n_i' is given by

$$P[\boldsymbol{n}'|\boldsymbol{n}] = \prod_{i=1}^{d} P[n_i'|n_i]. \qquad (7)$$

The collecting phase of the history then begins with $n' = \Sigma_{i=1}^{d} n_i'$ lineages, each in separate demes.

Collecting phase: The distribution among deme types of the n' lineages that enter the collecting phase will depend on the particular outcome of the scattering phase for each deme's sample. Let r_i be the number of lineages that are in type i demes. The vector $(r_1, \ldots, r_K)$ then denotes the configuration of the lineages among the different types of demes. The total number of lineages is equal to $r = \Sigma_{i=1}^{K} r_i$ and at the start of the collecting phase we have $r = n'$. Here it is shown that the time to a coalescent event does not depend on the starting value of $(r_1, \ldots, r_K)$ and is exponentially distributed as in Kingman's coalescent. First, as in Wakeley (2001), the time until two lineages are in the same deme is shown to be exponentially distributed. The coalescent result follows from this and the fact that the number of times two lineages must be in the same deme before a common ancestor event occurs is geometrically distributed.

Note that, from the perspective of a single lineage in a singly occupied deme, a migration event and an extinction/recolonization event are indistinguishable. Both simply move the lineage to another deme. Therefore, it is sufficient during the collecting phase to consider the combined effect of migration and extinction/recolonization: $h_{ij} = m_{ij} + e_{ij}$. It is assumed that h_{ij} is small, on the order of the reciprocal of the deme size. Thus, squared and higher-order terms in h_{ij}, which represent the movement of two or more of the lineages in a single generation, will be ignored. Note also that here there is no difference between migrant-pool and propagule-pool recolonization.

Looking back to the immediately previous generation, there are two kinds of events: changes in the configuration, $(r_1, \ldots, r_K)$, and the movement of a lineage into an occupied deme. Events of the first kind occur with probability

$$P\{(\ldots, r_i - 1, \ldots, r_j + 1, \ldots) \rightarrow (\ldots, r_i, \ldots, r_j, \ldots)\} = (r_j + 1)\, h_{ji}. \qquad (8)$$

The movement of a lineage into an occupied deme of type i occurs with probability

$$b_{i,(r_1,\ldots,r_K)} = r_i h_{ii}\left(\frac{r_i - 1}{\beta_i D}\right) + \sum_{j:\, j\neq i} r_j h_{ji}\left(\frac{r_i}{\beta_i D}\right). \qquad (9)$$

Clearly the first kind of event is much more likely to occur when D is large relative to r. The probability that the sample configuration is unchanged is equal to

$$P\{(r_1, \ldots, r_K) \rightarrow (r_1, \ldots, r_K)\} = 1 - \sum_{i=1}^{K} r_i \sum_{j:\, j\neq i} h_{ij} - \sum_{i=1}^{K} b_{i,(r_1,\ldots,r_K)} \qquad (10)$$

$$\approx 1 - \sum_{i=1}^{K} r_i \sum_{j:\, j\neq i} h_{ij}. \qquad (11)$$

As in Wakeley (2001), the essence of the separation of timescales is that, when $r \ll D$, an equilibrium for $(r_1, \ldots, r_K)$ is reached with respect to (8) and (11) before any event of the type in (9) occurs. Then, the waiting time to a movement event that places two lineages into the same type i deme is the average of (9) over the stationary distribution of $(r_1, \ldots, r_K)$. Möhle (1998a) provided a convergence theorem for processes such as this, which is used implicitly below.

Consider first the movement of just one lineage among demes in the population. This is determined by the matrix $\boldsymbol{Q}$, which has off-diagonal entries $q_{ij} = h_{ij}$. The diagonal entries are $q_{ii} = 1 - \Sigma_{j:j\neq i} h_{ij}$, which it is important to note are not equal to h_{ii} defined above. The h_{ii} do not directly affect the equilibrium configuration because such moves do not take the lineage into a different class of demes. Standard matrix theory shows that as long as the matrix $\boldsymbol{Q}$ is ergodic, *i.e.*, irreducible and aperiodic, a stationary distribution will exist. As Nagylaki (1998) notes, ergodicity in itself probably does not rule out very many plausible biological scenarios. Ergodicity requires only that lineages can eventually get from any deme type to any other and that lineages have some chance of staying in their current type of deme. If f_i is the equilibrium probability that a lineage is in a deme of type i, we have

$$f_i = \sum_{j=1}^{K} f_j q_{ji} \qquad (12)$$

or, equivalently,

$$f_i \sum_{j:\, j\neq i} q_{ij} = \sum_{j:\, j\neq i} f_j q_{ji}\,, \qquad (13)$$

where we may assume $0 < f_i < 1$ and $\Sigma_{i=1}^{K} f_i = 1$. The

quantity f_i can be interpreted as the average relative amount of time a lineage spends in demes of type i.

The stationary distribution of the full configuration $(r_1, \ldots, r_K)$ is multinomial:

$$p(r_1, \ldots, r_K) = \frac{r!}{r_1! \ldots r_K!} f_1^{r_1} \ldots f_K^{r_K}. \quad (14)$$

This is proved by induction over time. Using (8) and (11), we have

$$p(r_1, \ldots, r_K) = \sum_{i=1}^{K} \sum_{j,\, j \neq i} p(\ldots, r_i - 1, \ldots, r_j + 1, \ldots)(r_j + 1) q_{ji} + p(r_1, \ldots, r_K)\left(1 - \sum_{i=1}^{K} r_i \sum_{j,\, j \neq i} q_{ij}\right) \quad (15)$$

or equivalently,

$$p(r_1, \ldots, r_K) \sum_{i=1}^{K} r_i \sum_{j,\, j \neq i} q_{ij} = \sum_{i=1}^{K} \sum_{j,\, j \neq i} p(\ldots, r_i - 1, \ldots, r_j + 1, \ldots)(r_j + 1) q_{ji}. \quad (16)$$

If the stationary distribution of $(r_1, \ldots, r_K)$ is given by (14), then

$$p(r_1, \ldots, r_K) \sum_{i=1}^{K} r_i \sum_{j,\, j \neq i} q_{ij} = \sum_{i=1}^{K} \sum_{j,\, j \neq i} p(r_1, \ldots, r_K) \frac{r_i f_j}{(r_j + 1) f_i} (r_j + 1) q_{ji} \quad (17)$$

$$= p(r_1, \ldots, r_K) \sum_{i=1}^{K} \frac{r_i}{f_i} \sum_{j,\, j \neq i} f_j q_{ji}, \quad (18)$$

which is true because the f_i satisfy (13). The stationary distribution (14) is unique since the Markov chain is ergodic and has a finite number of states.

The total rate of events that put two lineages together in a deme of type i is the average of (9) over the stationary distribution of $(r_1, \ldots, r_K)$, that is,

$$g_{r,i} = \sum_{\{(r_1, \ldots, r_K) : \sum_{i=1}^{K} r_i = r\}} p(r_1, \ldots, r_K)\, b_{i,(r_1, \ldots, r_K)}. \quad (19)$$

Equation 19 is just the expectation of $b_{i,(r_1, \ldots, r_K)}$ over the multinomial distribution (14). Using (13) and the fact that $q_{ij} = h_{ij}$ for $j \neq i$, we have

$$g_{r,i} = \binom{r}{2} \frac{2 f_i^2 h_i}{\beta_i D} \quad (20)$$

in which $h_i = \sum_{j=1}^{K} h_{ij}$. The total rate of events that put two lineages into the same deme, regardless of the type, is the sum of (20) over all types of demes: $g_r = \sum_{i=1}^{K} g_{r,i}$. Given that such an event occurs, the probability that the two lineages are in a deme of type i is equal to $g_{r,i}/g_r$. Then, once two lineages are in the same deme, they either have a common ancestor or they again wind up in separate demes, either by migration or through an extinction/recolonization event. If they wind up in separate demes, there will be another exponentially distributed waiting time with rate g_r before two lineages are in one deme and again have a chance to coalesce.

Because each of the $\beta_i D$ demes of type i is equally likely to be the one that contains the two lineages, the overall chance that the two will have a common ancestor is equal to

$$\left\langle \frac{1 + E/k}{2M + 1 + E} \right\rangle_i = \frac{1}{D\beta_i} \sum_{\{j,\, j \in \Omega_i\}} \frac{1 + E_j/k_j}{2M_j + 1 + E_j}, \quad (21)$$

where Ω_i is the set of labels of the $D\beta_i$ demes of type i. As in Wakeley (2001), the number of movement events to occupied demes that must occur before a common ancestor event happens is geometrically distributed with probability of success equal to the average of (21) over the distribution $g_{r,i}/g_r$. Because the waiting time between these events is exponential with rate g_r, it follows (Wakeley 1999) that the time to coalescent event among the r lineages is exponentially distributed with rate

$$\binom{r}{2} \frac{2}{D} \sum_{i=1}^{K} \frac{f_i^2 h_i}{\beta_i} \left\langle \frac{1 + E/k}{2M + 1 + E} \right\rangle_i. \quad (22)$$

When a coalescent event occurs, the number of lineages decreases by one and the process continues.

This shows that the collecting phase is a Kingman-type coalescent process and is thus independent of the starting distribution of lineages among deme types, $(r_1, \ldots, r_K)$. The effective size of this coalescent process is given by

$$\frac{1}{2N_e} = \frac{2}{D} \sum_{i=1}^{K} \frac{f_i^2 h_i}{\beta_i} \left\langle \frac{1 + E/k}{2M + 1 + E} \right\rangle_i. \quad (23)$$

An equation like (23) can provide a framework for understanding the determinants of the effective population size. Wakeley (2001) discusses the effects of different factors in the case of a migration-only model. It is important to note that (23) determines the rate in a coalescent process that occurs in the history of every sample. This is different than the traditional effective sizes, which are descriptions of the equilibrium behavior of genetic drift in the population. However, (23) is by definition an inbreeding effective size and is essentially the same as the various effective metapopulation sizes that others have discussed in detail; for example, see Whitlock and Barton (1997). For purposes here, the significance of (23) is twofold. First, it will not be possible to differentiate among many different parameters of the model because they are buried in the composite parameter, N_e. Second, the collecting-phase coalescent result holds for many specific population structures.

Analytic predictions about DNA sequence polymorphisms: Because the collecting phase is a coalescent process, a natural way to incorporate neutral mutations is to define $\theta = 4N_e u$, where N_e is given by (23) and u is the neutral mutation rate at some genetic locus. With θ defined in this way, the history of one particular kind of sample $(n_1 = 1, n_2 = 1, \ldots, n_d = 1)$ will conform to the standard neutral coalescent model. All the usual coalescent results, for example, those found in Tavaré (1984), will apply directly to this sample when $\theta = 4N_e u$.

Predictions about levels of genetic variation for other samples will have to be averaged over the possible outcomes of the scattering phase and weighted by their probabilities as given in (7). If $S(\boldsymbol{n})$ is the number of segregating sites in the multideme sample, $\boldsymbol{n} = (n_1, \ldots, n_d)$, then

$$P[S(\boldsymbol{n}) = i] = \sum_{\boldsymbol{n}'} P[S(\boldsymbol{n}') = i]P[\boldsymbol{n}'|\boldsymbol{n}]. \qquad (24)$$

Under Watterson's (1975) infinite sites mutation model, $P[S(\boldsymbol{n}') = i]$ may be given by Tavaré's (1984) equation (9.5). Summing over all possible values of i gives the corresponding equation for the expectation of $S(\boldsymbol{n})$,

$$E[S(\boldsymbol{n})] = \sum_{\boldsymbol{n}'} E[S(\boldsymbol{n}')]P[\boldsymbol{n}'|\boldsymbol{n}], \qquad (25)$$

in which $E[S(\boldsymbol{n}')] = \theta\Sigma_{i=1}^{n'-1} 1/i$ (Watterson 1975). In the case of migration only, integral representations of $P[S(\boldsymbol{n}')]$ and $E[S(\boldsymbol{n}')]$ allow the sums in (24) and (25) to be evaluated, resulting in somewhat simpler expressions (Wakeley 2001), but this does not appear possible here.

The most basic prediction of this model, or of any subdivided population model, is that levels of polymorphism will be higher among than within demes. The simplest case, two sequences sampled either from the same deme or from two different demes, illustrates this point. If the two are from different demes, the expected number of pairwise differences will be equal to θ, identically for any pair of demes. A randomly chosen pair will have this same expectation because the chance of randomly sampling the same deme twice will be low when the number of demes is large. The expected number of differences between a pair of sequences from deme i will be equal to this, θ, times the probability that the scattering phase for this sample ends with two lineages. Call these expected values π_T and π_i, respectively. We can define an inbreeding coefficient for the deme,

$$F_i = \frac{\pi_T - \pi_i}{\pi_T} \qquad (26)$$

$$= \frac{1 + E_i/k_i}{2M_i + 1 + E_i}, \qquad (27)$$

which is simply the probability that the two lineages coalesce during the scattering phase. The inbreeding coefficient for the deme will be small when the migration rate is large or when the extinction/recolonization rate and the propagule size are both large. It will be large when the rates of migration and extinction/recolonization are both small or when the extinction/recolonization rate is large and the propagule size is small. It is important to note that here θ is assumed to remain constant and that these differences in F_i represent the possible differences among sample demes. If θ changes as well, *i.e.*, if demes do not differ in their characteristics, then the conclusions are different (Pannell and Charlesworth 1999; Wakeley 2001).

Equations 24 and 25 provide results for arbitrary samples. For example, Figure 3 plots (25) for a sample of five sequences from each of two different demes over a broad range of migration and extinction/recolonization rates. Shown are results for two different propagule sizes: $k = 1$ in Figure 3a and $k = 10$ in Figure 3b. In both cases, $\theta = 10$. As expected, the model predicts that samples from demes with larger backward migration rates will be more polymorphic. If within each deme all five lineages share a common ancestor by the end of the scattering phase, the collecting phase begins with just two lineages and the expected number of segregating sites will be equal to 10.0. At the other extreme, if, for example, migration is very frequent, the scattering phase could end with 10 sequences all in different demes. The collecting phase would then begin with 20 lineages, and the expected number of segregating sites will be equal to 28.3. These extremes are very nearly realized in Figure 3. In contrast to migration, the effect of changes in the rates of extinction/recolonization in the two demes depends on the propagule size. When the propagule size is small (Figure 3a), increases in the rate of extinction/recolonization make the expected number of segregating sites smaller, whereas when the propagule size is large (Figure 3b), increases in extinction/recolonization rates have a similar effect to increase in the migration rates, which is to increase levels of polymorphism. This is just restatement, within the framework of the coalescent, of Slatkin's (1977) conclusions about the dual roles of extinction/recolonization.

As in Wakeley (1999) it may be possible to find analytic expressions for the expected number of sites segregating at particular joint frequencies among the sampled demes. In addition, because the collecting phase is a coalescent, it is relatively straightforward to incorporate changes in effective population size over time. Again, existing expressions found for changing population sizes in the context of a Kingman-type coalescent, such as those in Slatkin and Hudson (1991) and Hey and Harris (1999), will apply directly to the sample where $n = d$ and can be averaged over (7) for other samples. It should also be possible to model diverging species, each of which conforms to the present metapopulation model, as in Wakeley (2000). Instead of pursuing these ideas any further here, the next section describes how genealogies can be simulated easily so that these and other questions can be addressed computationally.

SIMULATIONS

The coalescent process in the large-number-of-demes metapopulation is simulated as follows. First, the scattering phase is carried out for each sampled deme. This is done by simulating a series of coalescent and migration events, and possibly an extinction/recolonization event,

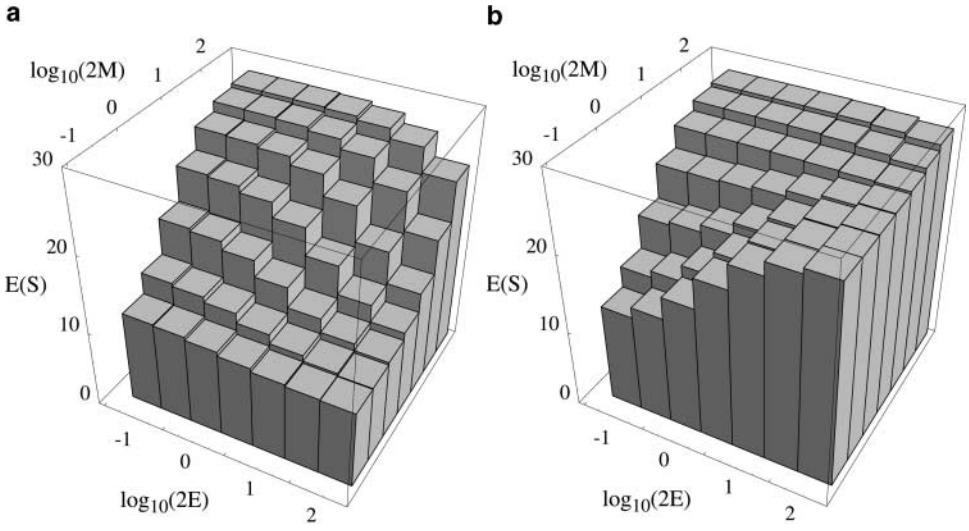


FIGURE 3.—The expected number of segregating sites in a sample of five sequences from each of two demes, plotted as a function of the migration and extinction/recolonization rates. These rates are assumed to be the same for both demes. In a the propagule size is equal to 1 and in b it is equal to 10.

for each deme according to the relative rates specified above Equation 1. The result for deme i is a number of lineages, n_i', to enter the collecting phase, and for each lineage the number of descendants it has in the sample, or its "size." For example, the sizes of the lineages in Figure 2 are two, five, and one. The sizes of lineages are required to simulate anything beyond the most simple properties of the model. When there is no extinction and recolonization, WAKELEY (1999) showed that the sizes of lineages for a single-deme sample have the same probability distribution as the allele counts in the EWENS (1972) sampling formula, with infinite-allele mutation replaced by infinite-deme migration. If, on the other hand, the rate of extinction/recolonization is very high relative to migration and coalescence, the distribution will be identical to the distribution of the occupancy numbers when n_i balls are thrown randomly into k_i boxes; see (5). With intermediate rates of extinction/recolonization, the size distribution will be some kind of mixture of these two extremes. The results presented below show that these two extremes produce very different patterns of polymorphism.

When this instantaneous scattering phase has been completed for every deme, the $n' = \Sigma_{i=1}^{d} n_i'$ lineages are thrown together into the usual coalescent process; for example, see the routine make_tree in HUDSON (1990). After a tree is generated, a Poisson-distributed number of mutations, with mean $T_{\text{tot}}\theta/2$, are placed randomly on its branches, where T_{tot} is the total length of the tree measured in units of $2N_e$ generations. According to the infinite-sites mutation model (WATTERSON 1975), each of these mutations produces a polymorphic site. The infinite sites mutation model is a good approximation for mutations in DNA sequences as long as the mutation rate per site is small. It is important to note that this could be replaced by any other neutral mutation model if needed. There is no recombination in the program, but it could also be included. The source code of this C program is available upon request.

One aspect of DNA sequence data that has received a lot of attention in recent years is the distribution of allele frequencies in a sample. These form the basis of many tests of the standard neutral coalescent model, such as TAJIMA's (1989) D, so it is of interest to know what other models predict about these "site frequencies." Figure 4 shows the predicted site-frequency distributions under different regimes of migration and extinction/recolonization. These are the "unfolded" site frequency distributions, that is, assuming it is known which is the ancestral and which is the mutant base at each polymorphic site. Figure 4, a–c, shows simulation results for a sample of 10 sequences from a single deme and are averaged over 1 million replicates. Figure 4a shows that as the migration rate becomes small when there is no extinction/recolonization, the site-frequency distribution becomes U-shaped. When migration is infrequent, typically only one or zero migration events will occur during the scattering phase of the sample. The genealogies with a single scattering-phase migration event will have a long internal branch, and it is most likely that this branch separates a single sequence from the rest of the sample. This U-shaped distribution under low migration appears to be a general property of subdivision with migration because it also occurs in a continuous-habitat model (J. F. WILKINS, unpublished results) and when the number of demes in the population is small (results not shown).

Figure 4b shows the site-frequency distribution expected when there is no migration among demes. In this case, as the rate of extinction/recolonization increases, the distribution develops a mode for $i > 1$. This results from the fact that $k = 2$ in the simulations that produced Figure 4b. As E increases, the scattering phase becomes equivalent to throwing 10 balls into two boxes,

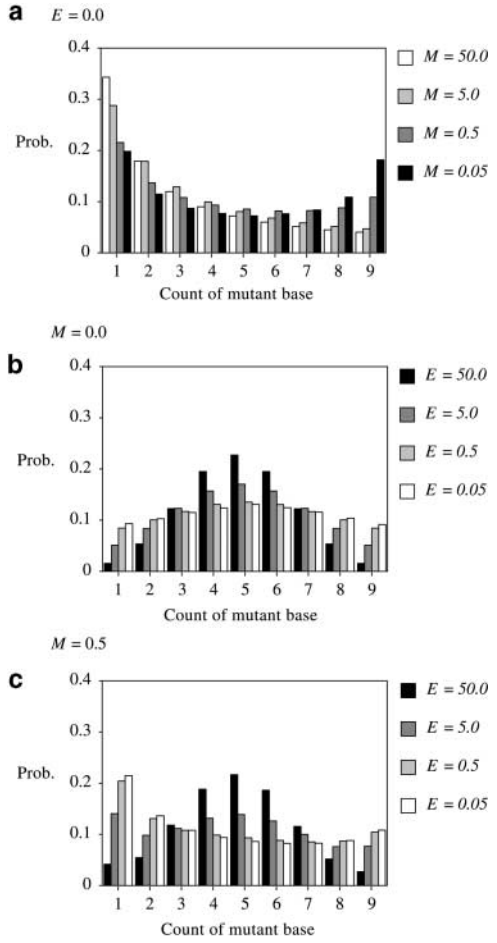


FIGURE 4.—The expected site frequency distribution in a sample of 10 sequences from a single population, as a function of the rates of migration and extinction/recolonization. The vertical bars represent the proportion of polymorphic sites expected to be segregating at various frequencies in the sample. In b and c the propagule size is equal to 2.

the result being a binomial ($n = 10$, $p = 0.5$) distribution for the "sizes" of the two collecting-phase lineages. For other values of k, the sizes will be multinomial, and the mode will be for site frequencies around n/k. While a U-shaped distribution of unfolded site frequencies can result from positive Darwinian selection, *e.g.*, see FAY and WU (2000) and KIM and STEPHAN (2000), this appears to be the first report of exchangeable lineages of a mode in the site-frequency distribution for anything other than the singleton-polymorphism ($i = 1$) class. Figure 4c shows that when both migration and extinction/recolonization occur, a mixture of the two above cases can result, and this can lead to a trimodal distribution of unfolded site frequencies.

DISCUSSION

The above work shows that samples of genetic data from a metapopulation that contains a large number of demes will reveal certain aspects of population structure but not others. Genetic variation will be shaped by the overall rates of migration, extinction/recolonization, and the number of colonists for the sampled demes only and by the scaled mutation parameter, θ, which hides all other aspects of demography. In particular, the geographic structure represented by the rates of movement of lineages, m_{ij} and e_{ij}, among the different classes of demes will not be discernible if the number of demes in each class is large. For this result to hold, the only restrictions on the movement matrix, $\boldsymbol{Q}$, are that lineages can get from any deme class to any other given enough time and that there is some chance a lineage does not switch deme classes in a single generation, *i.e.*, that $\boldsymbol{Q}$ is ergodic. Even highly structured populations, such as those with stepping-stone movement among deme classes (KIMURA and WEISS 1964), meet these simple criteria.

Analysis shows that samples from demes with higher rates of migration will be more polymorphic than samples from demes with lower rates of migration. Demes with higher rates of extinction/recolonization will be more polymorphic if the number of colonists is not small; otherwise they will be less polymorphic. These effects are shown in Figure 3. If demes vary in their characteristics, it is possible to observe a great reduction of variation within sampled demes even if the variability between demes is high. This may explain the recent observation of such a pattern, by LIU *et al.* (1998), in samples from populations of the plant Leavenworthia. In a large metapopulation, no structure will be visible in between-deme comparisons. This is analogous to the (total) lack of geographic structure in data under the strong-migration limit of NAGYLAKI (1980) and NOTOHARA (1993). However, the source of the limit here is that there are a large number of demes within each class, while the values of M and E per deme are not necessarily large.

It is interesting to note that a kind of isolation by distance could be observed under the present model. To see how this might develop, imagine a species with two types of demes: small and large. For simplicity, assume that every deme accepts the same fraction of migrants and has the same chance of extinction each generation. Thus, we have $m_{ij} = m$ and $e_{ij} = e$ for all i and j. The smaller demes will have smaller values of M and E than the larger demes and will thus be less polymorphic. For example, if the smaller demes have $M = E = 0.5$ and the larger demes are 10 times larger ($M = E = 5.0$), and if $\theta = 10.0$, then the average number of pairwise differences within small demes will be 4.94, and the average number of pairwise differences within large demes will be 7.74. These values are obtained from (25) with the additional assumption that $k = 2$ for all demes. The average number of pairwise differences between demes will be equal to θ, or 10.0 in this case. Then all that is needed to create a pattern of isolation by distance

would be for the larger demes to occupy larger geographic ranges and for the geographic distances between demes to be larger than the average geographic distance of within-larger-deme pairwise comparisons. Thus, while the model loses some of its structure in the large-number-of-demes limit, a correlation between geographic and genetic distances can develop under, arguably, reasonable biological conditions. This result does not depend strongly on the particular values of M, E, and k used here; *e.g.*, if $E = 0.0$ for all demes, the only differences are that the within-small-deme average becomes 4.90 and the within-large-deme average becomes 9.0.

It should be possible to estimate the relative contributions of migration *vs.* extinction/recolonization using DNA sequence data, due to the dramatically different effect these have on the frequencies of the segregating bases at polymorphic sites (Figure 4). The migration site-frequency distribution can be U-shaped, and this could be tested using Fay and Wu's (2000) statistic, which was designed to detect signatures of positive selection. In contrast, the extinction site-frequency distribution can have a mode for middle-frequency polymorphisms, which is quite unusual for exchangeable coalescent models. Note, however, that the colonization events in a metapopulation violate a fundamental premise of the coalescent: that no more than one common ancestor event can happen in a single generation. Thus, we should expect to see an identical middle-frequency mode in the site-frequency distribution in samples from species that recently experienced brief severe bottleneck events. Coalescent models of bottleneck events, in which θ becomes small for some period of time, never predict a nonsingleton mode in the site-frequency distribution. Instead, the distribution simply flattens out as a coalescent bottleneck becomes more severe. Thus, it will sometimes be inadequate to use Kingman's coalescent to model population bottlenecks.

A noteworthy consequence of there being a large number of demes in a population is that if the population mutation parameter, θ, is finite, then the demic mutation rates, $4N_i u$, will be vanishingly small. Conversely, if the demic mutation rates are not small, so that mutations occur during the scattering phase, then the population mutation rate will be infinite. Metapopulation studies often presume the former (Hanksi 1997). In fact, the latter can often be rejected using data because it predicts complete saturation (*i.e.*, multiple mutations per site) in among-deme samples. When the latter case does hold, then the approach of Slatkin (1982), which assumes an infinite-allele mutation model, could be used. In this model, migration and extinction/recolonization act like mutation, because every lineage that enters the collecting phase is guaranteed to represent a unique allele.

Recombination could easily be included in this metapopulation coalescent model. As with mutation, there will be no recombination events during the scattering phase if the total population recombination rate is finite. That is, in the same way mutations are modeled we assume that the population recombination rate, $R = 4N_e r$, where r is the rate per locus per generation and N_e is given by (23), is finite. Similar to the case of partial selfing (Nordborg 2000), the observable recombination rate will be smaller than this actual recombination rate. When a recombination event occurs in some collecting-phase lineage, it will split it into two lineages. The two lineages will necessarily be in the same deme, so there is some chance that they coalesce and the event is erased. The relationship between the actual and observable recombination rates in this case is

$$R_{\text{obs}} = R\sum_{i=1}^{K} f_i \left\langle \frac{1 + E/k}{2M + 1 + E} \right\rangle_i. \tag{28}$$

In words, relative to mutation, the recombination rate is decreased by a factor that is equal to the average chance that two ancestral lineages in the same deme either coalesce before one of them migrates or have a common ancestor during an extinction/recolonization event. This reduction in R will cause linkage disequilibrium among sites to be elevated in addition to that accrued more directly during the scattering phase via within-deme coalescent events.

The metapopulation model presented here has so far not addressed some well-known features of such species. The assumption that a recolonized population regains its previous size in a single generation is not realistic. It is known that delayed or slower growth can change some results (Whitlock 1992; Ingvarsson 1997). The effect here would be to increase the number of coalescent events during the scattering phase. Another point is that there is probably an inverse relationship between the size of a deme and its probability of going extinct (Foley 1997). In fact, the present model allows for this possibility already. One would simply have to add the assumption that deme classes with higher rates of extinction/recolonization also contained smaller-sized demes. Finally, one of the major concerns in metapopulation studies is whether the population under study is at equilibrium or not. The present work shows that if there are a large number of demes in the population, changes in demography over time will be manifest simply as changes in the effective size of the population. If the changes occur on a longer timescale than the scattering phase, they will affect only the collecting phase and can be modeled as a change in effective size of that coalescent process. Consequently, using samples of genetic data it will be impossible to distinguish between changes in population number and changes in the rates and patterns of migration and extinction/recolonization as explanations of variable effective size over time.

We thank Martin Möhle for continuing helpful discussions of his convergence results and two anonymous reviewers for helpful comments. Nicolas Aliacar participated in this project while visiting the lab for his research training period during April–July 2000 and was supported by l'Ecole Polytechnique, Paris. This work was supported by grant DEB-9815367 from the National Science Foundation.

LITERATURE CITED

Aldous, D. J., 1985 Exchangeability and related topics, pp. 1–198 in *École d'Été de Probabilités de Saint-Flour XII—1983*, Vol. 117 of *Lecture Notes in Mathematics*, edited by A. Dold and B. Eckmann. Springer-Verlag, Berlin.

Bahlo, M., and R. C. Griffiths, 2000 Inference from gene trees in a subdivided population. Theor. Popul. Biol. **57:** 79–95.

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.

Donnelly, P., 1984 The transient behaviour of the Moran model in population genetics. Math. Proc. Camb. Philos. Soc. **95:** 349–358.

Donnelly, P., and S. Tavaré, 1995 Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. **29:** 401–421.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3:** 87–112.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

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

Foley, P., 1997 Extinction models for local populations, pp. 215–246 in *Metapopulation Biology: Ecology, Genetics, and Evolution*, edited by I. Hanski and M. E. Gilpin. Academic Press, San Diego.

Hanski, I., 1997 Metapopulation dynamics: from concepts and observations to predictive models, pp. 69–91 in *Metapopulation Biology: Ecology, Genetics, and Evolution*, edited by I. Hanski and M. E. Gilpin. Academic Press, San Diego.

Hanski, I., 1998 Metapopulation dynamics. Nature **396:** 41–49.

Hanksi, I., and M. E. Gilpin, 1997 *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic Press, San Diego.

Hey, J., and E. Harris, 1999 Population bottlenecks and patterns of human polymorphism. Mol. Biol. Evol. **16:** 1423–1426.

Hudson, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, Vol. 7, edited by D. J. Futuyma and J. Antonivics. Oxford University Press, Oxford.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

Ingvarsson, P. K., 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, 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.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kimura, M., and G. H. Weiss, 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49:** 561–576.

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.

Kingman, J. F. C., 1982c On the genealogy of large populations. J. Appl. Probab. **19A:** 27–43.

Latter, B. D. H., 1973 The island model of population differentiation: a general solution. Genetics **73:** 147–157.

Levins, R., 1968 *Evolution in Changing Environments*. Princeton University Press, Princeton, NJ.

Levins, R., 1969 Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull. Entomol. Soc. Am. **15:** 237–240.

Liu, F., L. Zhang and D. Charlesworth, 1998 Genetic diversity in *Leavenworthia* populations with different inbreeding levels. Proc. R. Soc. Lond. Ser. B **265:** 293–301.

Maruyama, T., 1970 Effective number of alleles in a subdivided population. Theor. Popul. Biol. **1:** 273–306.

Maruyama, T., and M. Kimura, 1980 Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. Proc. Natl. Acad. Sci. USA **77:** 6710–6714.

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

Möhle, M., 1998b Robustness results for the coalescent. J. Appl. Probab. **35:** 438–447.

Nagylaki, T., 1980 The strong-migration limit in geographically structured populations. J. Math. Biol. **9:** 101–114.

Nagylaki, T., 1998 The expected number of heterozygous sites in a subdivided population. Genetics **149:** 1599–1604.

Nagylaki, T., 2000 Geographical invariance and the strong-migration limit in subdivided populations. J. Math. Biol. **41:** 123–142.

Neuhauser, C., and S. M. Krone, 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, pp. 56–76 in *Statistics in Molecular Biology and Genetics*, Vol. 33 of *IMS Lecture Notes—Monograph Series*, edited by F. Seillier-Moiseiwitsch. Institute of Mathematical Statistics, Hayward, CA.

Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial selfing. Genetics **154:** 923–929.

Nordborg, M., and P. Donnelly, 1997 The coalescent process with selfing. Genetics **146:** 1185–1195.

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 The effect of recombination on background selection. Genet. Res. **67:** 159–174.

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

Notohara, M., 1993 The strong migration limit for the genealogical process in geographically structured populations. J. Math. Biol. **31:** 115–122.

Notohara, M., 1997 The number of segregating sites in a sample of DNA sequences from a geographically structured population. J. Math. Biol. **36:** 188–200.

Pannell, J. R., and B. Charlesworth, 1999 Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution **53:** 664–676.

Pannell, J. R., and B. Charlesworth, 2000 Effects of metapopulation processes on measures of genetic diversity. Philos. Trans. R. Soc. Lond. Ser. B **355:** 1851–1864.

Slatkin, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. Theor. Popul. Biol. **12:** 253–262.

Slatkin, M., 1982 Testing neutrality in a subdivided population. Genetics **100:** 533–545.

Slatkin, M., 1985 Gene flow in natural populations. Annu. Rev. Ecol. Syst. **16:** 393–430.

Slatkin, M., 1996 Gene genealogies within mutant allelic classes. Genetics **143:** 579–587.

Slatkin, M., and R. R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129:** 555–562.

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tavaré, S., 1984 Lines-of-descent and genealogical processes, and their application in population genetic models. Theor. Popul. Biol. **26:** 119–164.

Wade, M. J., and D. E. McCauley, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution **42:** 995–1005.

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., 2000 The effect of population subdivision on the genetic divergence of populations and species. Evolution **54:** 1092–1101.

Wakeley, J., 2001 The coalescent in an island model of population subdivision with variation among demes. Theor. Popul. Biol. **59:** 133–144.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Whitlock, M. C., 1992 Temporal fluctuations in demographic parameters and the genetic variance among populations. Evolution **46:** 608–615.

Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics **146:** 427–441.

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., 1940 Breeding structure of populations in relation to speciation. Am. Nat. **74:** 232–248.

Wright, S., 1943 Isolation by distance. Genetics **28:** 114–138.

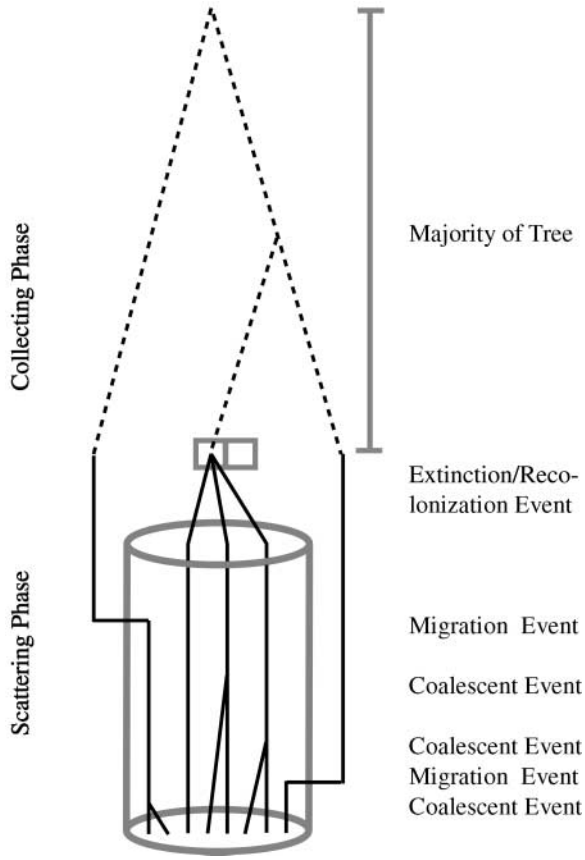
Communicating editor: D. Charlesworth

CORRIGENDA

In the article by Fernando Pardo-Manuel de Villena and Carmen Sapienza (Genetics **159:** 1179–1189) entitled "Female Meiosis Drives Karyotypic Evolution in Mammals," it should be noted that the authors made equal contributions to the work.

In the article by Mikkel H. Schierup, Barbara K. Mable, Philip Awadalla and Deborah Charlesworth (Genetics **158:** 387–399) entitled "Identification and Characterization of a Polymorphic Receptor Kinase Gene Linked to the Self-Incompatibility Locus of *Arabidopsis lyrata*," on page 389, line 10 of the paragraph entitled "**Identification and sequencing of *Aly*13 subtypes:**" the SLGR primer sequence is: ATCTGACATAAAGATCTTGACC.

In the article by John Wakeley and Nicolas Aliacar (Genetics **159:** 893–905) entitled "Gene Genealogies in a Metapopulation," Figure 2 should appear as it does here:



In the article by Rongling Wu, Chang-Xing Ma and George Casella (Genetics **160:** 779–792) entitled "Joint Linkage and Linkage Disequilibrium Mapping of Quantitative Trait Loci in Natural Populations," the following is added to the Acknowledgements:

This study is partially supported by the Outstanding Young Investigator Award of the National Natural Science Foundation of China (no. 30128017). This manuscript was approved for publication as journal series number R-07961 by the Florida Agricultural Experiment Station.

The following references are added to the Literature Cited:

Jiang, C. J., and Z-B. Zeng, 1997 Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetics **101:** 47–58.

Meng, X. L., and D. B. Rubin, 1993 Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika **80:** 267–278.