Theoretical Population Biology **51**, 148–164 (1997)
Article No. TP971302

TPB

Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization

Frédéric Austerlitz, Bernard Jung-Muller*, Bernard Godelle†, and Pierre-Henri Gouyon

Laboratoire Évolution et Systématique, Université Paris XI, Bâtiment 362, F-91405 Orsay Cedex, France

Received May 6, 1996

We consider the impact of a colonization process on the genetic diversity and spatial structure of a geographically subdivided population. A stepping-stone model combined with coalescence theory is used to predict the evolution of sequence divergence and genetic parameters. We first derive analytical results for coalescence times in a population undergoing logistic growth. We next consider a stepping-stone model in which demes are successively colonized, starting from a first deme at one of the borders of the metapopulation. We use recurrence equations to calculate coalescence times for two genes chosen either inside the same deme or in different demes. This allows us to obtain the distribution and the expectation of the coalescence times, and to deduce from them the distribution of the average pairwise differences and the evolution of F_{st}. Our results reflect the impact of the founder effect, which becomes stronger as the distance of the deme from the first deme increases. An increase in migration rate or growth rate generally leads to a decrease of the founder effect. F_{st} (i) increases during the beginning of the colonization, (ii) decreases when migration creates homogenization and (iii) increases again towards an equilibrium value. The distributions of pairwise coalescence times or differences between sequences show a peak corresponding to the colonization period. These results could help detect former colonization events in natural populations.

1. INTRODUCTION

Many species have experienced a recent colonization process. For instance, temperate forest trees have recently undergone a process of colonization following the last glaciation. This colonization terminated recently (Bennett, 1983), and because of the length of the life cycle of tree populations, those species are not at equilibrium. As pointed out by Kremer (1994), there have been only a hundred or so generations since the beginning of the recolonization by forest trees of Europe and North America, about 18,000 years ago.

Some species have also been introduced into new habitats, often consecutively to human migrations, and have from there colonized new places (Demelo and Hebert, 1994; Johnson, 1988). They have suffered several founder events. In some cases they are still continuing to invade new territories. Founder effect is defined as the reduction in genetic diversity caused by the founding of a new population by a few individuals.

Even human populations may have been submitted to a bottleneck, about 400,000 to 800,000 generations ago, when the population probably decreased to 10,000 people (Takahata *et al.*, 1995; Rogers, 1995). All the present populations descend from this small one, whose descendants colonized the whole world. These results have been obtained using coalescence theory (Kingman, 1982)

* E.N.G.R.E.F., 14 rue Girardet, F-54042 Nancy Cedex, France.
† École Nationale Supérieure d'Horticulture, F-49000 Angers, France.

0040-5809/97 $25.00

to predict the average sequence diversity of genes under infinite site model. But most of the theoretical work that has been done was made in a panmictic model, which does not reflect the process of subdivision of human populations after their expansion into new territories.

We extended the coalescent process, which is the study of the probability of gene genealogies, to populations undergoing a colonization process. This can give us results on genetic diversity, which is directly correlated with coalescence times: the older the common ancestor of two genes, the more likely they are to have accumulated different mutations in both lineages.

Numerous studies have been undertaken to investigate the impact of demographic changes on genetic diversity. Nei *et al.* (1975) give results about the genetic changes caused by a bottleneck followed by a logistic growth of the population size. They have shown that a high growth rate after the bottleneck or a relatively large size of population during this period limits the reduction of genetic diversity. Their results are based on numerical integrations and concern only isolated populations.

Nichols and Hewitt (1994) have simulated the consequences of the colonization process on allelic frequencies, showing the importance of long distance migration. Their populations were growing at a finite rate but they did not study the impact of the growth rate on the process.

Notohara (1990) gives analytical results concerning the coalescent process in geographically subdivided populations, but these populations are supposed to be at equilibrium demographically and genetically and when he deals with stepping-stone models, they are with an infinite number of demes. The results obtained are rather complicated and do not fit many situations where the expansions of populations is limited by natural barriers, nor do they take into account the fact that populations are not necessarily at equilibrium.

Slatkin (1991) gives some results about the coalescence times of island and circular stepping-stone populations: the average intra-deme coalescence time does not depend on the migration rate whereas the inter-deme coalescence time does. He also gives the relationships between the coalescence times and parameters such as F_{st}, which is inversely proportional to the migration rate. He gives equilibrium values and rate of approach to equilibrium, starting with all demes being full and beginning that moment to exchange migrants.

Rogers and Harpending (1992) and Rogers (1995) study the influence of a sudden population growth on the distribution of pairwise differences between sequences. They showed that this process produces waves in the distribution that fit well available data on some human populations. The limitation of their approach is that they do not take into consideration the geographical subdivision of human populations, which is one of the factors that causes bias in their estimators.

Marjoram and Donnelly (1994) study the impact of geographic subdivision in populations that grow exponentially after a bottleneck. They show that a low migration rate often leads to multimodality in distributions of pairwise differences between sequences obtained by simulation. Their model is however not a model of colonization, since all subpopulations have the same size at every moment and the migration model is an island one.

The aim of the present work is to provide theoretical results for systems where colonization takes place, with a finite number of demes experiencing successive colonization events and then continue to exchange migrants and grow logistically. This situation is typical of many species including forest trees or invaders.

We model the colonization process with a metapopulation where demes are colonized one after another, in a linear, one-dimensional stepping-stone process, starting from a first deme at one of the borders, supposed to be full and at demographic and genetic equilibrium. Our results have been obtained by a recursion method and have been confirmed by simulations.

We obtain the distributions and the expectations of the various coalescence times for neutral genes. From those results we can obtain some information about the evolution of diversity, with parameters such as S, the number of pairwise differences between sequences, F_{st} or R_{st}, which is an equivalent to F_{st} for microsatellites, introduced by Slatkin (1995). From a theoretical point of view, our aim is to better understand how genetic diversity evolves and is structured by the colonization process. The other goal is to determine to which extent it is possible, using the pattern of within and between populations diversity, to conclude that a colonization event occurred at a given moment in the past.

In order to better explain our numerical results, we developed an analytical approach in a simple case: an isolated population whose size follows a logistic growth. The formulas we get allow a study of the impact of the different parameters and then an easier understanding of what happens for the general model.

2. ISOLATED POPULATION UNDER LOGISTIC GROWTH

2.1. Formulas

We suppose we have a population of deterministically varying size. We define t as the real time expressed in

number of generations. The population size at time t is denoted $N(t)$. The population is panmictic, composed of diploid individuals, with non-overlapping generations. For a sample of i genes ($i \geqslant 2$ and $\forall t\ i \ll N(t)$), the times $T_t(i)$, $T_t(i-1)$, ..., $T_t(2)$, during which there are respectively $i, i-1, ..., 2$ lineages (see Fig. 1), are random variables with joint distribution depending on time t. This joint distribution can be derived using the formulas given by Griffiths and Tavaré (1994),

$$p(t_i, ..., t_2) = \prod_{j=2}^{i} \frac{\binom{j}{2}}{2N(t-s_j)} \exp\left(-\binom{j}{2}(\Lambda_t(s_j) - \Lambda_t(s_{j+1}))\right), \quad (1)$$

where $\Lambda_t(s) = \int_0^s (\mathrm{d}s')/(2N(t-s'))$ and $s_{i+1} = 0$, $s_i = t_i$ and $s_j = t_j + t_{j-1} + \cdots + t_2, j = 2, ..., i-1$. Unlike t, the t_j are times measured backward. To calculate the expectation of the coalescence time $Tc_{i,t}$ and of the total tree length $L_{i,t}$ of a sample of i genes, we use the formulas of Slatkin (1996) which are modified using our notations into

$$\mathrm{E}(Tc_{i,t}) = \int_0^\infty (1 - g_{i,1}(\Lambda_t(s)))\, ds. \quad (2)$$

$$\mathrm{E}(L_{i,t}) = \sum_{j=2}^{i} j \int_0^\infty g_{i,j}(\Lambda_t(s))\, ds. \quad (3)$$

The $g_{i,j}$ process is defined in Tavaré (1994) as follows: $g_{i,j}(t)$ is the probability for i genes sampled in a population of constant size to have j distinct ancestors t generations ago. Replacing the $g_{i,j}$'s by their formula given in Tavaré (1984) leads to

$$\mathrm{E}(T_{i,t}) = \sum_{k=2}^{i} \frac{(-1)^k (2k-1)\, i_{[k]} \Psi(k(k-1)/2, t)}{i_{(k)}}, \quad (4)$$

$$\mathrm{E}(L_{i,t}) = \sum_{j=2}^{i} j \sum_{k=j}^{i} \frac{(2k-1)(-1)^{k-j} j_{(k-1)} i_{[k]} \Psi(k(k-1)/2, t)}{j!\,(k-j)!\, i_{(k)}}, \quad (5)$$

where $\Psi(n, t) = \int_0^\infty \exp(-n\Lambda_t(s))\, \mathrm{d}s$ for whatever integer n and

$$i_{[k]} = i(i-1)\cdots(i-k+1), k \geqslant 1; \qquad i_{[0]} = 1,$$
$$i_{(k)} = i(i+1)\cdots(i+k-1), k \geqslant 1; \qquad i_{(0)} = 1.$$

Simplifying, (5) becomes

$$\mathrm{E}(L_{i,t}) = 2 \sum_{k=1}^{\mathrm{Int}(i/2)} \frac{(4k-1)\, i_{[2k]} \Psi(k(2k-1), t)}{i_{(2k)}}, \quad (6)$$

where $\mathrm{Int}(x)$ denotes the integer part of a real number x.

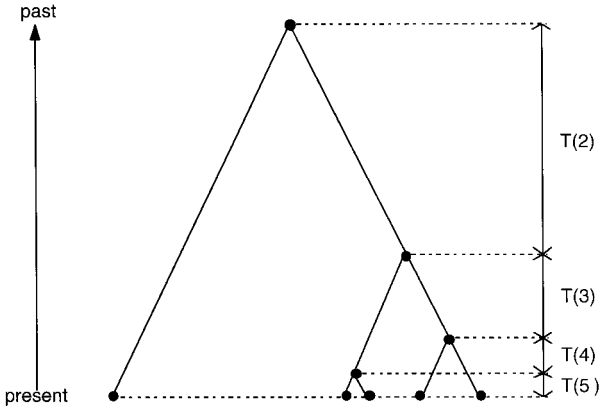


FIG. 1. Example of coalescent tree for five genes, with the times $T(j)$, $j = 2$, 3, 4 or 5, during which there are exactly j lineages.

We suppose that our population has the following demographic history: for $t < 0$ its size is constant: K_c. At $t = 0$, its population size decreases instantaneously to K_0 to rise again following logistic growth to its carrying capacity K.

The differential equation of the population size for $t > 0$ is

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right). \quad (7)$$

We can integrate it:

$$N(t) = K_c \qquad (t < 0),$$
$$N(t) = \frac{KK_0}{K_0 + (K - K_0)\, e^{-rt}} \qquad (t > 0).$$

In this case we can calculate $\Lambda_t(s)$,

$$\Lambda_t(s) = \frac{1}{2}\left(\frac{s}{K} + \left(\frac{1}{K_0} - \frac{1}{K}\right)\frac{e^{-rt}(e^{rs} - 1)}{r}\right) \qquad (t < s),$$

$$\Lambda_t(s) = \frac{1}{2}\left(\frac{t}{K} + \left(\frac{1}{K_0} - \frac{1}{K}\right)\frac{1 - e^{rt}}{r} + \frac{s - t}{K_c}\right) \qquad (t > s).$$

The integral $\Psi(k, t)$ is calculated in the Appendix 1:

$$\Psi(k, t) = e^{-(k/2)(\alpha - \alpha e^{-rt} + (t/K))}\left(\frac{2K_c}{k} - \frac{\mathrm{U}(1, 1 - (k/2Kr), \alpha(k/2))}{r}\right) + \frac{1}{r}\mathrm{U}\left(1, 1 - \frac{k}{2Kr}, \frac{\alpha k e^{-rt}}{2}\right), \quad (8)$$

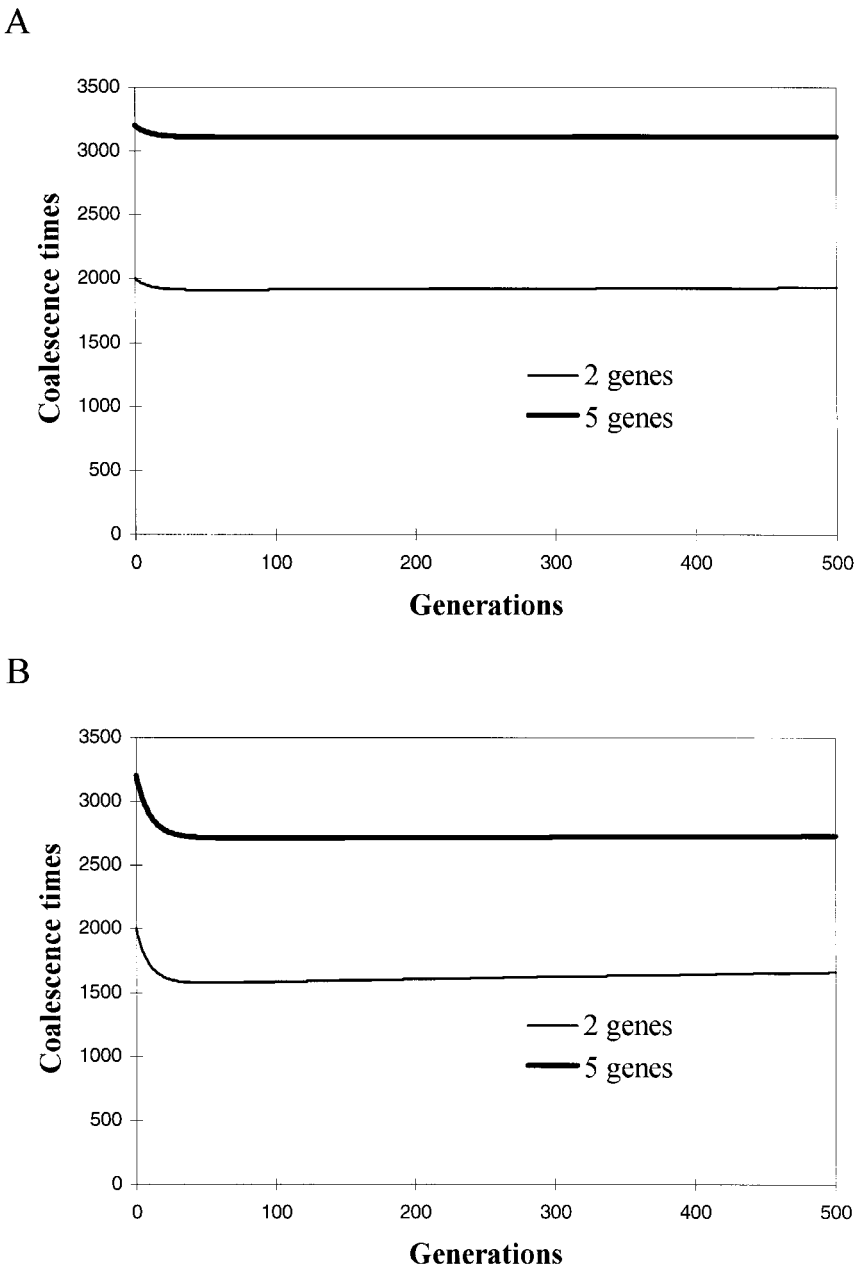


FIG. 2. Expectation of coalescence times for a sample of two or five genes during the 500 first generations for a population following logistic growth after a bottleneck event. The parameters are $K_c = K = 1000$, $K_0 = 100$ (A) or $K_0 = 20$ (B), $r = 0.01$.

where U is the confluent hypergeometric function (see Abramowitz and Stegun, 1964, pp. 504–505) defined by

$$\mathrm{U}(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} \, \mathrm{d}t,$$

and so (4) and (6) allows us to calculate $\mathrm{E}(Tc_{i,t})$ and $\mathrm{E}(L_{i,t})$.

2.2. Results

We can plot the expectations of coalescence times (Fig. 2) or of total three lengths (Fig. 3) against the number of generations since the bottleneck event. We see in every case a decrease in the coalescence times, which then increase again to reach their equilibrium value. This can be understood as a founder effect: the probability for two or more genes to have an ancestor early in the past

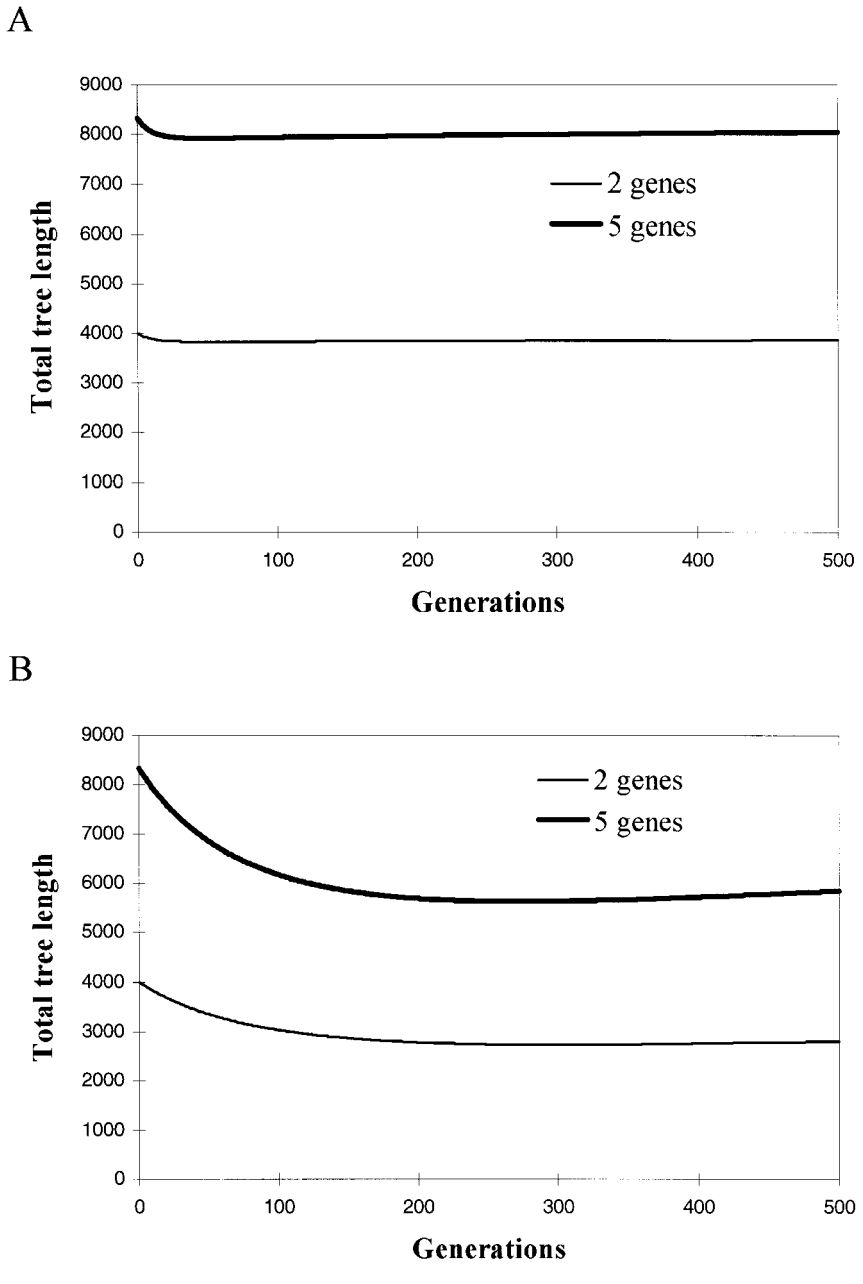


FIG. 3. Expectation of total tree length for a sample of two or five genes during the 500 first generations for a population following logistic growth after a bottleneck event. The parameters are $K_c = K = 1000$, $K_0 = 100$, $r = 0.1$ (A) or $r = 0.01$ (B).

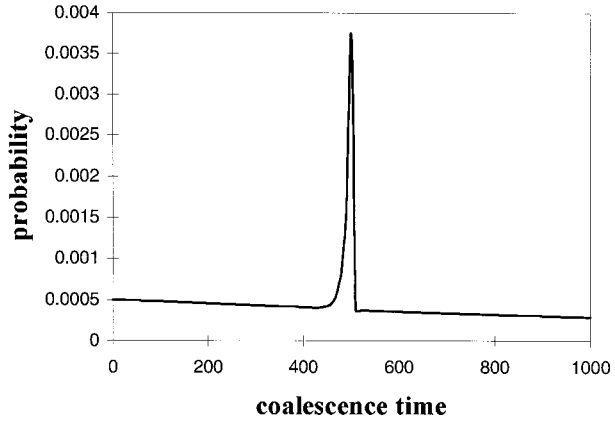
is higher if the population is smaller and therefore the diversity is reduced.

We found results similar to those of Nei *et al.* (1975) for the average heterozygosity. The decrease in coalescence times is lower if the growth rate r or the minimum population size K_0 is higher. It is interesting to notice that K_0 has no influence on the time at which the minimum is reached, whereas this decreases with r.

An other interesting point is that the reduction in coalescence times or total tree lengths is higher in absolute value for a sample of five genes than for a sample of two. The parameters K_0 and r also have a greater effect for a larger sample.

If we look at the distribution of the times separating two nodes of the tree (Fig. 4 gives an example for two genes), we see a peak in the distribution corresponding to

A



B

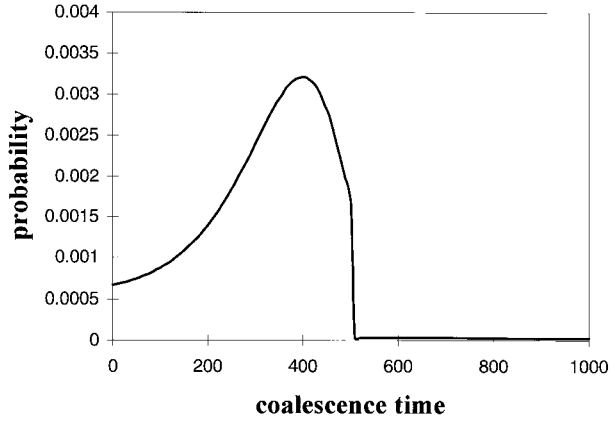


FIG. 4. Distribution of coalescence times for a population 500 generations after the bottleneck event. The parameters are $K_c = K = 1000$, $K_0 = 20$, $r = 0.1$ (A) or $r = 0.01$ (B).

the coalescence events that would have occurred during the growth period. This is important because, as we will see later, this peak can be seen also in the number of pairwise differences between sequences.

It is interesting to compare our curves to the Poisson-like given by Slatkin and Hudson (1991), which pertain to single populations in a long period of exponential growth, or those of Rogers and Harpending (1992) or Rogers (1995), where the populations experience a rapid change in size. As in their case we find a peak corresponding to the period of expansion of the population. The difference is that because of the instantaneous decrease of the population at the instant of the bottleneck, the right corner of the peak is very sharp in our case.

3. THE COLONIZATION PROCESS

3.1. Model

We use a linear stepping stone model with a finite number of demes d and nonoverlapping generations. Individuals are diploid and each deme is panmictic. The genes are selectively neutral. Every deme sends migrants to the demes on the left and right, except those at the ends of the metapopulation. At the beginning, we have a first deme at a border that is supposed to be full and at equilibrium, which means that its size is at the carrying capacity K and that the distribution of coalescence time follows a geometric law with parameter $1/(2K)$ (for a review, see Hudson, 1990). The other demes are empty and are colonized one after another.

The demography of each deme is supposed to be discrete logistic, i.e. the population size at time t is obtained from the one at time $t-1$ following

$$N_t = \mathrm{f}(N_{t-1}) = N_{t-1} + rN_{t-1}(1 - N_{t-1}/K), \qquad (9)$$

where r is the growth rate and K the carrying capacity. This discrete logistic growth is correctly approximated by the continuous one we used in the previous section as long as r is not too large (approximately $r < 2$). We define $N_{i,t}$ as the size of deme i at time t. $N_{i,t}$ are real numbers. In each generation, the population size in each deme changes, following (9), and then a proportion m of migrants is sent in the two neighboring demes equitably (except for the two demes of the border that send only a proposition of $m/2$ migrants). The laws that determine the variation of the size of the demes are obtained as follows:

$$N_{1,t} = (1 - m/2)\, f(N_{1,t-1}) + (m/2)\, f(N_{2,t-1}), \qquad (10a)$$

$$N_{i,t} = (1 - m)\, f(N_{i,t-1}) + (m/2)\, f(N_{i-1,t-1}) + (m/2)\, f(N_{i+1,t-1}) \qquad (1 < i < d), \qquad (10b)$$

$$N_{d,t} = (1 - m/2)\, f(N_{d,t-1}) + (m/2)\, f(N_{d-1,t-1}). \qquad (10c)$$

Actually the $N_{i,t}$ can be less than one. In this case the deme does not send any migrants. Therefore a deme is colonized only when the size of the preceding deme exceeds one.

3.2. Iterative Method

Our method is based on the use of the recurrence equations, as in Li (1977). He gave theoretical results

for average pairwise differences between sequences in a single population, not submitted to demographic changes. We adapted this method to the coalescence times of a system of several demes with migration and changes in the size of the demes.

We define $x_{i,j,t}$ as the proportion, at time t, of individuals of deme i whose parents were in deme j the generation before, the values of $x_{i,j,t}$ can be calculated as follows:

$$x_{i,i,t}=\frac{f(N_{i,t-1})\times(1-m)}{N_{i,t}} \quad (1<i<d), \tag{11a}$$

$$x_{1,1,t}=\frac{f(N_{1,t-1})\times(1-m/2)}{N_{1,t}}, \tag{11b}$$

$$x_{d,d,t}=\frac{f(N_{d,t-1})\times(1-m/2)}{N_{d,t}}, \tag{11c}$$

$$x_{i,i+1,t}=\frac{f(N_{i+1,t-1})\times(m/2)}{N_{i,t}} \quad (i<d), \tag{11d}$$

$$x_{i,i-1,t}=\frac{f(N_{i-1,t-1})\times(m/2)}{N_{i,t}} \quad (i>1), \tag{11e}$$

$$x_{i,j,t}=0 \quad (j<i-1 \text{ or } j>i+1). \tag{11f}$$

We use $T_{i,j,t}$ to denote the coalescence time of two genes from demes i and j taken t generations after the beginning of the colonization and $p_{i,j,t}(s)=P(T_{i,j,t}=s)$ to denote the distribution of $T_{i,j,t}$. The gene coming from i and the gene from j can coalesce in the previous generation only if the parents of both genes were in the same deme at this generation.

For $s=1$, the equation is:

$$p_{i,j,t}(1)=\sum_{k=1}^{d} x_{i,k,t}\times x_{j,k,t}\times\frac{1}{2N_{k,t-1}}, \tag{12a}$$

where $x_{i,k,t}\times x_{j,k,t}$ is the probability that the parent of the gene of i and the parent of the gene of j were in k, and $1/2N_{k,t-1}$ is the probability that these parents were the same individual.

For $s>1$ we have a recurrence equation:

$$p_{i,j,t}(s)=\sum_{k=1}^{d}\sum_{l=1}^{d} x_{i,k,t}\times x_{j,l,t}\times p_{k,l,t-1}(s-1)\times\left(1-\frac{\delta_{k,l}}{2N_{k,t-1}}\right), \tag{12b}$$

where $\delta_{k,l}=1$ if $k=l$ (i.e. if the parent belong to the same deme) and $\delta_{k,l}=0$ if $k\neq l$ and $x_{i,k,t}\times x_{j,l,t}$ is the probability that the parent of the gene of i was in k and that the parent of the gene of j was in l. $(1-(\delta_{k,l}/2N_{k,t-1}))$ is the probability that these parents were not the same individual. (If the parents were not in the same deme, they could not be the same and if they were, the probability that they were the same is as given previously).

A time-homogeneous Markov chain is obtained. From (12a) and (12b), we can derive the recursion equations for the expectations of the coalescence times, $\mathrm{E}(T_{i,j,t})$:

$$\mathrm{E}(T_{i,j,t})=1+\sum_{k=1}^{d}\sum_{l=1}^{d} x_{i,k,t}\times x_{j,l,t}\times\mathrm{E}(T_{k,l,t-1})\times\left(1-\frac{\delta_{k,l}}{2N_{k,t-1}}\right). \tag{13}$$

The distributions and expectations are calculated at each step of the process only in the demes where the population size exceeds one.

The number of differences accumulated between two sequence is linked to their coalescence time. Let $S_{i,j,t}$ be the number of differences between two sequences chosen respectively in demes i and j at time t; in an infinite-site model, for a given coalescence time $T_{i,j,t}$, $S_{i,j,t}$ has a Poisson distribution with mean $2\mu T_{i,j,t}$, where μ is the average number of mutation events per generation that occur in each lineage (see Hudson, 1990). The distribution of $S_{i,j,t}$ can therefore be deduced from the distribution of $T_{i,j,t}$:

$$P(S_{i,j,t}=k)=\sum_{s=0}^{\infty}\mathrm{Poiss}_{2\mu s}(k)\times p_{i,j,t}(s), \tag{14}$$

where Poiss_{θ} is the Poisson distribution with parameter θ.

The expectation of $S_{i,j,t}$ is given by

$$\mathrm{E}(S_{i,j,t})=2\mu\mathrm{E}(\mathrm{T}_{i,j,t}). \tag{15}$$

The F_{st} (or the R_{st}) can be calculated according to Slatkin (1991), in an infinite-allele or K-alleles model (provided the mutation rate is much less than one):

$$F_{st}=(\overline{T}-\overline{T_0})/\overline{T}, \tag{16}$$

where $\overline{T}$ is the average coalescence time of two genes of the whole metapopulation and $\overline{T_0}$ is the average intra-deme coalescence time.

Analytical results cannot be determined for the colonization period, because of the time dependency of the transition matrix of the Markov chain. Therefore we have written a C program that calculates the distributions and the expectations of coalescence times generation after

generation, using (12a), (12b) and (13), starting with a situation of a unique deme at one border, genetically at equilibrium, with a geometric distribution of pairwise coalescence times.

3.3. Equilibrium Values

Analytical results can be obtained, using the method given in Hey (1990). At equilibrium (12a) and (12b) become (the quantities without the time index are the equilibrium values):

$$p_{i,j}(1) = \sum_{k=1}^{d} x_{i,k} \times x_{j,k} \times \frac{1}{2K}, \tag{17a}$$

$$p_{i,j}(s) = \sum_{k=1}^{d} \sum_{l=1}^{d} x_{i,k} \times x_{j,l} \times p_{k,l}(s-1) \times \left(1 - \frac{\delta_{k,l}}{2K}\right) \quad (s>1). \tag{17b}$$

(17b) can be written in matrix form. We denote by $\mathbf{p}(s)$ the $d \times d$ matrix of the $p_{i,j}(s)$, $\mathbf{p_0}(s)$ the $d \times d$ matrix that has the $p_{i,i}(s)$ values on its diagonal and 0 elsewhere. $\mathbf{x}$ is the $d \times d$ matrix of $x_{i,j}$ ($^t\mathbf{x}$ means the transposed matrix of $\mathbf{x}$):

$$\mathbf{p}(s) = \mathbf{x} \cdot (\mathbf{p}(s-1) - \mathbf{p_0}(s-1)/2K) \cdot {}^t\mathbf{x}. \tag{18}$$

(13) becomes:

$$\mathrm{E}(T_{i,j}) = 1 + \sum_{k=1}^{d} \sum_{l=1}^{d} x_{i,k} \times x_{j,l} \times \mathrm{E}(T_{k,l}) \times \left(1 - \frac{\delta_{k,l}}{2K}\right), \tag{19}$$

which can be written in matrix form:

$$\mathrm{E}(\mathbf{T}) = \mathbf{1} + \mathbf{x} \cdot (\mathrm{E}(\mathbf{T}) - \mathrm{E}(\mathbf{T_0})/2K) \cdot {}^t\mathbf{x}, \tag{20}$$

where $\mathbf{T}$ is the $d \times d$ matrix of the $T_{i,j}$ and $\mathbf{T_0}$ the $d \times d$ matrix that has the $T_{i,i}$ values on its diagonal and 0 elsewhere. $\mathbf{1}$ is a matrix whose elements are 1. The quantities without the time index are the equilibrium values. At demographic equilibrium, the matrix $\mathbf{x}$ depends only on m. For example its expression is as follows, given $d = 5$:

$$\begin{pmatrix} 1-m/2 & m/2 & 0 & 0 & 0 \\ m/2 & 1-m & m/2 & 0 & 0 \\ 0 & m/2 & 1-m & m/2 & 0 \\ 0 & 0 & m/2 & 1-m & m/2 \\ 0 & 0 & 0 & m/2 & 1-m/2 \end{pmatrix}.$$

We can solve formally the recursion relation between the $\mathbf{p}(s)$ and $\mathbf{p}(s-1)$ given in (18). To do that the matrix $\mathbf{p}(s)$ has to be written as a one dimensional vector and then the transition matrix that gives the relationship between $\mathbf{p}(s)$ and $\mathbf{p}(s-1)$ can be diagonalised. This gives us an analytical result for the $p_{i,j}(s)$. The values of the $\mathrm{E}(T_{i,j})$ can be found by solving the linear system given by (12). The details of the calculations are given in Appendix 2.

4. RESULTS FOR THE EQUILIBRIUM VALUES

The equilibrium values obtained formally for the expectations of the coalescence times were in agreement with that obtained via the iteration process (see Table 1 for a comparison). We can see that the values of the intra-deme coalescence times are higher when those demes are in the middle of the population. This is due to the limitation imposed by the borders of the population, since individuals in these demes have fewer immigrants in their ancestry.

These values allow us to calculate the F_{st} (or R_{st}) values at equilibrium, using (16). For instance, the result for 10 demes is:

$$F_{st} = \frac{\begin{pmatrix} 1 + 10.1Km + 36K^2m^2 \\ +53.3K^3m^3 + 26.9K^4m^4 \end{pmatrix}}{\begin{pmatrix} 1 + 11.5Km + 50.1K^2m^2 + 102K^3m^3 \\ +96.7K^4m^4 + 32.6K^5m^5 \end{pmatrix}}.$$

The analytic formulas of the equilibrium distribution for systems with a low number of demes can be calculated using the method given in the previous section. For instance for two demes, we can obtain a discrete version of equation (25) in Takahata (1988):

$$p_{11}(s) = p_{22}(s) = \frac{\begin{pmatrix} (\gamma+1)(1-2m-1/(4K)-\gamma/(4K))^s \\ +((\gamma-1)(1-2m-1/(4K)+\gamma/4K))^s \end{pmatrix}}{4K\gamma},$$

$$p_{12} = \frac{\begin{pmatrix} 2m((1-2m-1/(4K)+\gamma)^s \\ -(1-2m-1/(4K)-\gamma)^s) \end{pmatrix}}{\gamma},$$

with $\gamma = \sqrt{1 + 64N^2m^2}$.

TABLE 1

Formal Equations for the Intra- and Inter-deme Coalescence Times at Equilibrium, for $d = 5$, and Comparison of a Numerical Example with the Results Obtained by Iteration

Couples of demes	Formal equations of the coalescence times	Example with $N = 100$ and $m = 0.001$	Iteration with the same values
1-1 and 5-5	$\frac{69N + 460N^2m + 750N^3m^2}{11 + 61Nm + 75N^2m^2}$	8700.3	8700.68
1-2 and 4-5	$\frac{47 + 476\,Nm + 1520\,N^2m^2 + 1500\,N^3m^3}{2m(11 + 61Nm + 75N^2m^2)}$	12047.3	12051
1-3 and 3-5	$\frac{97 + 782Nm + 1970N^2m^2 + 1500N^3m^3}{2m(11 + 61Nm + 75N^2m^2)}$	14788.7	14792.5
1-4 and 2-5	$\frac{137 + 1011Nm + 2270\,N^2m^2 + 1500\,N^3m^3}{2m(11 + 61Nm + 75N^2m^2)}$	16724	16727.9
1-5	$\frac{159 + 1133Nm + 2420N^2m^2 + 1500\,N^3m^3}{2m(11 + 61Nm + 75N^2m^2)}$	17724	17727.9
2-2 and 4-4	$\frac{132N + 685N^2m + 750N^3m^2}{11 + 61Nm + 75N^2m^2}$	10660.1	10659.9
2-3 and 3-4	$\frac{63 + 615Nm + 1820N^2m^2 + 1500N^3m^3}{2m(11 + 61Nm + 75N^2m^2)}$	13594.8	13598.6
2-4	$\frac{111 + 874Nm + 2120N^2m^2 + 1500N^3m^3}{2m(11 + 61Nm + 75N^2m^2)}$	15659.4	15663.3
3-3	$\frac{158N + 760N^2m + 750N^3m^2}{11 + 61Nm + 75N^2m^2}$	11279.3	11278.9

These distributions are linear combinations of two geometric distributions, the first with both coefficients positive, the second with one coefficient positive and one negative. The inter-deme coalescence time shows that the probability of a small coalescence time is very low, due to the separation of the demes. A graphical representation is given in Fig. 5.

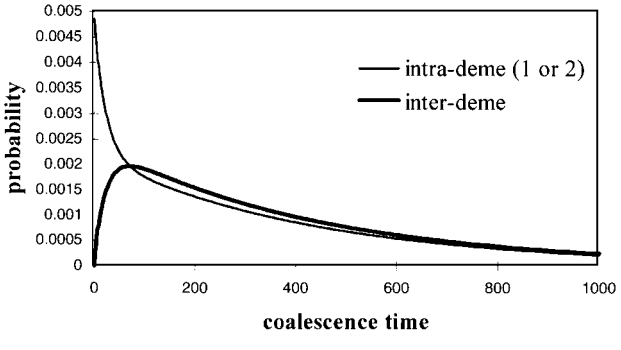


FIG. 5. Distributions of coalescence times for two genes sampled either in the same deme or one in each deme in a two-deme system ($d = 2$) with $K = 100$, $m = 0.01$. These distributions are linear combinations of geometric ones.

5. IMPACT OF THE COLONIZATION PROCESS

5.1. Intra-deme Coalescence Times

As in the case of an isolated population under logistic growth after a bottleneck, the curves of intra-deme coalescence times show clearly a founder effect, which consists of a decrease in the coalescence times inside a deme during the first generations after the beginning of colonization, then an increase to equilibrium. The founder effect becomes stronger with the distance from the first deme.

The coalescence time decreases more sharply in each deme as the migration rate m is lowered (see Fig. 6), consequently the loss of diversity in each deme is greater. The difference can be very important, showing non-linear effects of the impact of m. The minimum average intra-deme coalescence time is about 10 generations for the fifth deme for $K \cdot m = 0.1$ and about 700 generations for $K \cdot m = 10$.

The effect of the growth rate is not always the same. In some cases, the minimum average coalescence time

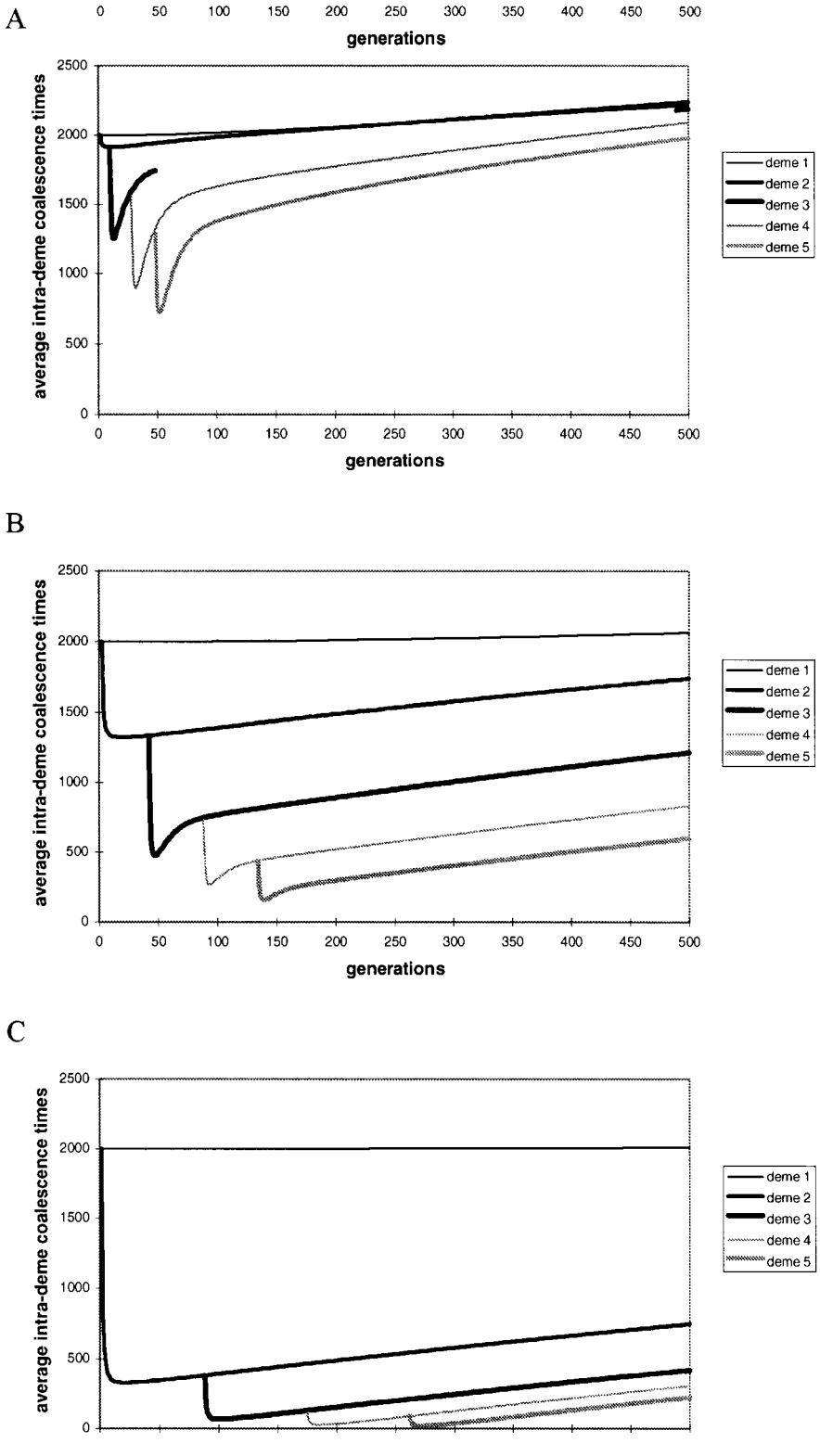


FIG. 6. Impact of the migration rate on average intra-deme coalescence times. The parameters are: $d = 5$, $K = 1000$, $r = 0.1$ and $m = 0.01$ (A), $m = 0.001$ (B), $m = 0.0001$ (C). A higher migration rate decreases the founder effect inside the demes.

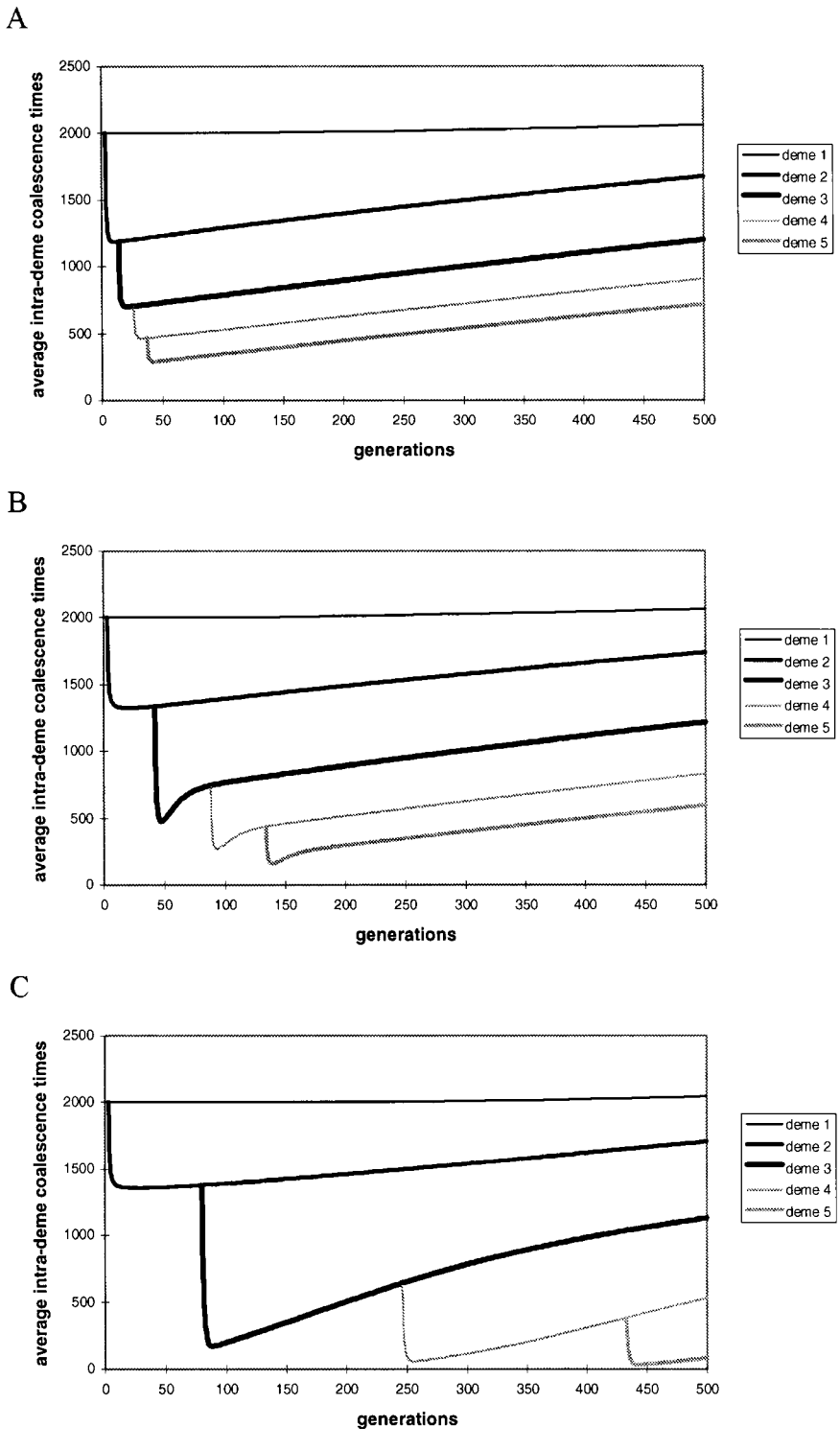


FIG. 7. Impact of the growth rate on average intra-deme coalescence times. The parameters are: $d = 5$, $K = 1000$, $m = 0.001$ and $r = 1$ (A), $r = 0.1$ (B), $r = 0.01$ (C). In this case, except for deme two, the founder effect is weaker when r increases.

decreases slightly with r, whereas in most cases it increases at a rate that can be quite high (see Fig. 7). The results are different for deme two, showing a slight decrease of the minimum coalescence time with r when the migration rate is high and an increase when it is low.

5.2. Evolution of the F_{st} (*or the* R_{st})

Using (16) we followed the evolution of the F_{st} throughout the colonization process. Examples for F_{st} are given in Figs. 8 and 9, for different parameter sets. During the colonization period, the F_{st} increases because the number of demes is increasing and because the migrants founding the new demes have less and less variability, so new demes differ more from the average deme. The sawtooth effect that can be seen during this phase comes from the sudden increase caused by the foundation of a new deme which is then followed by a short homogenization by migration until a new deme is founded. In the real world this should also happen although it should be less regular. Then after the first migrant has arrived in the last deme, the F_{st} decreases, corresponding to homogenization by migration. This process continues long after all demes are full. Finally the value of the F_{st} increases again, towards its equilibrium value.

The comparison of the curves during and after the colonization for different values of the migration rate m (see Fig. 8) shows that the maximum that F_{st} reaches at the end of the colonization process decreases with m.

If we make the same comparison for different values of the growth rate r (see Fig. 9), we can see that this parameter has far less impact on the maximum F_{st} value that is reached, and the effect is not clear: in our example these maximum values are 0.4 for $r = 0.01$, 0.37 for $r = 0.1$ and

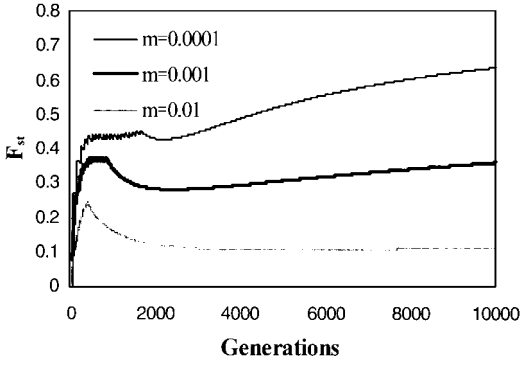


FIG. 8. Impact of the migration rate on the F_{st}. The shape of the curves during the colonization period is due to successive foundation events. Parameters are $d = 20$, $K = 1000$, $r = 0.1$, $m = 0.0001$, $m = 0.001$ or $m = 0.01$.

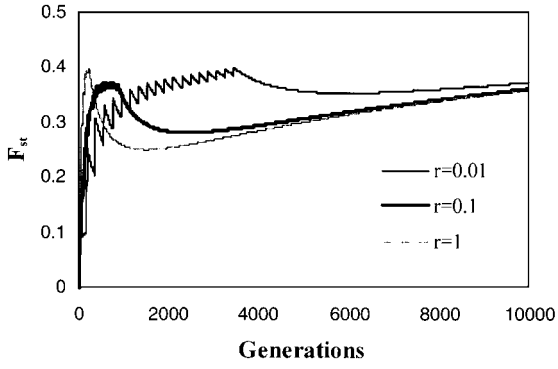


FIG. 9. Impact of the growth rate on the F_{st}, $d = 20$, $K = 100$, $m = 0.001$, $r = 0.01$, $r = 0.1$ or $r = 1$.

0.4 for $r = 1$. Afterwards the decrease of the F_{st} is much stronger when r is large. Then the curves rejoin to rise to an equilibrium independent of r.

5.3. Evolution of the Distributions of Coalescence Times

The distribution of coalescence times allows us, in simple cases, to retrace the history of a deme. We show here the results for a system of two demes, the first one colonizing the second. The distribution of coalescence times in the donor deme is not modified at the beginning of the process, whereas the newly colonized deme shows clearly the impact of the founder effect. Deme two is full 87 generations after the arrival of the first migrant with $r = 0.01$ and after 47 generations with $r = 0.1$. At the opposite extreme, 3500 generations are needed for the distribution of coalescence times to reach its equilibrium. Therefore we can conclude that the effect of colonization can be seen on the genetic level long after demography is at equilibrium.

The colonization has a rapid effect on coalescence times of the donor deme. For instance with $r = 0.1$, the distribution of the intra-deme coalescence times in the donor deme seems to be already slightly affected at $t = 40$ generations (Fig. 10), rather early after colonization of deme two, where the distribution continues to show the consequences of the founder effect with a high probability of a very short coalescence times. For the inter-deme coalescence time, we can see that a short coalescence time becomes less and less probable, indicating the separation of the demes. Continuing to follow the evolution of the distribution, we can see that very shortly (about 200 generations after the beginning of colonization) the distributions in deme one and two become similar. Both

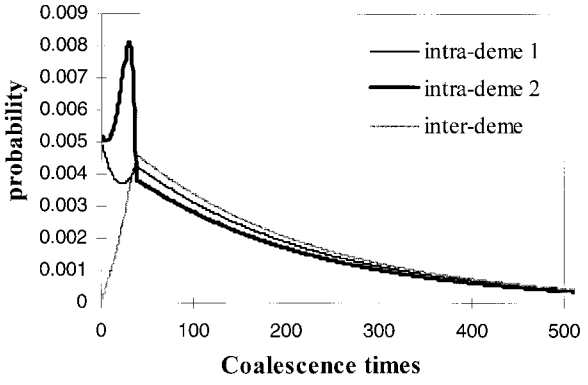


FIG. 10. Distributions of intra- and inter-deme coalescence times (s), 40 generations after the beginning of colonization, with $d=2$, $K=1000$, $m=0.01$, $r=0.1$.

intra- and inter-deme coalescence times show the impact of the colonization period. For instance, the curves given in Fig. 11, which correspond to 300 generations after the beginning of the colonization, can be divided in three parts. The first part ($s<s_e$) corresponds to values for which coalescence occurred after the end of the colonization. This portion of the curves resembles the equilibrium curves (see the section equilibrium values). The second part ($s_e<s<s_b$) corresponds to events of coalescence that would have occurred during the colonization period. We can see that there is an increase in the probability of coalescence during this period, and that this effect is maintained hundreds of generations after this period is over. The third part of the curve ($s>s_b$) corresponds to coalescence occurring in deme one before colonization. Therefore this part of the curves resembles the initial curve of deme one. The peak associated with high probability of coalescence times becomes smaller when we look at the curves obtained more generations after the colonization process.

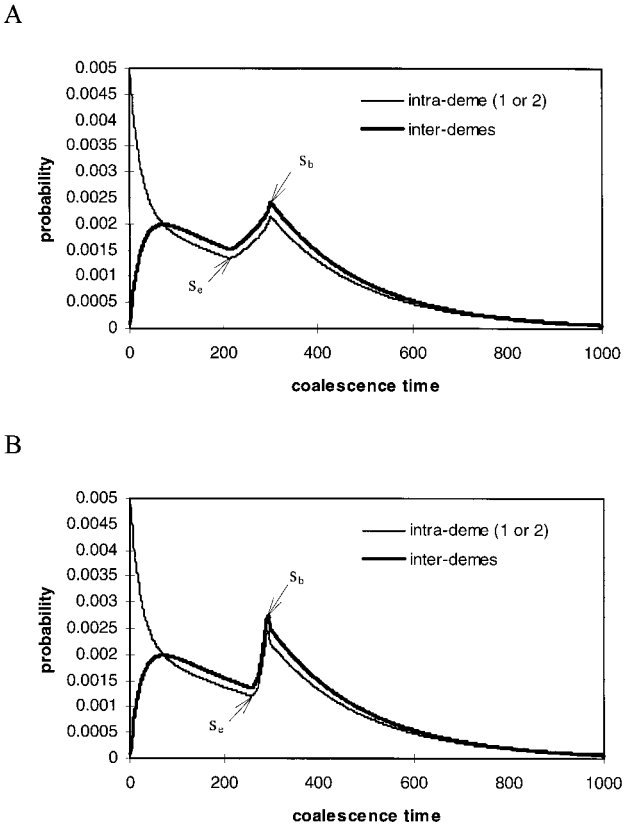


FIG. 11. Distributions of intra- and inter-deme coalescence times (s), 300 generations after the beginning of colonization, with $d=2$, $K=1000$, $m=0.01$, and $r=0.01$ (A) or $r=0.1$ (B). s_b indicates coalescence occurring at the beginning of the colonization and s_e at the end of it. The peaks correspond clearly to the colonization period.

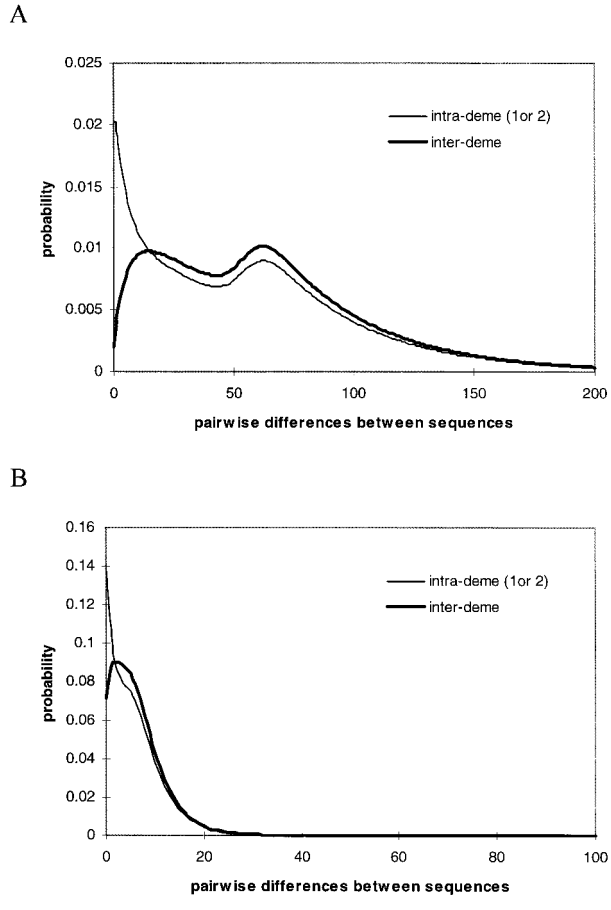


FIG. 12. Distributions of intra- and inter-deme pairwise differences between sequences, 300 generations after the beginning of colonization, with $d=2$, $K=1000$, $m=0.001$, $r=0.01$, $\mu=0.01$ (A) or $\mu=0.001$ (B). With a rather high mutation rate (about $K\mu=10$), colonization process is evident, otherwise the differentiation between the different alleles is too limited to allow detection of the impact of colonization, which will be hidden by the variability generated by the stochasticity of the process.

The shape of the peak is dependent on the speed and synchrony of colonization. When population growth rate after colonization is small, this peak is much broader and a little lower than when colonization is faster (see Fig. 11). This difference holds until the equilibrium curve is reached.

5.4. Distribution of Average Pairwise Genetic Differences

The expectation of S, the number of differences between two sequences, can be deduced directly from the expectation of the coalescence times, using (15). So the results that we obtained concerning the coalescence times can be applied to genetic diversity as well, both at the intra- and inter-deme level. The effect of parameters such as the growth rate or the migration rate will be the same, as long as we are interested only in the expectations.

With (14) and the distribution of coalescence times, we can obtain the distribution of S, for two genes located either in the same deme or in different demes. We give an example in Fig. 12. With $K\mu = 10$ (i.e. the product of the number of individuals in each deme at equilibrium by the average number of mutations in a sequence between two generations) this distribution is approximately the same as the distribution of coalescence times, but with $K\mu = 1$ the range of S is too limited to have a visible impact on coalescence times.

6. DISCUSSION AND CONCLUSION

6.1. Intra-deme Coalescence Times

Recurrence methods allow us to study the impact of colonization on the coalescence times and consequently on diversity during and after the colonization period. We have seen that the migration rate (Fig. 6) and the growth rate (Fig. 7) both have an influence on the founder effect, which can produce a very sharp decrease in diversity. The decrease is higher with a low migration rate, which was an expected result: if the newly founded demes receive more migrants during the first generation after the founder event, their individuals will have a pattern of ancestry closer to that of the donor deme.

Population growth rate has a more complicated effect, generally decreasing the founder effect when the growth rate is higher. On one hand, with a large growth rate, the descendants of the first migrant(s) will rapidly fill the new deme and therefore overwhelm the effect of further migrants. Consequently the coalescence time should decrease more sharply, because all the inhabitants should be the descendants of a few individuals.

On the other hand, a higher growth rate allows the deme to grow faster, which, by reducing drift, reduces the probability of a short coalescence time. Moreover, when a deme i is colonized, the deme $(i-1)$ that sends the migrants is still in a transition state, where its own coalescence times and size are changing. Therefore it sends more and more migrants each generation, thus increasing the coalescence times. Therefore there is a combination of effects that increases the diversity of the newly founded deme which receives more and more migrants with a pattern of ancestry that goes deeper into the past.

These effects usually overwhelm the others. As we have shown in Section 2, even for an isolated population subjected to a bottleneck, a high growth rate after the bottleneck limits the loss of diversity. Therefore we can conclude that at least in some cases, the simple fact that the demes have reached a large population size more rapidly is enough to reduce the founder effect.

6.2. Evolution of the F_{st}

The first maximum reached (see Fig. 8) decreases when the migration rate increases. This can be understood in view of the fact that with a low migration rate, demes have fewer founders and consequently new demes differ more from the old ones, and they do not have the time during the first phase to reduce this differentiation by migration. The later homogenization is stronger with a high migration rate, i.e., the decrease in inter-deme differentiation is very strong for a high value of m, making the newly created deme similar to the old ones.

Concerning the effect of the growth rate, there is as in Section 6.1 a compensation of effects (see Fig. 9) that makes the maximum reached at the end of the colonization independent of r. Then homogenization is higher when the growth rate is low, which can be attributed to the fact that the demes are growing faster and consequently exchange more migrants before they reach their equilibrium size.

6.3. Distributions of Coalescence Times and Pairwise Differences

In our example with two demes, we have seen that the shape of the intra-deme distribution in the donor deme

is very rapidly affected by the colonization process (Fig. 10). And we have shown that these distributions become similar to the distribution of the receiver deme (Fig. 11). This means that colonization has a strong effect on both demes and that it is not possible to tell which deme colonized the other. The peak we obtain is consistent with the one obtained from the analytical results of Section 2, where we also found a peak whose shape depended on the colonization speed. This is to be compared with Tajima's (1989) results for an equilibrium two-subpopulation model, where he shows that the distribution of the number of differences between two sequences chosen in the same deme is independent of the migration rate. On the contrary, in the case of a transitory colonization process, this distribution is disturbed by the colonization process.

6.4. Detecting a Colonization Process in Natural Populations

Our results tend to suggest that colonization processes could be detected by the analysis of the structure of the diversity of neutral markers. As we have seen high F_{st} can be an indication of a colonization period but this is not always the case provided that there is sufficient gene flow from one population to another. Nonetheless, if gene flow is high, the F_{st} will be lower than with more restricted gene flow but nevertheless higher than the expected equilibrium value obtained for a system of stable populations.

A decrease in diversity along a given spatial axis will also be a good indicator that a colonization took place. If the more diverse populations are at an end of the axis rather than in the middle, this would provide a clear indication of the non equilibrium state of the populations. Variation in diversity must be tested using a rather large number of markers because of the stochasticity of the process and because some markers can be, directly or indirectly through hitchhiking, subjected to environmental selection that can generate particular patterns of diversity.

The fact that the time needed to reach the genetic equilibrium is much higher than the time necessary to reach the demographic equilibrium shows us that results concerning populations that are supposed to be at equilibrium must be evaluated with caution.

Experimental distributions of the pairwise number of differences between sequences could be used to determine if an event of colonization occurred inside a group of demes. We must as before be however very careful when using this kind of approach, because, as before, other phenomena, like bottlenecks or hitchhiking effects, could generate distributions which would share some resemblance with the ones we obtained, especially if only results for a few individuals at a few loci are available.

It is however important to notice that it is necessary to dispose of highly variable or very long sequences, or to work with populations large enough, in order to fulfill the condition $K\mu = 10$. Takahata and Slatkin (1990) have found similar results for their method of reconstructing the history of samples for two populations that both descend from an ancestral population which divided itself into two.

6.5. Analysis of Experimental Results

Hamrick *et al.* (1992) point out that the F_{st} based on allozymes is small for most forest trees species, which he attributes to a high level of gene flow and low mutation rate. On the other hand the global genetic diversity of trees is larger than for other species. This indicates, according to our results, a rather high level of migration. We have to take into consideration that the pollen exchange considerably increases gene flow, which reduces F_{st} and limits the effects of successive founder events, allowing the maintenance of a high level of genetic diversity.

Demelo and Hebert (1994) give numerical values for the intra-deme diversity and F_{st} in a species of the freshwater invertebrate *Bosmina coregoni*. Their results show rather high F_{st} values (0.18 to 0.36) but not a great loss of diversity within the populations. This is consistent with our results, provided the migration rate is sufficiently high. In this case we have shown that the F_{st} increases to a value which is far higher than the equilibrium value but that the diversity does not decrease much in the newly founded populations.

Many other factors could be taken into consideration, for instance gene flow mediated by pollen or selective effects. The iteration approach developed here should be applicable to those situations and allow, thanks to the rapidity of this method, an exhaustive study of the impact of different phenomena. It might also be useful to work with the distributions of coalescence times in samples of more than two genes for the general model as we did for the isolated population suffering a bottleneck. This would be useful because the pairwise coalescence times of genes inside a sample are correlated. The problem is that the transition matrices become very difficult to determine in this case.

APPENDIX 1

Derivation of the Function $\Psi(k, t) = \int_0^\infty \exp(-k\Lambda_t(s))\,ds$

In order to calculate $\Psi(k, t)$, we have to decompose it into two parts: $I_1 = \int_0^1 \exp(-k\Lambda_t(s))\,ds$ and $I_2 = \int_t^\infty \exp(-k\Lambda_t(s))\,ds$. The first part, I_1, can be calculated as follows:

$$I_1 = \int_0^t \exp\left(-\frac{k}{2}\left(\frac{s}{K} + \left(\frac{1}{K_0} - \frac{1}{K}\right)\frac{e^{-rt}(e^{rs}-1)}{r}\right)\right) ds.$$

We denote $\alpha = 1/r(1/K_0 - 1/K)$ and make the variable transform $u = e^{rs}$:

$$I_1 = \frac{\exp((\alpha k e^{-rt})/2)}{r} \int_1^{e^{rt}} u^{-1-(k/2Kr)} \exp\left(-\frac{\alpha k e^{-rt}}{2} u\right) du.$$

Defining $\beta = (\alpha k e^{-rt})/2$, $c = k/2Kr$ and making the new variable transform $v = \beta\mu$:

$$I_1 = \frac{e^\beta}{r}\beta^{-c} \int_\beta^{\beta e^{rt}} v^{-1-c} e^{-v}\, dv.$$

According to Gradshteyn and Ryshik (1980, p. 318),

$$\int_u^\infty v^{-a} e^{-v}\, dv = W_{(-(a/2),\,((1-a)/2))}(u),$$

where W is Whittaker's function (see Abramowitz and Stegun, 1964, p. 505). So I_1 can be written as:

$$I_1 = \frac{\beta^c e^\beta}{r}\left(\beta^{-((c+1)/2)} e^{-(\beta/2)} W_{(-((c+1)/2),\,-(c/2))}(\beta) - (\beta e^{rt})^{-((c+1)/2)} e^{-(\beta e^{rt}/2)} W_{(-((c+1)/2),\,-(c/2))}(\beta e^{rt})\right). \quad \text{(A1)}$$

According to Abramowitz and Stegun (1964, p. 505),

$$W_{(a,\,b)}(z) = e^{-z/2} z^{1/2+b} \mathrm{U}(1/2 + b - a,\, 1 + 2b,\, z).$$

where U is the confluent hypergeometric function defined by

$$\mathrm{U}(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1}\, dt,$$

and so substituting into (A1) and simplifying:

$$I_1 = -e^{-(k/2)(\alpha - \alpha e^{-rt} + (t/K))} \frac{\mathrm{U}(1,\, 1-(k/2Kr),\, (\alpha k/2))}{r} + \frac{1}{r}\mathrm{U}\left(1,\, 1 - \frac{k}{2Kr},\, \frac{\alpha k e^{-rt}}{2}\right).$$

I_2 can be calculated much more easily.

$$I_2 = e^{-(k/2)(\alpha - \alpha e^{-rt} + (t/K))} \frac{2K_i}{k},$$

and consequently

$$\Psi(k, t) = e^{-(k/2)(\alpha - \alpha e^{-rt} + (t/K))}\left(\frac{2K_i}{k} - \frac{\mathrm{U}(1,\, 1-(k/2Kr),\, (\alpha k/2))}{r}\right) + \frac{1}{r}\mathrm{U}\left(1,\, 1 - \frac{k}{2Kr},\, \frac{\alpha k e^{-rt}}{2}\right).$$

APPENDIX 2

Analytical Formulas for the Equilibrium Distributions and Expectations of Coalescence Times in a Linear Stepping-Stone Model with Finite Number of Demes

In order to calculate the formal values of $\mathbf{p}(s)$ at equilibrium, we have to transform this $d \times d$ matrix into a one-dimensional vector $\mathrm{p}(s)$. We take into consideration that there are some symmetries in our problem: $p_{i,j} = p_{j,i}$ and $p_{i,j} = p_{d+1-j,\,d+1-i}$. So $\mathrm{p}(s)$ can be defined as:

$$\mathrm{p}(s) = (p_{1,1}(s),\, p_{1,2}(s),\, ...,\, p_{1,d}(s),\, p_{2,2}(s),\, ...,\, p_{2,d-1}(s),\, ...,\, p_{(d+1)/2,\,(d+1)/2}(s)),$$

if d is even, and

$$\mathrm{p}(s) = (p_{1,1}(s),\, p_{1,2}(s),\, ...,\, p_{1,d}(s),\, p_{2,2}(s),\, ...,\, p_{2,d-2}(s),\, ...,\, p_{d/2,\,d/2}(s),\, p_{d/2,\,d/2+1}(s)),$$

if d is odd.

Let us denote by σ the size of $\mathrm{p}(s)$. $\sigma = (d+1/2)^2$ if d is even and $\sigma = d/2(d/2+1)$ if d is odd.

$\mathrm{p}(1)$ is obtained by saying that $p_{i,i}(1) = 1/(2K)$ and $p_{i,j}(1) = 0$ for $i \neq j$. We denote by S_u the set of pairs (i, j)

for which $p_u = p_{i,j}$ (p_u is the uth coordinate of p). S_u contains two to four elements.

(18) can be written in terms of p(s).

$p(s) = \mathbf{A} \cdot p(s-1)$, where $\mathbf{A}$ is a $\sigma \times \sigma$ transition matrix that can be deduced from the $x_{i,j}$. Its coefficients can be calculated as follows:

$$A_{u,v} = \sum_{(k,l) \in S_v} x_{i,k} x_{j,l} \left(1 - \frac{\delta_{k,l}}{2K}\right)$$

where (i, j) is whatever element of S_u.

This allows us to pack all the symmetries of the problem inside $\mathbf{A}$. Then we can diagonalize $\mathbf{A}$ and write it $\mathbf{A} = \mathbf{P} \cdot \mathbf{D} \cdot \mathbf{P}^{-1}$, where $\mathbf{D}$ is the diagonalised matrix of $\mathbf{A}$ and $\mathbf{P}$ is the transition matrix between $\mathbf{A}$ and $\mathbf{D}$. And the we can calculate $p(s)$:

$$p(s) = \mathbf{P} \cdot \mathbf{D}^{s-1} \cdot \mathbf{P}^{-1} p(1).$$

All this can be solved formally using a computer algebra program. Concerning the expectations of the coalescence times, (20) can be written: $E(T) = \mathbf{A} \cdot E(T) + 1$, where T is the vector of size s constructed from $\mathbf{T}$ in the same way that p(s) is constructed from $\mathbf{p}(s)$ and 1 is the vector of size σ whose all values equals 1. As previously this system can be solved formally.

ACKNOWLEDGMENTS

The analytical work was done while F.A. was visiting Montgomery Slatkin's group in the Department of Integrative Biology, University of California at Berkeley. We thank him for his helpful remarks and suggestions on this work as well as for his comments on an early version of this paper. We also thank N. Frascaria-Lacoste, D. Taneyhill, J. Shykoff, T. Wiehe, and three anonymous reviewers for their helpful comments, criticisms and suggestions.

REFERENCES

Abramowitz, M., and Stegun, I. A. 1964. "Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables," Washington, DC: U. S. Department of Commerce.

Bennett, K. D. 1983. Postglacial population expansion of forest trees in Norfolk, UK, *Nature* **303**, 164–167.

Demelo, R., and Hebert, P. D. N. 1994. Founder effects and geographical variation in the invading cladoceran *Bosmina* (*Eubosmina*) *coregoni Baird* 1857 in North America, *Heredity* **73**, 490–499.

Gradshteyn, I. S., and Ryzhik, I. W. 1980. "Tables of Integrals, Series and Products, Corrected and Enlarged Edition," Academic Press, New York.

Griffiths, R. C., and Tavaré, S. 1994. Sampling theory for neutral alleles in a varying environment, *Phil. Trans. R. Soc. Lond. B* **344**, 403–410.

Hamrick, J. L., Godt, M. J. W., and Sherman-Broyles, S. L. 1992. Factors influencing levels of genetic diversity in woody plant species, *New forests* **6**, 95–124.

Hey, J. 1990. A multi-dimensional coalescent process applied to multi-allelic selection models and migration models, *Theor. Pop. Biol.* **39**, 30–48.

Hudson, R. 1990. Gene genealogies and the coalescent process, *Oxford Surveys in Evolutionary Biology* **7**, 1–44.

Johnson, M. S. 1988. Founder effects and geographically variation in the land snail Theba pisana, *Heredity* **61**, 133–142.

Kingman, J. F. C. 1982. The coalescent, *Stochast. Proc. Appl.* **13**, 235–248.

Kremer, A. 1994. Diversité génotypique et variabilité des caractères phénotypiques chez les arbres forestiers, *Genet. Sel. Evol.* **26**, Suppl. 1, 105s–123s.

Li, W.-H. 1977. Distribution of nucleotide differences between two randomly chosen cistrons in a finite population, *Genetics* **85**, 331–337.

Marjoram, P., and Donnelly, P. 1994. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution, *Genetics* **136**, 673–683.

Nei, R., Maruyama, T., and Chakraborty, R. 1975. The bottleneck effect and genetic variability in populations, *Evolution* **29**, 1–10.

Nichols, R. A., and Hewitt, G. M. 1994. The genetic consequences of long distance dispersal during colonization, *Heredity* **72**, 312–317.

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

Rogers, A. R. 1995. Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608–615.

Rogers, A. R., and Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences, *Mol. Biol. Evol.* **9**, 552–569.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times, *Genet. Res. Camb.* **58**, 167–175.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies, *Genetics* **139**, 457–462.

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

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

Tajima, F. 1989. DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model, *Genetics* **123**, 229–240.

Takahata, N. 1988. The coalescent in two partially isolated diffusion populations, *Genet. Res. Camb.* **52**, 213–222.

Takahata, N., and Slatkin, M. 1990. Genealogy of neutral genes in two partially isolated populations, *Theor. Pop. Biol.* **38**, 331–350.

Takahata, N., Satta, Y., and Klein, J. 1995. Divergence time and population size in the lineage leading to modern humans, *Theor. Pop. Biol.* **48**, 608–615.

Tavaré, S. 1984. Line-of-descent and genealogical processes, and their applications in population genetics models, *Theor. Pop. Biol.* **26**, 119–164.