J. Math. Biol. 48, 275–292 (2004)
Digital Object Identifier (DOI):
10.1007/s00285-003-0230-x

Mathematical Biology

Sabin Lessard · John Wakeley

The two-locus ancestral graph in a subdivided population: convergence as the number of demes grows in the island model

Received: 15 October 2002 / Revised version: 22 April 2003 /
Published online: 20 August 2003 –

Abstract. We study the ancestral recombination graph for a pair of sites in a geographically structured population. In particular, we consider the limiting behavior of the graph, under Wright's island model, as the number of subpopulations, or demes, goes to infinity. After an instantaneous sample-size adjustment, the graph becomes identical to the two-locus graph in an unstructured population, but with a time scale that depends on the migration rate and the deme size. Interestingly, when migration is gametic, this rescaling of time increases the population mutation rate but does not affect the population recombination rate. We compare this to the case of a partially-selfing population, in which both mutation and recombination depend on the selfing rate. Our result for gametic migration holds both for finite-sized demes, and in the limit as the deme size goes to infinity. However, when migration occurs during the diploid phase of the life cycle and demes are finite in size, the population recombination rate does depend on the migration rate, in a way that is reminiscent of partial selfing. Simulations imply that convergence to a rescaled panmictic ancestral recombination graph occurs for any number of sites as the number of demes approaches infinity.

1. Introduction

It is of great biological interest to understand the dynamics of recombination within genomes and the effects these can have on patterns of genetic variation. Restricted recombination between loci causes their fates to be correlated, slowing the breakdown of gametic associations (Lewontin and Kojima 1960) and reducing the effectiveness of natural selection (Hill and Robertson 1968). There is a correlation between rates of recombination and levels of polymorphism in humans (Nachman 2001) and in *Drosophila* (Begun and Aquadro 1992), and the rate and

S. Lessard: Département de mathématiques et de statistique, Université de Montréal.
e-mail: lessards@dms.umontreal.ca

J. Wakeley: Department of Organismic and Evolutionary Biology, Harvard University

Send offprint request to: Sabin Lessard

S. Lessard was supported by grants from the Natural Sciences and Research Council of Canada, the Fonds Québécois de la Recherche sur la Nature et les Technologies, and the Université de Montréal.

J. Wakeley was supported by a Career Award (DEB-0133760) and by a grant (DEB-9815367) from the National Science Foundation.

Key words or phrases: Coalescent – Island model – Migration – Recombination

pattern of recombination have important consequences for linkage studies of human diseases (Kruglyak 1999; Reich et al. 2001). In addition, several methods of estimating the population recombination rate have been developed (Hudson 1987; Hey and Wakeley 1997; Wall 1999; Fearnhead and Donnelly 2001; Hudson 2001). An understanding of recombination in populations is also important in gene mapping studies using linkage disequilibrium (Jorde 1995; Risch and Merikangas 1996; Lander 1996).

The bulk of theoretical studies of recombination have assumed a panmictic Wright-Fisher model (Fisher 1930; Wright 1931), while studies of recombination in structured populations have been rare. A structured population is a population that is subdivided into groups which contain one or more individuals, and the hallmark of subdivision is greater relatedness within groups than between groups. We present results for the two-locus ancestral process in the island model of population subdivision (Wright 1931; Moran 1959; Maruyama 1970; Latter 1973), in which groups are local populations (demes) and structure is mediated by restricted migration. The island model may not adequately represent the geographic structure of some species as it does not predict a pattern of "isolation by distance" (Wright 1943). Still, it is the most frequently applied migration model in empirical population genetic studies. Our results for the island model can be compared to those obtained for a population structured by partial selfing, in which case the groups are diploid individuals. In both models, the inbreeding coefficient F (Wright 1943) is important in determining the effects of structure, but results for the two models can be quite different.

In the context of the ancestral process for a sample, F is defined to be the probability that two lineages currently in the same group will coalesce before one or the other of them moves to a different group. In a partially selfing population, the effect of structure is to increase the rate of coalescence/drift by the factor $1+F$ relative to a panmictic population of the same total size (Pollack 1987; Nordborg and Donnelly 1997). By contrast, in the island migration model with a large number of large demes, analysis of dynamics at a single locus shows that the rate of coalescence/drift is decreased by the factor $1 - F$ (Wakeley 1998; Wright 1943). This difference is explained by the fact that even strictly selfing individuals can produce more than one offspring, whereas if no migration occurs in the island migration model demes cannot replace other demes. We extend the results for the island model to a pair of loci and also show that, if demes are small and migration occurs during the diploid phase of the life cycle, the effects of island model structure can be similar to those of partial selfing.

In particular, we show that, in the limit of a large number of demes, the ancestral process for a pair of selectively neutral loci in an island-model population can be described analogously to a panmictic population, as is possible in the single-locus case (Wakeley 1998). The effective size of the limiting ancestral process is given by $2ND/(MF)$, where N is the deme size, D is the number of demes and $M = 4Nm$, where m is the proportion of each deme replaced by migrants every generation. When migration occurs during the haploid, or gametic, phase of the life cycle, or if the deme size is large, then $MF = 1 - F$, whereas under diploid migration and small deme size MF is close to $1 + F$. Note that these effective sizes hold

when each lineage is in a separate deme. If some demes contain multiple lineages, a much faster process of migration and within-deme coalescence occurs until each remaining lineage is in a separate deme (Wakeley 1998).

In a population with neutral mutation rate u per generation at some locus, the effective size of the ancestral process sets the scaled population mutation rate — here, for the island model $\theta = 4NDu/(MF)$ — which in turn determines the rate at which polymorphisms will be found in a sample of DNA sequences at the locus. However, population structure affects the rate of recombination differently. Under both partial selfing and island model migration, and likely under any model of structure with F appropriately defined, a fraction F of recombination events are repaired by within-deme coalescent events. This is true because, looking backwards in time, the two chromosomal segments that result from a recombination event are initially present in the same group (deme or individual) and thus have a chance to coalesce before migration or outcrossing makes the recombination event potentially observable. Alternatively, looking forward in time, higher levels of relatedness within than between groups in a structured population increase the chance that a recombination event will involve homozygotes at one or both loci and thus be unobservable.

Under gametic migration, we find that $R = 4NDr$ is the appropriate population recombination rate in the island model with a large number of demes, where r is the per-generation rate of recombination between two loci. In this case, the opportunity for recombination is at once increased by the factor $1 - F$, due to the lengthening of coalescent times, and decreased by the same factor, due to the fact that only a fraction $1 - F$ of recombination events are observable. This result holds both when demes are finite in size and in the limit of large demes size (with $\lim_{N\to\infty} 4Nm = M$) as long as migration is gametic.

When migration occurs during the diploid phase of the life cycle and demes are finite in size, the fraction of recombination events that are repaired by coalescent events is not identical to the factor by which coalescent times are scaled. In the limit of a large number of demes, we find that levels of polymorphism will depend on $4NDu/(MF_1(N))$ and the population recombination rate is given by $4NDr(1 - F_1(N))/(MF_1(N))$, where $1 - F_1(N) \leq MF_1(N) \leq 1 + F_1(N)$ and $F_1(N)$ is the inbreeding coefficient for an individual. Thus, the results for diploid migration are intermediate between those for gametic migration and the analogous results of Nordborg (2000) for partial selfing. The similarity to partial selfing is due to the fact that a pair of chromosomes can travel together via diploid migration if they are in the same individual. A proof of Nordborg's (2000) result, at least in the case of two loci, which Nordborg and Krone (2002) point out is lacking, could follow along the same lines as the one we give here for the island model.

In this paper, we consider changes backward in time in the distribution of the ancestral material of sampled chromosomal segments at two sites with recombination in a variety of island models and show convergence of the discrete-time transition matrices to continuous-time approximations as the number of demes D tends to infinity and $2ND$ generations are taken as unit of time where N is the deme size. Our convergence results are essentially based on a theorem for Markov chains due to Möhle (1998) and originally used to study the coalescent with partial selfing,

the number of ancestral genes as well as the different identity classes at a single locus without recombination backward in time as in Kingman's (1982) coalescent. They are applied in the same way and under the same conditions.

In the first part of the paper (Sections 2 and 3), we let N tend to infinity and then D tend to infinity, which is less cumbersome and more intuitive than the opposite case or the case N finite, to introduce the notation and illustrate the method. In the second part (Section 4 for the case of gametic migration and Section 5 for the case of diploid migration), we let D go to infinity for N fixed, which has an interest on its own, and we further check that this result is consistent with the previous one when we let N go to infinity in the limit. In Section 7, we explain that convergence must actually occur in all cases, irrespectively of the way N and D tend to infinity, and we discuss some biological applications.

2. The ancestral recombination graph for pairs of sites in a structured population

Assume discrete, non-overlapping generations in a diploid population structured into D demes with Nc_μ individuals in deme μ. Under gametic migration, at the beginning of every generation, each individual in the population produces the same very large number of gametes, which then disperse independently. Let $m_{\mu\nu}$ be the proportion of gametes in deme μ that come from deme ν. These backward migration rates per generation are assumed to be constant over time. After migration there is random union of gametes and population regulation within demes to form the diploid individuals of the next generation. The genealogical process for this model for a single locus, or site, is known as the structured Wright-Fisher coalescent (Wilkinson-Herbots 1998; Nordborg 2001). We consider two chromosomal sites and assume a recombination rate r per generation between them.

Looking backward in time at the genealogy of a sample of chromosomal segments at the two sites, the number and location of the ancestral segments in a given generation can be described by the vector

$$\mathbf{n} = (\mathbf{n}_1, ..., \mathbf{n}_D),$$

where

$$\mathbf{n}_\mu = (n_\mu^{(1)}, n_\mu^{(2)}, n_\mu^{(3)}),$$

for $\mu = 1, ..., D$. That is, deme μ contains $n_\mu^{(1)}$ segments ancestral at site 1, $n_\mu^{(2)}$ ancestral at site 2 and $n_\mu^{(3)}$ ancestral at both sites, for $\mu = 1, ..., D$.

Here we make the usual coalescent assumption that N is large and $m_{\mu\nu}$ is correspondingly small. Then, measuring time in units of $2ND$ generations, any segment in deme μ may migrate to deme ν at a rate $DM_{\mu\nu}/2$, where $M_{\mu\nu} = 4Nm_{\mu\nu}$ is a scaled migration rate, or recombine at a rate $R/2$, where $R = 4NDr$ is a scaled recombination rate. Recombination will change the sample configuration only if it occurs on a segment that is ancestral at both sites, thus creating two new segments that are each ancestral at one site. A segment may coalesce with another ancestral segment in the same deme at a rate $(n_\mu - 1)D/c_\mu$, where

$$n_\mu = n_\mu^{(1)} + n_\mu^{(2)} + n_\mu^{(3)}$$

is the total number of ancestral segments in deme μ. When two segments ancestral at the same site coalesce, the number of ancestral segments (at that site) decreases by one. Coalescent events can also change the linkage patterns in the sample by joining together two segments that are not both ancestral at both sites.

As N goes to infinity, only one of these events can occur at a time according to a continuous-time Markov chain. The chain remains in state $\mathbf{n}$ for an exponentially distributed length of time with parameter

$$\lambda_{\mathbf{n}} = \sum_{\mu} \left\{ \frac{n_\mu D M_\mu}{2} + \frac{n_\mu^{(3)} R}{2} + \frac{n_\mu (n_\mu - 1) D}{2 c_\mu} \right\},$$

where

$$M_\mu = \sum_{\nu \neq \mu} M_{\mu\nu}.$$

Then, there is a transition to another state $\mathbf{n}'$ with probability $Q_{\mathbf{n}\mathbf{n}'}$ given by

$$\begin{cases}
\frac{n_\mu^{(k)} D M_{\mu\nu}}{2\lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_\mu^{(k)} + \mathbf{e}_\nu^{(k)}, \text{ for } \nu \neq \mu \text{ and } k = 1, 2, 3, \\
\frac{n_\mu^{(3)} R}{2\lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} + \mathbf{e}_\mu^{(1)} + \mathbf{e}_\mu^{(2)} - \mathbf{e}_\mu^{(3)}, \\
\frac{n_\mu^{(1)} n_\mu^{(2)} D}{c_\mu \lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_\mu^{(1)} - \mathbf{e}_\mu^{(2)} + \mathbf{e}_\mu^{(3)}, \\
\frac{2 n_\mu^{(k)} n_\mu^{(3)} D + n_\mu^{(k)} (n_\mu^{(k)} - 1) D}{2 c_\mu \lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_\mu^{(k)}, \text{ for } k = 1, 2, \\
\frac{n_\mu^{(3)} (n_\mu^{(3)} - 1) D}{2 c_\mu \lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_\mu^{(3)},
\end{cases}$$

where $\mathbf{e}_\mu^{(k)}$ designates a vector of all zero D triplets except the μ-*th*, which is $(1, 0, 0)$ if $k = 1$, $(0, 1, 0)$ if $k = 2$ and $(0, 0, 1)$ if $k = 3$.

The above formulation describes the structured ancestral recombination graph at two neutral sites. Notice that such a graph is symmetric with respect to the two sites. It is a generalization of the structured coalescent at a single neutral site (Wilkinson-Herbots 1998; Nordborg 2001), which in turn generalized Kingman's (1982) coalescent. Recombination was first incorporated in the (unstructured) coalescent by Hudson (1983) and Hudson and Kaplan (1985), and was considered later on in a context of selection, which shares some similarity with geographic subdivision (Hudson and Kaplan 1988; Kaplan, Darden and Hudson 1988; Kaplan, Hudson and Iizuka 1991). The terminology 'ancestral recombination graph' was introduced by Griffiths and Marjoram (1996) to describe the lineages of chromosomal segments at a continuum of neutral sites with recombination in a random mating population.

3. The ancestral recombination graph for pairs of sites in the island model with a large number of demes

In the island model, the D demes are assumed to be of the same size and the backward migration rates to other demes all equal. Therefore, we have $c_\mu = 1$ for all

μ and $M_{\mu\nu} = M/(D-1)$ for all $\nu \neq \mu$. Now the state of the ancestral material at two sites, given by the vector $\mathbf{n} = (\mathbf{n}_1, ..., \mathbf{n}_D)$, where $\mathbf{n}_\mu = (n_\mu^{(1)}, n_\mu^{(2)}, n_\mu^{(3)})$, for deme $\mu = 1, ..., D$, is the same by symmetry under any permutation of the demes. Moreover, only the ancestral demes matter, that is, the demes that contain ancestral segments. In the following, segments that are ancestral at site 1 will be said of type 1, those ancestral at site 2 of type 2 and those ancestral at both sites of type 3. When a deme contains only one ancestral segment, it is said of the type of the ancestral segment that it contains.

If we order the demes in such a way that the number of ancestral segments is decreasing, and then the number of segments of type 3, 2, 1 in this order in case of equalities, we find a standard state $\tilde{\mathbf{n}}$ that is equivalent to $\mathbf{n}$ and we note $\mathbf{n} \cong \tilde{\mathbf{n}}$. If $d = d(\mathbf{n}) = d(\tilde{\mathbf{n}})$ is the number of ancestral demes, then $\tilde{n}_\mu \geq 1$ for $\mu \leq d$ and $\tilde{n}_\mu = 0$ for $\mu > d$, where

$$\tilde{n}_\mu = \tilde{n}_\mu^{(1)} + \tilde{n}_\mu^{(2)} + \tilde{n}_\mu^{(3)}$$

is the number of ancestral segments in deme μ with demes ordered such that $\mu < \nu$ if

$$\tilde{n}_\mu > \tilde{n}_\nu$$

or

$$\tilde{n}_\mu = \tilde{n}_\nu \text{ and } \tilde{n}_\mu^{(3)} > \tilde{n}_\nu^{(3)}$$

or

$$\tilde{n}_\mu = \tilde{n}_\nu \text{ and } \tilde{n}_\mu^{(3)} = \tilde{n}_\nu^{(3)} \text{ and } \tilde{n}_\mu^{(2)} > \tilde{n}_\nu^{(2)}$$

or

$$\tilde{n}_\mu = \tilde{n}_\nu \text{ and } \tilde{n}_\mu^{(3)} = \tilde{n}_\nu^{(3)} \text{ and } \tilde{n}_\mu^{(2)} = \tilde{n}_\nu^{(2)} \text{ and } \tilde{n}_\mu^{(1)} \geq \tilde{n}_\nu^{(1)}.$$

Notice that the standard state $\tilde{\mathbf{n}}$ does not actually depend on D, since the number of positive entries is always bounded by twice the number of ancestral segments in the original sample.

Let S be the set of all possible standard states of the ancestral material, ordered by convention such that the states with at most one ancestral segment in each deme ($\tilde{n}_\mu = 1$ for $\mu \leq d$ and 0 for $\mu > d$) come first, those with two ancestral segments in one deme and at most one in each other ($\tilde{n}_1 = 2$ and $\tilde{n}_\mu = 1$ for $2 \leq \mu \leq d$ and 0 for $\mu > d$) come next, and all other states come last. Let us denote by S_1, S_2 and S_3 the corresponding subsets of states.

Since every ancestral demes describing a state in S_1 contains only one ancestral segment, such a state can be described by a vector $\mathbf{d} = (d^{(1)}, d^{(2)}, d^{(3)})$, where $d^{(k)}$ is the number of demes containing one segment of type k for $k = 1, 2, 3$ and $d^{(1)} + d^{(2)} + d^{(3)} = d$. Furthermore, a state in S_2 can be represented by a vector $\mathbf{d} = (d^{(1)}, d^{(2)}, d^{(3)}, d^{(11)}, d^{(12)}, d^{(13)}, d^{(22)}, d^{(23)}, d^{(33)})$, where $d^{(kl)} = 1$ if one deme contains a segment of type k and another of type l for $k, l = 1, 2, 3$ and only one is equal to 1 with $\sum_{k=1}^{3}(d^{(k)} + \sum_{l=k}^{3} d^{(kl)}) = d$.

Taking $2ND$ as the unit of time where N is the deme size and letting N go to infinity, the transition matrix of the Markov chain on S over a period of time t in the past can be written in the form

$$\mathbf{P}(t) = \exp\left\{\left[\mathbf{A} + \frac{\mathbf{B}(D)}{D}\right] Dt\right\},$$

where the element $A_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}'}$ of the matrix $\mathbf{A}$ is given by

$$\begin{aligned}
&\sum_{\mu\in\Gamma_1^{(k)}}\left\{\frac{\tilde{n}_\mu^{(k)}M}{2}\right\} && \text{if one migration of a segment of type } k \text{ from a deme with 2 or more ancestral segments to a non ancestral deme, for } k = 1, 2, 3,\\
&\sum_{\mu\in\Gamma_2}\left\{\tilde{n}_\mu^{(1)}\tilde{n}_\mu^{(2)}\right\} && \text{if one coalescence between segments of type 1 and 2, respectively,}\\
&\sum_{\mu\in\Gamma_3^{(k)}}\left\{\tilde{n}_\mu^{(k)}\tilde{n}_\mu^{(3)}+\frac{\tilde{n}_\mu^{(k)}(\tilde{n}_\mu^{(k)}-1)}{2}\right\} && \text{if one coalescence between one segment of type } k \text{ and one of type } k \text{ or 3, for } k = 1, 2,\\
&\sum_{\mu\in\Gamma_3^{(3)}}\left\{\frac{\tilde{n}_\mu^{(3)}(\tilde{n}_\mu^{(3)}-1)}{2}\right\} && \text{if one coalescence between two segments of type 3,}\\
&-\sum_{\mu\in\Gamma_4}\left\{\frac{\tilde{n}_\mu M+\tilde{n}_\mu(\tilde{n}_\mu-1)}{2}\right\} && \text{if the same standard state,}
\end{aligned}$$

and 0 otherwise, while the element $B_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}'}(D)$ of $\mathbf{B}(D)$ is given by

$$\begin{aligned}
&-\sum_{\mu\in\Gamma_1^{(k)}}\left\{\frac{\tilde{n}_\mu^{(k)}MD(d-1)}{2(D-1)}\right\} && \text{if one migration of a segment of type } k \text{ from a deme with 2 or more ancestral segments to a non ancestral deme, for } k = 1, 2, 3,\\
&\sum_{\mu\in\Gamma_5^{(k)}}\left\{\frac{\tilde{n}_\mu^{(k)}MD}{2(D-1)}\right\} && \text{if one migration of a segment of type } k \text{ to an ancestral deme, for } k = 1, 2, 3,\\
&\sum_{\mu\in\Gamma_6}\left\{\frac{\tilde{n}_\mu^{(3)}R}{2}\right\} && \text{if one recombination of a segment of type 3,}\\
&-\sum_{\mu\in\Gamma_7}\left\{\frac{\tilde{n}_\mu^{(3)}R}{2}\right\}-\sum_{\mu\in\Gamma_8}\left\{\frac{\tilde{n}_\mu MD(d-1)}{2(D-1)}\right\} && \text{if the same standard state,}
\end{aligned}$$

and 0 otherwise, with

$$\begin{aligned}
\Gamma_1^{(k)} &= \left\{\mu : \tilde{\mathbf{n}}-\mathbf{e}_\mu^{(k)}+\mathbf{e}_{d+1}^{(k)} \cong \tilde{\mathbf{n}}' \text{ and } \tilde{n}_\mu^{(k)} \geq 1 \text{ with } \tilde{n}_\mu \geq 2\right\},\\
\Gamma_2 &= \left\{\mu : \tilde{\mathbf{n}}-\mathbf{e}_\mu^{(1)}-\mathbf{e}_\mu^{(2)}+\mathbf{e}_\mu^{(3)} \cong \tilde{\mathbf{n}}' \text{ and } \tilde{n}_\mu^{(1)}, \tilde{n}_\mu^{(2)} \geq 1\right\},\\
\Gamma_3^{(k)} &= \left\{\mu : \tilde{\mathbf{n}}-\mathbf{e}_\mu^{(k)} \cong \tilde{\mathbf{n}}' \text{ and } \tilde{n}_\mu^{(k)} \geq 1\right\},\\
\Gamma_4 &= \left\{\mu : \tilde{n}_\mu \geq 2\right\},\\
\Gamma_5^{(k)} &= \left\{\mu : \tilde{\mathbf{n}}-\mathbf{e}_\mu^{(k)}+\mathbf{e}_\nu^{(k)} \cong \tilde{\mathbf{n}}' \text{ different from } \tilde{\mathbf{n}}\right.\\
&\qquad \left.\text{with } \tilde{n}_\mu^{(k)} \geq 1 \text{ and } \tilde{n}_\nu \geq 1 \text{ for some } \nu \neq \mu\right\},
\end{aligned}$$

$$\Gamma_6 = \left\{\mu : \tilde{\mathbf{n}} - \mathbf{e}_\mu^{(3)} + \mathbf{e}_\mu^{(1)} + \mathbf{e}_\mu^{(2)} \cong \tilde{\mathbf{n}}' \text{ and } \tilde{n}_\mu^{(3)} \geq 1\right\},$$

$$\Gamma_7 = \left\{\mu : \tilde{n}_\mu \geq 1\right\},$$

$$\Gamma_8 = \left\{\mu : \tilde{n}_\mu = 1\right\}.$$

Notice that the matrix $\mathbf{A}$ is an infinitesimal generator, actually D times the infinitesimal generator of the Markov chain on S when there is no recombination and migration occurs only to non ancestral demes, with time measured in units of $2N$ generations and N going to infinity. For this chain, the states in S_1 are absorbing and all others are transient. Therefore, the infinitesimal generator $\mathbf{A}$ is in the block form

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{pmatrix}$$

with respect to the subsets of states S_1, S_2 and S_3 in this order, where $\mathbf{0}$ denotes a null matrix, while

$$\mathbf{P} = \lim_{t\to\infty} \exp\{\mathbf{A}t\}$$

is in the block form

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{21} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{31} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{I}$ is an identity matrix, and the elements of $\mathbf{P}_{21}$ and $\mathbf{P}_{31}$ are probabilities of fixation from states in S_2 and S_3, respectively, into states in S_1. The elements of $\mathbf{P}_{21}$ will be 0 or F or $1 - F$, where $F = 1/(1 + M)$ is the probability that two segments in the same deme coalesce before migrating. Actually, from a state $\mathbf{d}$ in S_2 with $d^{(kl)} = 1$ for $k \leq l$, we will have absorption either to the state with $d^{(k)}$ and $d^{(l)}$ increased by 1 ($d^{(l)}$ increased by 2 in the case $k = l$) with probability $1 - F$, or to the state in S_1 with $d^{(l)}$ increased by 1 ($d^{(3)}$ increased by 1 in the case $k = 1$ and $l = 2$) with probability F. Similarly, the elements of $\mathbf{P}_{31}$ can be obtained by looking at the number of segments of each type that can migrate from each ancestral deme as in Wakeley (2001).

On the other hand, $\mathbf{B}(D)$ converges to a matrix $\mathbf{B}$ as D goes to infinity, whose element $B_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}'}$ is given by

$$-\sum_{\mu\in\Gamma_1^{(k)}} \left\{\frac{\tilde{n}_\mu^{(k)} M(d-1)}{2}\right\} \quad \text{if one migration of a segment of type } k \text{ from a deme with 2 or more ancestral segments to a non ancestral deme, for } k = 1, 2, 3,$$

$$\sum_{\mu\in\Gamma_5^{(k)}} \left\{\frac{\tilde{n}_\mu^{(k)} M}{2}\right\} \quad \text{if one migration of a segment of type } k \text{ to an ancestral deme, for } k = 1, 2, 3,$$

$$\begin{array}{ll} \sum_{\mu\in\Gamma_6}\left\{\frac{\tilde{n}_\mu^{(3)}R}{2}\right\} & \text{if one recombination of a segment of type 3,} \\ -\sum_{\mu\in\Gamma_7}\left\{\frac{\tilde{n}_\mu^{(3)}R}{2}\right\}-\sum_{\mu\in\Gamma_8}\left\{\frac{\tilde{n}_\mu M(d-1)}{2}\right\} & \text{if the same standard state,} \end{array}$$

and 0 otherwise. This matrix can be written in the block form

$$\mathbf{B}=\begin{pmatrix} \mathbf{D} & \mathbf{B}_{12} & \mathbf{0} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_{33} \end{pmatrix}$$

with respect to the subsets of states S_1, S_2 and S_3 in this order, where $\mathbf{D}$ is a diagonal matrix with $-[d^{(3)}R + d(d-1)M]/2$ on the main diagonal for a state in S_1 in the form $(d^{(1)}, d^{(2)}, d^{(3)})$ with $d = d^{(1)} + d^{(2)} + d^{(3)}$, while $\mathbf{B}_{12}$ contains rates of recombination and rates of migration to ancestral demes from states in S_1 to states in S_2, actually, $d^{(3)}R/2$ for a recombination, decreasing $d^{(3)}$ by 1 and leading to $d^{(12)} = 1$, $d^{(1)}d^{(2)}M$ for a migration from a deme of type 1 to a deme of type 2 or the opposite, decreasing $d^{(1)}$ and $d^{(2)}$ by 1 and leading to $d^{(12)} = 1$, $d^{(k)}d^{(3)}M$ for a migration from a deme of type k to a deme of type 3 or the opposite, decreasing $d^{(k)}$ by 1 and leading to $d^{(k3)} = 1$, for $k = 1, 2$, and $d^{(k)}(d^{(k)} - 1)M/2$ for a migration from a deme of type k to another deme of type k, decreasing $d^{(k)}$ by 1 and leading to $d^{(kk)} = 1$, for $k = 1, 2, 3$. These correspond to rates of recombination and M times rates of coalescence in a random mating population of size D when D is taken as the unit of time and let to go to infinity.

Moreover, we have

$$\exp\left\{\mathbf{A}+\frac{\mathbf{B}(D)}{D}\right\}=\sum_{n\geq 0}\frac{\left(\mathbf{A}+\frac{\mathbf{B}(D)}{D}\right)^n}{n!}=\exp\{\mathbf{A}\}+\frac{\mathbf{C}(D)}{D},$$

where $\exp\{\mathbf{A}\}$ is a stochastic matrix and

$$\mathbf{C}(D)=\sum_{n\geq 1}\left\{\frac{\sum_{k=0}^{n-1}\mathbf{A}^k\mathbf{B}(D)\mathbf{A}^{n-k-1}}{n!}\right\}+O(1/D)$$

converges to the matrix

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

as D goes to infinity. Then, a lemma due to Möhle (1998) guarantees that

$$\lim_{D\to\infty}\exp\left\{\left[\mathbf{A}+\frac{\mathbf{B}(D)}{D}\right]Dt\right\}=\mathbf{P}\exp\{t\mathbf{G}\},$$

where

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

But actually, we have

$$\mathbf{G} = \mathbf{PBP},$$

since

$$\mathbf{PA} = \mathbf{AP} = \mathbf{0}.$$

We get

$$\mathbf{G} = \begin{pmatrix} \mathbf{D} + \mathbf{B}_{12}\mathbf{P}_{21} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{21}\mathbf{D} + \mathbf{P}_{21}\mathbf{B}_{12}\mathbf{P}_{21} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{31}\mathbf{D} + \mathbf{P}_{31}\mathbf{B}_{12}\mathbf{P}_{21} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Therefore, we have instantaneous transitions from states in S_2 and S_3 to states in S_1 and, once in S_1, a Markov chain with infinitesimal generator $\mathbf{D} + \mathbf{B}_{12}\mathbf{P}_{21}$.

From a state $(d^{(1)}, d^{(2)}, d^{(3)})$ in S_1, there is recombination or migration to ancestral demes according to the rates given by the elements of $\mathbf{B}_{12}$, but then the two ancestral segments that happen to coexist momentarily in the same deme either coalesce with probability F or end up in different demes with probability $1 - F$, which are the non zero elements of $\mathbf{P}_{21}$. There will be a change in the former case only if the two segments in the same deme were the result of migration, and in the latter only if they were the result of recombination. Notice that the factor M in the rates of migration times F gives $M/(1 + M) = 1 - F$. Then, we get the rates of change

$$\begin{aligned}
d^{(3)}R(1-F)/2 &\quad \text{to } (d^{(1)}+1, d^{(2)}+1, d^{(3)}-1),\\
d^{(1)}d^{(2)}(1-F) &\quad \text{to } (d^{(1)}-1, d^{(2)}-1, d^{(3)}+1),\\
(d^{(1)}d^{(3)} + d^{(1)}(d^{(1)}-1)/2)(1-F) &\quad \text{to } (d^{(1)}-1, d^{(2)}, d^{(3)}),\\
(d^{(2)}d^{(3)} + d^{(2)}(d^{(2)}-1)/2)(1-F) &\quad \text{to } (d^{(1)}, d^{(2)}-1, d^{(3)}),\\
d^{(3)}(d^{(3)}-1)(1-F)/2 &\quad \text{to } (d^{(1)}, d^{(2)}, d^{(3)}-1).
\end{aligned}$$

These are $1 - F$ times the rates of change in a random mating population. If the unit of time is divided by $1 - F$, we get the generator for a random mating population with the same coalescence rates and the scaled recombination rate R.

4. The case of small deme size

In this section, we consider the island model with a large number of demes in the case where the deme size N is small and the backward migration rate per generation m is not necessarily small, and we define $M = 2Nm(2 - m)$, where $m(2 - m) = 1 - (1 - m)^2$ is the probability that at least one of two segments in the same deme comes from another deme in the previous generation. Looking at the ancestral material at two sites for a given sample of segments, there may be migration, coalescence and recombination events and some of these events may

occur simultaneously. Nevertheless, the transition matrix on S from one generation to the previous one can still be written in the form

$$\mathbf{I} + \frac{\mathbf{A}(N)}{2N} + \frac{\mathbf{B}(N, D)}{2ND},$$

where $\mathbf{I}+\mathbf{A}(N)/(2N)$ is the transition matrix that we would get assuming no recombination and migration only to different non ancestral demes, and $\mathbf{B}(N, D)/(2ND)$ contains all other transition probabilities minus terms added to get $\mathbf{A}(N)$. Defining S_1, S_2 and S_3 as previously, the states in S_1 are absorbing for $\mathbf{I} + \mathbf{A}(N)/(2N)$ and

$$\lim_{t\to\infty} \left[\mathbf{I} + \frac{\mathbf{A}(N)}{2N}\right]^t = \mathbf{P}(N),$$

where $\mathbf{P}(N)$ can be written in the block form

$$\mathbf{P}(N) = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{21}(N) & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{31}(N) & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

with the non null entries of $\mathbf{P}_{21}(N)$ being

$$F(N) = \frac{(1-m)^2}{(1-m)^2 + M}$$

or $1 - F(N)$. Notice that $F(N)$ is the probability that two lineages starting in the same deme coalesce before one or the other migrates. On the other hand, we have

$$\lim_{D\to\infty} \mathbf{B}(N, D) = \mathbf{B}(N),$$

where

$$\mathbf{B}(N) = \begin{pmatrix} \mathbf{B}_{11}(N) & \mathbf{B}_{12}(N) & \mathbf{0} \\ \mathbf{B}_{21}(N) & \mathbf{B}_{22}(N) & \mathbf{B}_{23}(N) \\ \mathbf{B}_{31}(N) & \mathbf{B}_{32}(N) & \mathbf{B}_{33}(N) \end{pmatrix}.$$

The entries of $\mathbf{B}_{12}(N)$ on the line corresponding to the state $(d^{(1)}, d^{(2)}, d^{(3)})$ in S_1 are $d^{(3)}R/2$ for a recombination, decreasing $d^{(3)}$ by 1 and leading to $d^{(12)} = 1$, $d^{(1)}d^{(2)}M(2N-1)/(2N)$ for a migration from a deme of type 1 to a deme of type 2 without coalescence, or the opposite, decreasing $d^{(1)}$ and $d^{(2)}$ by 1 and leading to $d^{(12)} = 1$, $d^{(k)}d^{(3)}M(2N-1)/(2N)$ for a migration from a deme of type k to a deme of type 3 without coalescence, or the opposite, decreasing $d^{(k)}$ by 1 and leading to $d^{(k3)} = 1$, for $k = 1, 2$, and $d^{(k)}(d^{(k)}-1)M(2N-1)/(4N)$ for a migration from a deme of type k to another deme of type k without coalescence, decreasing $d^{(k)}$ by 1 and leading to $d^{(kk)} = 1$, for $k = 1, 2, 3$. On the other hand, the entries of $\mathbf{B}_{11}(N)$ off the main diagonal are $d^{(1)}d^{(2)}M/(2N)$ for a migration from a deme of type 1 to a deme of type 2 with coalescence, or the opposite, decreasing $d^{(1)}$ and

$d^{(2)}$ by 1 and increasing to $d^{(3)}$ by 1, $d^{(k)}d^{(3)}M/(2N)$ for a migration from a deme of type k to a deme of type 3 with coalescence, or the opposite, decreasing $d^{(k)}$ by 1 and increasing $d^{(3)}$ by 1, for $k = 1, 2$, and $d^{(k)}(d^{(k)} - 1)M/(4N)$ for a migration from a deme of type k to another deme of type k with coalescence, decreasing $d^{(k)}$ by 1, for $k = 1, 2, 3$. The entries of $\mathbf{B}_{11}(N)$ on the main diagonal are simply minus the sums of all other entries on the same rows in $\mathbf{B}_{11}(N)$ and $\mathbf{B}_{12}(N)$.

Taking $2ND$ as the unit of time, the transition matrix from 0 to time t in the past as D goes to infinity is

$$\lim_{D\to\infty}\left[\mathbf{I} + \frac{\mathbf{A}(N)}{2N} + \frac{\mathbf{B}(N, D)}{2ND}\right]^{[2NDt]} = \mathbf{P}(N)\exp\{t\mathbf{G}(N)\},$$

where

$$\mathbf{G}(N) = \mathbf{P}(N)\mathbf{B}(N)\mathbf{P}(N)$$

and $[2NDt]$ designates the integer of $2NDt$. We find

$$\mathbf{G}(N) = \begin{pmatrix} \mathbf{B}_{11}(N) + \mathbf{B}_{12}(N)\mathbf{P}_{21}(N) & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{21}(N)\mathbf{B}_{11}(N) + \mathbf{P}_{21}(N)\mathbf{B}_{12}(N)\mathbf{P}_{21}(N) & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{31}(N)\mathbf{B}_{11}(N) + \mathbf{P}_{31}(N)\mathbf{B}_{12}(N)\mathbf{P}_{21}(N) & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

After instantaneous transitions from states in S_2 and S_3 to states in S_1, we get the rates of change in a random mating population multiplied by

$$1 - F(N) = \frac{M(2N-1)F(N)}{2N} + \frac{M}{2N}.$$

As N goes to infinity, $F(N)$ goes to F and $\mathbf{P}(N)$ goes to $\mathbf{P}$, and the above result is in agreement with the previous one.

5. The case of diploid migration

When the deme size is small, there is a difference between diploid migration and gametic migration even in the case of a large number of demes. Assume that migration occurs after mating within demes, but before reproduction. In this case, the states in S_2 can be of two possible kinds: those with the two ancestral segments in one deme in the same individual and those with these segments in two individuals. Then, the non null entries of $\mathbf{P}_{21}(N)$ will be $F_1(N)$, $1 - F_1(N)$, $F_2(N)$ or $1 - F_2(N)$, where $F_1(N)$ and $F_2(N)$ are the probabilities of coalescence before migration when we consider two segments at a single site in the same individual and in two individuals in the same deme, respectively. We find easily, assuming $N \geq 2$, that

$$F_2(N) = (1-m)^2 F_1(N)$$

and

$$F_1(N) = \frac{1}{M + 1 - 4m + 2m^2},$$

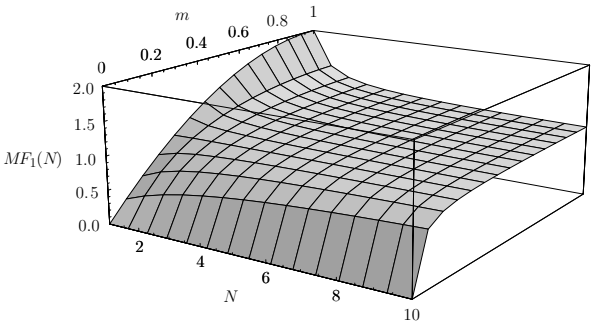

Fig. 1. Plot of $MF_1(N)$, the factor by which coalescent times are scaled under diploid migration, as a function of the migration rate and the deme size.

where again $M = 2Nm(2 - m)$. That is, M is equal to $2N$ times the probability that at least one of two individuals in the same deme comes from another deme in the previous generation.

Moreover, the entries of $\mathbf{B}_{12}(N)$ corresponding to migration events have the factor $M(N - 1)/N$ in the case of segments originating from separate individuals in the same deme and the factor $M/(2N)$ in the case of segments that are copies of the two segments in the same individual, while the entries of $\mathbf{B}_{11}(N)$ off the main diagonal have the same factor $M/(2N)$ as in the case of gametic migration. As a result, after instantaneous transitions from states in S_2 and S_3 to states in S_1, we get the rates of change in a random mating population multiplied by $(1 - F_1(N))$ in the case of recombination and

$$MF_1(N) = \frac{M(N-1)F_2(N)}{N} + \frac{MF_1(N)}{2N} + \frac{M}{2N}$$

in the case of coalescence. We have the inequalities

$$1 - F_1(N) \leq MF_1(N) \leq 1 + F_1(N)$$

with equalities in the cases $m = 0$ (or when N goes to infinity) and $m = 1$, respectively.

Figure 1 shows $MF_1(N)$ over a range of values of the migration rate and the deme size. When deme sizes are small and migration rates large, $MF_1(N)$ is close to $1 + F_1(N)$, and coalescent times are shortened relative to those in a panmictic population. When migration rates are smaller and deme sizes larger, $MF_1(N)$ is close to $1 - F_1(N)$, and the effect of subdivision under diploid migration is to lengthen coalescent times, similar to the case of gametic migration.

6. Simulations

To provide an additional check of our results and to investigate the convergence properties of the ancestral recombination graph for multiple sites, we performed coalescent simulations using software developed for Hudson et al. (1992), which Dick Hudson has kindly made available. These simulations assume that N is large,

but allow for an arbitrary number of sites and level of recombination. Importantly, they also allow D to be varied. We assumed a genetic locus composed of ten thousand sites (10Kb) between which recombination events could occur, and considered a range of recombination rates, sample sizes, and values of M. Most of the results are not shown, but all of them were completely consistent with our analytical predictions. In addition, they imply that the rescaled ancestral recombination graph, in which the rate of coalescence is rescaled by MF but the rate of recombination is not, holds for any number of sites.

Figure 2 presents the results of two sets of simulations for a pair of sequences taken from two different demes in a population with migration parameter $M = 1$. The two sets of simulations assumed different values of the recombination parameter, $R = 0$ and $R = 10$, between the two ends of the 10Kb locus. In each simulation replicate, the total length of the genealogy (the sum of the tree lengths for all ten thousand sites) was recorded, and in one million replicates these were compiled into the histograms shown in figure 2 for each value of D. Time is measured in units of $2ND$ generations, and on this timescale the expected total tree length is

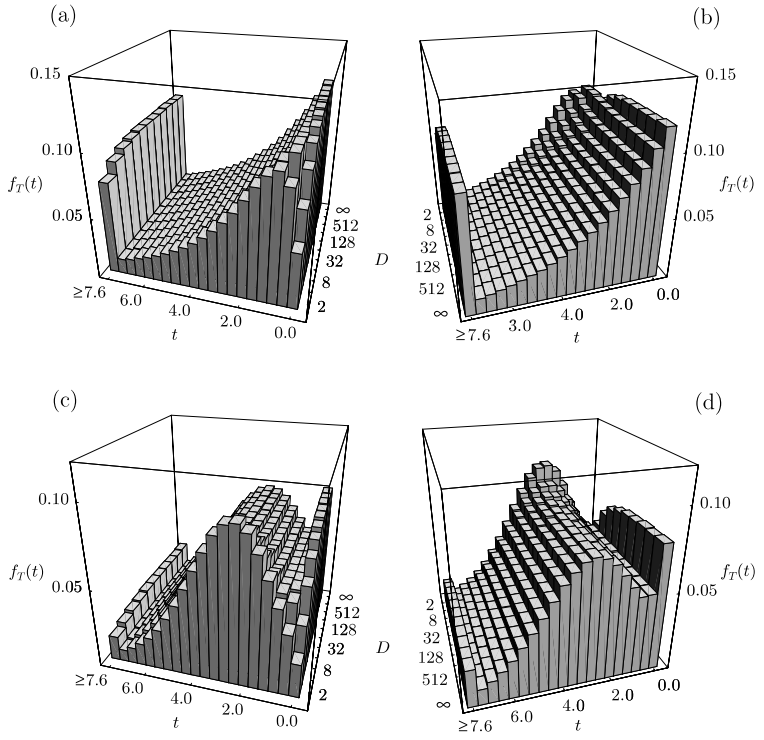

Fig. 2. The left and right panels give two alternative views of the distribution of the total length of the genealogy in simulations of a pair of sequences sampled from an island-model population, over a range of D, with $M = 1.0$, and $R = 0.0$ ((a) and (b)) or $R = 10.0$ ((c) and (d)).

$E[T] = 2(1 + 1/M)$, or 4 when $M = 1$. The number of demes was varied from its minimum value of $D = 2$ for this kind of sample up to 1024 on a log scale. The values for $D = \infty$ in the figure were obtained by simulating the genealogy for a pair of sequences sampled from a single panmictic population that had exactly those properties predicted for the limiting ancestral process described above, namely, with a coalescent rate rescaled by $MF = M/(1 + M)$. Clearly, the probability distribution of genealogical tree lengths approaches the large-D limiting result quite rapidly as the number of demes increases.

7. Discussion

In this paper, we have shown convergence of the discrete-time transition matrices, backward in time, for the distribution of the ancestral material of a sample of genes at two linked loci assuming an island model with either gametic or diploid migration and letting the total number of demes D go to infinity while measuring time in units of $2ND$ generations where N is any fixed deme size. Moreover, as N goes to infinity, before or after D goes to infinity, there is still convergence and the limit is the same in both cases. Actually, there is convergence to the same limit as N and D tend simultaneously to infinity, since the convergence as N goes to infinity is uniform with respect to D. As a matter of fact, the transition matrix on the state space S from one generation to the previous one given in Section 4 can also be written in the form

$$\mathbf{I} + \frac{\mathbf{A}}{2N} + \frac{\mathbf{B}(D)}{2ND} + \mathbf{O}(N, D),$$

where the matrices $\mathbf{A}$ and $\mathbf{B}(D)$ are as defined in Section 3 and the entries of $\mathbf{O}(N, D)$ are the transition probabilities involving two or more events of coalescence, migration or recombination. Since the number of such events is bounded and the probability of any two of these events, keeping $M = 4Nm$ and $R = 4NDr$ fixed, is also bounded by some constant times $(1/N^2)$, we have that $N\mathbf{O}(N, D)$ tends to zero uniformly with respect to D as N tends to infinity. Moreover, $[2NDt]/(2N)$ converges to Dt uniformly with respect to D as N tends to infinity, and therefore,

$$\lim_{N\to\infty} \left[\mathbf{I} + \frac{\mathbf{A}}{2N} + \frac{\mathbf{B}(D)}{2ND} + \mathbf{O}(N, D)\right]^{[2NDt]} = \exp\left\{\left[\mathbf{A} + \frac{\mathbf{B}(D)}{D}\right] Dt\right\},$$

the convergence being uniform with respect to D.

Numerical studies on linkage disequilibrium at two linked loci in the island model were performed and reported in a companion paper submitted to Genetics (Wakeley and Lessard 2003). These are in agreement with our convergence results and can explain some genomic data in humans.

The results presented in the present paper imply that methods developed for panmictic populations, for instance to estimate the population recombination rate (Hudson 1987; Hey and Wakeley 1997; Wall 1999; Fearnhead and Donnelly 2001; Hudson 2001), can be applied to samples in which each sequence comes from a different deme as long as an island model with a large number of demes remains a good approximation. It is only necessary to realize that estimates from these "scattered"

samples will depend on whether migration is gametic or demes are finite in size and migration occurs in the diploid phase of the life cycle. Under gametic migration, the only effect of population structure is to lengthen genealogies, and thus increase levels of polymorphism. The above methods will estimate the population rate of recombination $R = 4NDr$, which does not depend on the migration rate. However, under diploid migration it is possible (see figure 3) that these methods will estimate something substantially smaller than R. When demes are small and migration rates relatively high, diploid migration is similar to partial selfing, in which case the effective recombination rate is given by $R(1 - F)/(1 + F)$ (Nordborg 2000; Golding and Strobeck 1980). Although we have not proven convergence for the ancestral recombination graph of an entire linked sequence of sites (but only for a pair), simulations suggest that the same result holds for any number of sites.

In the case where multiple samples come from the same deme, it will be necessary to consider the instantaneous, stochastic adjustment to the sample size called the scattering phase (Wakeley 1999). The previous description of the scattering phase (Wakeley 1998) assumed that the deme size was large and the migration rate small, so that $M = 4Nm$ served as the migration parameter. Under this assumption, events within a deme cannot co-occur in the same generation. However, when N is not large and m not necessarily small, it is possible that several coalescent events and/or migration events will occur in the same generation. The description of the scattering phase would need to be modified accordingly. We note that such a description will not be as simple as it is in the case of large N, and that it will depend upon the order in which events take place in the life cycle and upon whether migration occurs in the diploid or the haploid phase.

Even without an analytical description of the scattering phase, our results suggest an efficient method of simulating genealogies of samples when there is recombination and the population is divided into a large number of demes. First, the scattering phase would be simulated and the descendants in the sample would be recorded for each remaining lineage (which are all in different demes at the end of the scattering phase). Next the remaining lineages would be fed into any standard coalescent simulation algorithm that includes recombination, such as that of Hudson (1983). Of course, the appropriate effective population rates of mutation and recombination would have to be computed and used as input in the simulation. Whenever the assumption of a large number of demes is appropriate, this method would give the desired distribution of genealogies and would be much more efficient than simulations in which all the demes in the population are represented.

References

Begun, D.J., Aquadro, C.F.: Levels of naturally occurring polymorphism correlate with recombination rate in Drosophila melanogaster. Nature **356**, 519–520 (1992)

Fearnhead, P., Donnelly, P.: Estimating recombination rates from population genetic data. Genetics **159**, 1299–1318 (2001)

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

Golding, G.B., Strobeck, C.: Linkage disequilibrium in a finite population that is partially selfing. Genetics **94**, 777–789 (1980)

Griffiths, R.C., Marjoram, P.: Ancestral inference from samples of DNA sequences with recombination. J. Comp. Biol. **3**, 479–502 (1996)

Hey, J., Wakeley, J.: A coalescent estimator of the population recombination rate. Genetics **145**, 833–846 (1997)
Hill, W.G., Robertson, A.R.: Linkage disequilibrium in finite populations. Theoret. Appl. Genet. **38**, 226–231 (1968)
Hudson, R.R.: Properties of a neutral allele model with intragenic recombination. Theoret. Pop. Biol. **23**, 183–201 (1983)
Hudson, R.R.: Estimating the recombination parameter of a finite population model without selection. Genet. Res. Camb. **50**, 245–250 (1987)
Hudson, R.R.: Two-locus sampling distributions and their application. Genetics **159**, 1805–1817 (2001)
Hudson, R.R., Kaplan, N.L.: Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**, 147–164 (1985)
Hudson, R.R., Kaplan, N.L.: The coalescent process in models with selection and recombination. Genetics **120**, 831–840 (1988)
Hudson, R.R., Slatkin, M., Maddison, W.P.: Estimation of levels of gene flow from DNA sequence data. Genetics **132**, 583–589 (1992)
Jorde, L.B.: Linkage disequilibrium as a gene mapping tool. Am. J. Hum. Genet. **56**, 11–14 (1995)
Kaplan, N.L., Darden, T., Hudson, R.R.: Coalescent process in models with selection. Genetics **120**, 819–829 (1988)
Kaplan, N.L., Hudson, R.R., Iizuka, M.: Coalescent processes in models with selection, recombination and geographic subdivision. Genet. Res. Camb. **57**, 83–91 (1991)
Kingman, J.F.C.: The coalescent. Stochastic Process. Appl. **13**, 235–248 (1982)
Kruglyak, L.: Prospects for whole genome linkage disequilibrium mapping of common diseases. Nature Genetics **22**, 139–144 (1999)
Lander, E.S.: The new genomics: global views of biology. Sci. **274**, 536–539 (1996)
Latter, B.D.H.: The island model of population differentiation: a general solution. Genetics **73**, 147–157 (1973)
Lewontin, R.C., Kojima, K.: The evolutionary dynamics of complex polymorphisms. Evolution **14**, 450–472 (1960)
Maruyama, T.: Effective number of alleles in a subdivided population. Theoret. Pop. Biol. **1**, 273–306 (1970)
Möhle, M.: A convergence theorem for Markov chains arising in population genetics and the coalescent with partial selfing. Adv. Appl. Prob. **30**, 493–512 (1998)
Moran, P.A.P.: The theory of some genetical effects of population subdivision. Austr. J. Biol. Sci. **12**, 109–116 (1959)
Nachman, M.W.: Single nucleotide polymorphisms and recombination rate in humans. Trends in Genetics **17**, 481–485 (2001)
Nordborg, M.: Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial selfing. Genetics **154**, 923–929 (2000)
Nordborg, M.: Coalescent theory. In: Handbook of Statistical Genetics, D. J. Balding, M. J. Bishop, C. Cannings, (eds.), John Wiley & Sons, Chichester, England, 2001
Nordborg, M., Donnelly, P.: The coalescent process with selfing. Genetics **146**, 1185–1195 (1997)
Nordborg, M., Krone, S.M.: Separation of time scales and convergence to the coalescent in structured populations. In: Modern Developments in Theoretical Population Genetics, M. Slatkin, M. Veuille, (eds.), Oxford University Press, Oxford, UK, 2002
Pollack, E.: On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117**, 353–360 (1987)
Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Farhadian, S.F., Ward, R., Lander, E.S.: Linkage disequilibrium in the human genome. Nature **411**, 199–204 (2001)
Risch, N., Merikangas, K.: The future of genetic studies of complex human diseases. Sci. **273**, 1516–1517 (1996)

Wakeley, J.: Segregating sites in Wright's island model. Theoret. Pop. Biol. **53**, 166–175 (1998)
Wakeley, J.: Non-equilibrium migration in human history. Genetics **153**, 1863–1871 (1999)
Wakeley, J.: The coalescent in an island model of population subdivision with variation among demes. Theoret. Pop. Biol. **59**, 133–144 (2001)
Wakeley, J., Lessard, S.: Theory of the effects of population structure and sampling on patterns of linkage disequilibrium applied to genomic data from humans. Genetics **164**, 1043–1053 (2003)
Wall, J.D.: A comparison of estimators of the population recombination rate. Mol. Biol. Evol. **17**, 156–163 (1999)
Wilkinson-Herbots, H.M.: Genealogy and subpopulation differentiation under various models of population structure. J. Math. Biol. **37**, 535–585 (1998)
Wright, S.: Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)
Wright, S.: Isolation by distance. Genetics **28**, 114–138 (1943)