Heredity (2003) 90, 282–290

www.nature.com/hdy

ORIGINAL ARTICLES

Modelling the impact of colonisation on genetic diversity and differentiation of forest trees: interaction of life cycle, pollen flow and seed long-distance dispersal

F Austerlitz and PH Garnier-Géré
Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA, Domaine de l'Hermitage, B.P. 45, Pierroton, F-33611 Gazinet cedex, France

It was shown previously that the long lifespan and juvenile phase of trees strongly attenuate founder effects during colonisation in a diffusive dispersal model. However, this model yielded too slow a colonisation rate in comparison with palynological data for temperate forest trees. Since rare long-distance dispersal events have been shown to increase considerably colonisation rates in population dynamics models, we investigate here the impact of long-distance dispersal on within-population diversity (H_S) and among-population differentiation (F_{ST}) during the colonisation process. We use a stochastic approach and compare several dispersal strategies, ranging from very rare dispersal events of large amplitude to more frequent events of smaller amplitude. Using a simulation approach, which takes into account tree life-history traits, we show that long-distance dispersal events increase colonisation speed, and yield much larger founder effects in comparison with the diffusive model. The two models that include intermediate- and long-distance dispersal events show stronger deviations from experimental F_{ST} values during and at the end of the colonisation process than the model with more frequent events of smaller dispersal variance. Furthermore, the introduction of a high level of pollen flow has a much more limited impact on models that include long-distance dispersal than on a diffusive dispersal model. The relatively high H_S values that were obtained in all models are discussed according to the assumed mutation rate and effective population size. This study is an example of how observed genetic data can provide additional evidence on the best demographic model for a given species or group of species. *Heredity* (2003) **90,** 282–290. doi:10.1038/sj.hdy.6800243

Keywords: founder effect; diffusive dispersal; stratified dispersal; model; F_{ST}; demography

Introduction

Temperate forest trees have recolonised very large areas, following the last glacial period, in Europe as well as North America. Until 18 000 years ago, temperate forests were confined to the southern-most parts of these continents, and recolonisation of higher latitudes has taken place within a few hundreds to a few thousands of generations (Huntley and Birks, 1983; Webb, 1985; Huntley, 1990). Several population dynamics models (Skellam, 1951; Le Corre *et al*, 1997b; Clark, 1998) have shown that such a rapid colonisation process would have been impossible without the occurrence of rare long-distance dispersal events. Even very small proportions (from 10^{-6} to 10^{-4}) of these long-distance dispersal events were sufficient to accelerate drastically the colonisation process (Le Corre *et al*, 1997b), which then becomes compatible with the colonisation speeds estimated from pollen records (150–500 m/year, according to Huntley and Birks, 1983).

Correspondence: F Austerlitz, Laboratoire Ecologie Systématique et Evolution, UPRESA CNRS 8079, Bâtiment 360, Université Paris-Sud, F-91405 Orsay cedex, France. E-mail: Frederic.Austerlitz@ese.u-psud.fr

In the case of trees with heavy seeds, long-distance dispersal events have been reported and attributed to birds: for example, jays in the case of oaks (*Quercus* spp.), in Europe (Bossema, 1979) as well as in Northern America (Darley-Hill and Johnson, 1981), and passenger pigeons in the case of Northern American beeches (*Fagus* spp., Webb, 1987). Alternatively, other tree species such as maple (*Acer* spp.) and ash (*Fraxinus* spp.) have light, winged seeds, which allow them to be dispersed for long distances, mostly by wind.

Le Corre *et al* (1997b) showed that colonisation can be equally accelerated either by very rare long-distance dispersal events of very large amplitude (~50 km), or by more frequent events of more limited amplitude (~20 km). The latter model seems more realistic; for example, in the case of European oaks, long-distance dispersal events are frequent, but have never been observed on a scale of more than 8 km (Schuster, 1950). Very rare events of larger amplitude may have been unnoticed, of course, hence these observations are inconclusive. Also, seeds of some species could have been dispersed either along rivers over such long distances (Thebauld and Debussche, 1991), or as consequence of human activity.

From a genetic point of view, forest trees show, whatever their life-history traits, high levels of within-population diversity and low levels of among-population differentiation compared to herbaceous plants (Hamrick *et al*, 1992; Hamrick and Godt, 1996). This may appear surprising because colonisation frequently leads to an excess of differentiation among populations, when compared with the equilibrium situation (Austerlitz *et al*, 1997; Le Corre and Kremer, 1998), as a consequence of successive founder effects that accompany the creation of new populations, each genetically differentiated from the previous populations. However, Austerlitz *et al* (2000) have shown that the long lifespan of trees, and more importantly, their long juvenile phase, dramatically reduce the founder effects observed during colonisation by allowing more migrants to arrive, and eventually participate in reproduction when it takes place.

One problem of Austerlitz *et al*'s (2000) model was that it did not take into account the events of long-distance seed dispersal, which led to very slow colonisation rates. In contrast, models accounting for long-distance events, but for species with nonoverlapping generations and without a juvenile phase, predict an increased differentiation among populations for nuclear and cytoplasmic genes (Nichols and Hewitt, 1994; Le Corre *et al*, 1997b), which is inconsistent with experimental data in the case of nuclear genes. Furthermore, Nichols and Hewitt's (1994) model was adapted to animal species, and thus did not account for pollen flow, a key determinant of the genetic structure of nuclear loci (Petit *et al*, 1993; Ennos, 1994).

In short, models have been developed that either account for the extended juvenile phase of tree species and its effect on genetic structure but with a far too slow colonisation process, or integrate long-distance dispersal events but without taking into account life cycle and pollen flow. The aim of this paper is to assess the impact of long-distance dispersal during and after the colonisation period, on genetic diversity and differentiation of tree species, taking all the latter key components of their life cycle into account.

We thus develop a stochastic modelling approach for colonisation of trees, using overlapping generations, which allows us to integrate their long life cycle and juvenile phase. Populations are situated on a line of 300 km. At the beginning of each simulation, we have only one population at one of the extremities of the line. Colonisation occurs through seed migration that follows a dispersal distribution that integrates a portion of long-distance dispersal. Then populations remain connected through seed and pollen flow.

We test several colonisation scenarios, ranging from very rare long-distance dispersal events of very large amplitude to more frequent events of more limited amplitude, to see which scenario is more consistent with trees' high within-population diversity, as defined by Nei (1973) and low among-populations differentiation, as measured by F_{ST} (Wright, 1951; Nei, 1973). We also study the effect of pollen flow. Finally, we assess the impact of the numbers of loci and alleles per locus on the variance among simulations of diversity and differentiation levels. This allows to evaluate the robustness of the conclusions regarding the likelihood of one scenario or another.

The model

General description

We assumed that we had $d = 60$ populations situated on a line of length 300 km. In context, we visualise the populations being aligned south (on the left) to north (on the right). At time $t = 0$, only the left-most (southernmost) population was full and at demographic equilibrium. All other populations were empty. Since we needed to integrate rare long-distance dispersal events, which are by definition stochastic, we used a fully stochastic approach for demography and genetics. This is an important difference from Austerlitz *et al*'s (2000) model, which was deterministic as far as demography was concerned. Thus, we considered here a one-dimensional (1D) model, because simulations that integrate tree life-history characteristics are very time intensive in (2D) models.

In each population, individuals were sorted into $c = 25$ different size classes corresponding logically to increasing basal area values. Each year, each individual died or survived, either remaining in the same class or moving to the next class. Thus, each class contained individuals of different ages. Only the individuals of the last class produced seeds and pollen, thus trees could not reproduce before they were 25 years old. A density-dependent regulation made the size of each class increase towards an equilibrium size. The equilibrium distribution of individuals in classes corresponded to $N_e = 1000$, as computed using the method given in Orive (1993), and a generation time of 100 years at equilibrium. Otherwise in each population, the generation time varied and was shorter during its growth period because competition was less intense.

As in Le Corre *et al* (1997b), we assumed that the seed dispersal function (F_m) was a weighted sum of two zero-centred normal distributions:

$$F_m = (1 - a)N(0, \sigma_1) + aN(0, \sigma_2)$$

The first distribution $N(0, \sigma_1)$ corresponds to short-distance dispersal and the second $N(0, \sigma_2)$ to long-distance dispersal, the parameter a being the proportion of these long-distance dispersal events. Since we were dealing with discrete populations, we had to transform the continuous function F_m into a $d \times d$ matrix, $\mathbf{M} = (M_{ij})$, where M_{ij} represents the proportion of seeds that migrates from population i to population j. To compute the coefficients M_{ij}, the function F_m was integrated over all points in the intervals corresponding to populations i and j, assuming that every population occupied a 5 km interval (60 populations along 300 km).

The first model that we studied (model 1, see Table 1) was a diffusive dispersal model, with $\sigma_1 = 250$ m and no long-distance seed dispersal ($a = 0$). Since we assumed that each population occupied a 5 km range, there was no dispersal beyond the adjacent population in that case; that is, it was a pure stepping-stone process, with a fraction of ~ 0.02 of the seeds being dispersed to each of the two neighbouring populations.

We then studied several stratified dispersal models, in which $\sigma_1 = 250$ m, but a ranged from 5×10^{-6} to 10^{-2}, while the long-distance dispersal standard deviation (σ_2) ranged from 7.5 to 50 km (see Table 1). We retained models that yielded similar colonisation rates: two

Table 1 Length of the colonisation period (in years) under various demographic models

Model no.	*Dispersal parameters*		*Colonisation time (years)*	*Ratio to model 1*	*Colonisation rate (m/year)*
	a	σ_2 *(km)*			
1	0		4019	1.000	74.6
2	5×10^{-6}	50	1784	0.444	168
3	10^{-4}	20	1533	0.381	195
4	10^{-2}	7.5	1960	0.488	153

In all cases, the short-distance parameter was σ_1=250 m. Results are given for various values of the proportion (*a*) of long-distance seed dispersal events and of the variance of these long-distance dispersal events (σ_2). Colonisation times as a fraction of the time observed for the diffusive dispersal model (model 1) are also given.

contrasted models with either very rare long-distance dispersal events (model 2) or more frequent medium-distance dispersal events (model 4), and an intermediate case (model 3). For model 2, this meant that migrants could reach any population, but with a very small probability, from 1.95×10^{-7} for the second neighbours to 2.23×10^{-9} for the 59th neighbours. For model 4 this probability was 1.12×10^{-3} for the second neighbours, and dropped to zero by the 14th neighbours.

We also studied the impact of pollen flow, using a dispersal function, F_p, corresponding to a diffusive dispersal model, that is, a normal distribution with standard deviation $\sigma_p=1$ km The function F_p was transformed into a $d\times d$ matrix $\mathbf{P}=(P_{ij})$, using the same method as for seed dispersal. Applying F_p, 84% of the pollen remained in the population, while 8% migrated to each of the two neighbouring populations, a very small portion (1.07×10^{-8}) reached the second neighbours, but pollen did not go any further.

We simulated n_L independent, diploid, nuclear loci, with n_a alleles, and a symmetric mutation model, with mutation rate $\mu=10^{-6}$. In the initial population, all alleles were set to equal frequency and the genotypes to Hardy–Weinberg frequencies. Knowing the demographic and genetic situation at time t, we randomly sampled the number of carriers of each genotype for each class in each population at time t+1, as described below. This process was repeated iteratively, starting from $t=0$.

Within-population demographic model

For any population p, $p=1,\ldots,d$, at time t, denote $\mathbf{N}_p(t)=(N_{p,1}(t),\ N_{p,2}(t),\ldots,\ N_{p,c}(t))$ as the vector of the numbers of trees in each size class. Also denote by $\mathbf{A}$ the density-independent annual transition matrix, which is the same for all populations. $\mathbf{A}$ gave the expected numbers of individuals in each class of the population at time t+1, as a function of those same numbers at time t: $E(\mathbf{N}(t+1))=\mathbf{A}\cdot\mathbf{N}(t)$.

$$\mathbf{A}=\begin{pmatrix} \mathbf{A_{11}} & \mathbf{0} & \mathbf{0} & \ldots & f \\ \mathbf{A_{12}} & \mathbf{A_{22}} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A_{23}} & \mathbf{A_{33}} & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{A_{c-1,c}} & \mathbf{A_{cc}} \end{pmatrix}$$

where A_{ii} was the proportion of trees that remained both alive and in class i from t to t+1; $A_{i,i+1}$ the proportion of trees of class i that remained alive and moved from the class i to i+1 during the time interval t to t+1; and f was the number of offspring produced by each individual of the last class at each generation. Thus, we had $c-1$ classes of juveniles and only one class of adult trees that could reproduce. The proportion of individuals that died in class i during 1 year was $1-A_{ii}-A_{i,i+1}$.

Density dependence was added to the model by assigning to the individuals of each class a given stand basal area (a measure of the space occupied by an individual). Then, at each time t, fecundity f and all probabilities $A_{i,i+1}$ to move to the upper class were multiplied by a reduction coefficient $\alpha_p(t)$ ($0\leqslant\alpha_p(t)\leqslant1$), computed so that the basal area G_t of the whole population grew logistically towards an equilibrium value G_e (for details, see Austerlitz *et al*, 2000). Thus, density dependence acted on all classes except the last one. The chosen values for the coefficients of matrix $\mathbf{A}$ and for the standard basal area of the individuals of each class are given in the Appendix.

The expected proportion of individuals passing to the next class was therefore $\tilde{A}_{ii+1}=\alpha_p(t)A_{ii+1}$. For each population p, $p=1,\ \ldots,\ d$ and each class i, $i=1,\ \ldots,\ c$, the number of individuals ($R_{pi}(t)$) that survived and remained in class i, the number of individuals ($T_{pi}(t)$) that survived and moved to class i+1, and the number of individuals ($D_{pi}(t)$) of class i that died were drawn from a trinomial distribution with parameters ($N_{pi}(t)$, A_{ii}, $\tilde{A}_{ii+1}$, $1-A_{ii}-\tilde{A}_{ii+1}$). Thus, the total number of individuals in class $i>1$ of population p at time t+1 was

$$N_{pi}(t+1)=R_{pi}(t)+T_{p,i-1}(t)$$

The value for class 1, $N_{p1}(t+1)$, is computed in the next section.

Migration model

We have defined $\mathbf{M}=(M_{ij})$ and $\mathbf{P}=(P_{ij})$ above as the $d\times d$ seed and pollen migration matrices, respectively. Before the action of density dependence, the number of seeds produced, on average, by the adults of population q was $f\ N_{qc}(t)$. A given proportion of these seeds will migrate to each population p, according to $\mathbf{M}$. So the expected number of seeds that migrated from q to p was $n_{qp}(t)=fN_{qc}(t)M_{qp}\alpha_p(t)$, where $\alpha_p(t)$ is the density-dependence correction factor in population p (defined above), because these seeds, although originating from population q, arrived in population p and were subsequently subjected to competition there. To integrate stochasticity into this process, we drew the number of seeds ($n_{qp}(t)$) that migrated from q to p from a Poisson distribution with mean $n_{qp}(t)$.

The total number, $N_{p1}(t+1)$, of individuals in class 1 of population p at time t+1, was constituted by the $R_{p1}(t)$ individuals that survived and remained in class 1 and by the newborn individuals coming from all populations. Thus, the total number of individuals was

$$N_{p1}(t+1)=R_{p1}(t)+\sum_{q=1}^{d}n_{qp}(t)$$

Concerning pollen flow, the proportion $x_{pq}(t)$ of the pollen cloud of population p that came from

population q was

$$x_{pq}(t) = \frac{P_{qp}N_{qc}(t)}{\sum_{q'=1}^{d} P_{q'p}N_{q'c}(t)}$$

Thus, $x_{pq}(t)$ was proportional to the amount of pollen flow between q and p and to the relative size, compared to the other populations, of the adult class in population q.

Genetic model

For each of the n_L nuclear loci, since there were n_a alleles, the number of genotypes per locus was $n_g = n_a(n_a+1)/2$. Define $N_{pij}(t)$ as the number of carriers of each genotype j, $1 \leqslant j \leqslant n_g$, within the N_{pi}(t) individuals of class i, $1 \leqslant i \leqslant c$, in each population p, $1 \leqslant p \leqslant d$, thus

$$\sum_{j=1}^{n_g} N_{pij}(t) = N_{pi}(t).$$

For $i>1$, each class i of each population p at time t+1 consisted of the $R_{pi}(t)$ individuals that survived and remained in class i, and the $T_{p,i-1}(t)$ individuals that survived and moved from $i-1$ to i (see above). We knew the number of carriers, $N_{pij}(t)$, of each genotype j among the $N_{pi}(t)$ individuals, in each class i of population p. The $R_{pi}(t)$ individuals were sampled without replacement among the $N_{pi}(t)$ individuals. Thus, the n_g values

$$(R_{pi1}(t), R_{pi2}(t), \ldots, R_{pin_g}(t))$$

of the number of carriers of genotypes 1 to n_g within the $R_{pi}(t)$ individuals were drawn from a hypergeometric distribution, $H(R_{pi}(t), N_{pi1}(t), N_{pi2}(t), \ldots, N_{pin_g}(t))$, whose probability distribution function is defined as (Feller, 1957, pp 41–45)

$$p(R_{pi1}(t), R_{pi2}(t), \ldots, R_{pin_g}(t)) = \frac{\binom{N_{pi1}(t)}{R_{pi1}(t)}\binom{N_{pi2}(t)}{R_{pi2}(t)}\cdots\binom{N_{pin_g}(t)}{R_{pin_g}(t)}}{\binom{N_{pi}(t)}{R_{pi}(t)}}$$

The $R_{pij}(t)$ individuals were then withdrawn from the $N_{pij}(t)$ individuals, for each genotype j. The n_g values

$$(T_{pi1}(t), T_{pi2}(t), \ldots, T_{pin_g}(t))$$

of carriers of each genotype j, among the $T_{pi}(t)$ individuals of class i, of population p, that survived and moved to i+1 were then drawn without replacement from among the remaining individuals, following the same procedure as above. Then, the number of carriers of each genotype j, in each class $i>1$ of each population p, at time t+1 was calculated as

$$N_{pij}(t+1) = R_{pij}(t) + T_{p,i-1,j}(t), \quad (1)$$

For class 1, the number of carriers $R_{p1j}(t)$ of each genotype j among the individuals that survived and remained in it was drawn as above. To compute the number of carriers of each genotype among the newborns, we first calculated the allelic frequencies $f_{pk}(t)$ for the individuals in the adult class of each population p, $p=1,\ldots, d$, for each allele k, $k=1,\ldots, n_a$. Then, because of the n_a-alleles symmetrical model with mutation rate μ, these frequencies were modified to

$$f_{pk}'(t) = (1-\mu)f_{pk}(t) + \mu\frac{1-f_{pk}(t)}{n_a - 1}$$

Then the allelic frequencies ($f_{pk}^{\text{poll}}(t)$) in the pollen cloud of each population p was

$$f_{pk}^{\text{poll}}(t) = \sum_{q=1}^{d} x_{pq}(t) f_{qk}'(t)$$

where $x_{pq}(t)$ is the proportion of pollen from population q within the pollen cloud of population p (see above). Assuming random fertilisation within the population, the following genotypic frequencies were obtained, among the seeds produced in population p, for each genotype kk', with $1 \leqslant k \leqslant k' \leqslant n_a$:

$$f_{pkk'}(t) = f_{pk}'(t) f_{pk'}^{\text{poll}}(t)$$

These genotypes were then numbered, as above, with j, $1 \leqslant j \leqslant n_g$, and we denote $f_{pj}^{g}(t)$, the frequency of each genotype j in the newborns of population p. Then, since $n_{qp}(t)$ seeds migrated from population q to population p, the numbers of carriers of each genotype $(n_{pi1}(t), n_{pi2}(t), \ldots, n_{pin_g}(t))$, within these $n_{qp}(t)$ seeds, were drawn in a multinomial distribution with parameter $n_{qp}(t)$ and the genotypic frequencies $f_{qj}^{g}(t)$. Thus, adding the $R_{p1j}(t)$ individuals that survived and remained in class 1 yielded the total number ($N_{p1j}(t)$) of carriers of each genotype j in class 1 of each population p at time t+1:

$$N_{p1j}(t+1) = R_{p1j}(t) + \sum_{q=1}^{d} n_{qpj}(t) \quad (2)$$

Then, from equations (1) and (2), the number of carriers of each genotype in each class of each population at each time t, we computed the allelic frequency $f_{pk}(t)$ of each allele k within each population p. Then the total diversity, $h_p(t)$, within each population p was

$$h_p(t) = 1 - \sum_{k=1}^{n_a} f_{pk}^2(t)$$

and the average within-population genetic diversity, $h_S(t)$, was the average of $h_p(t)$ over every nonempty population p at time t. The total diversity, $h_T(t)$, was calculated as

$$h_T(t) = 1 - \sum_{k=1}^{n_a} \overline{f_k(t)}^2$$

where $\overline{f_k(t)}$ was the average of $f_{pk}(t)$ over every population p at time t. $h_S(t)$ and $h_T(t)$ were computed for each locus, and averaged over all n_L loci, to obtain the mean within-population diversity ($H_S(t)$) and total diversity ($H_T(t)$). The genetic differentiation, $F_{ST}(t)$, among populations was calculated as

$$F_{ST}(t) = 1 - \frac{H_S(t)}{H_T(t)}$$

Outcomes of the model

The simulations were run for 10 000 years for the various colonisation models. A period of 10 000 years is smaller than the estimated length (18 000 years) of the postglacial

period (Huntley, 1990), but was chosen because the simulations are time-intensive. The consequence is that the colonised area simulated by the model is smaller than the actual area. For each simulation, the time at which the 60 populations were colonised was recorded, and for each year t, the within-population genetic diversity, $H_S(t)$, and the genetic differentiation, $F_{ST}(t)$, among populations were also recorded. As a result of computational limitations, only 20 simulations were performed for each set of parameters and the results given below are the averages and standard deviations over these 20 repetitions.

Results

Colonisation dynamics

Table 1 gives the colonisation times for the various demographic models. In comparison with the diffusive dispersal model (model 1), the introduction of very small proportions of long-distance dispersal events considerably accelerated the colonisation process. Colonisation times less than half that of model 1 were reached under the three contrasted models of colonisation that included long-distance dispersal events. For example, the differences in colonisation times between model 2 (with infrequent but large-amplitude dispersal events) and model 4 (with frequent but small-amplitude dispersal events) were much smaller that between them and model 1. Rates of spread of $\sim$180 m/year on average were obtained for the three dispersal models including long-distance seed dispersal, whereas it was only 75 m/year in the case of the pure diffusion model.

Genetic diversity within and among populations

As in previous studies (Austerlitz *et al*, 1997; Le Corre *et al*, 1997b; Le Corre and Kremer, 1998; Austerlitz *et al*, 2000), decreased within-population diversity (H_S) and increased differentiation (F_{ST}) were observed during the colonisation period (Figure 1), as a consequence of genetic drift and founder effects. Owing subsequent homogenisation, H_S slowly increased and F_{ST} slowly decreased. The effects of the various colonisation models on H_S and F_{ST} are summarised in Table 2.

Founder effects were smallest in a pure diffusive dispersal model (model 1); with or without pollen flow, the minimum H_S values were only slightly below the values at the end of the 10 000 years and conversely the maximum F_{ST} values were only slightly above the end value, indicating a minimal loss of diversity or increase in differentiation because of founder events in each population. There was, however, an important quantitative difference between values reached with and without pollen flow (Figures 1 and 2, model 1). In the diffusive model, gene exchange through pollen was a powerful factor of homogenisation between populations with F_{ST} never increasing above 0.067 (Figure 2).

The introduction of rare seed long-distance dispersal events had a strong impact on the evolution of the genetic structure of trees. The rates of colonisation were higher in the stratified models (models 2–4) than in the diffusive dispersal models. These models thus showed a decrease in H_S, which occurred in less than half the time of model 1 (Figures 1 and 2). This decrease was greatest for model 3, especially in the case of pollen flow.

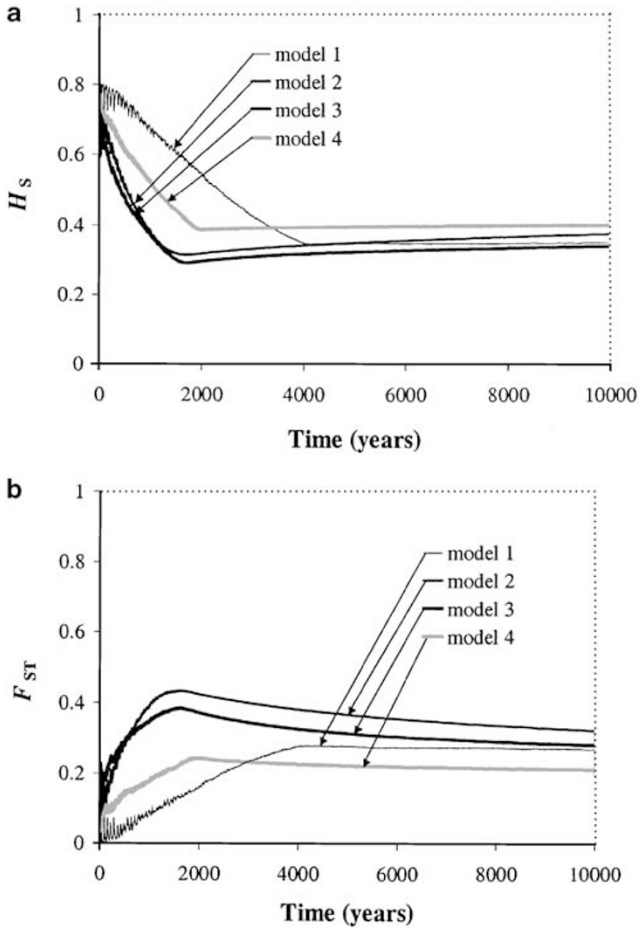

Figure 1 Evolution of (**a**) within-population diversity (H_S) and (**b**) among-population differentiation (F_{ST}) for all models (see Table 1 for the seed-dispersal parameters of each model), with no pollen flow ($\sigma_p = 0$). A minimum value was reached for H_S at the end of colonisation, which subsequently increased to its equilibrium value. Conversely, the maximum F_{ST} value was reached at the end of colonisation, which subsequently decreased to its equilibrium value.

Conversely, a more rapid increase in F_{ST} towards F_{STmax} was observed, with clearly higher values, for models 2 and 3 than for model 4, with or without pollen flow (Figures 1 and 2). These two models were the ones with rare, long-distance dispersal events of larger amplitude (50 and 20 km, for models 2 and 3, respectively). In the case of pollen flow for models 2 and 3, F_{ST} reached an average maximum of $\sim$0.34–0.35, which was reduced, after subsequent homogenisation, to $\sim$0.23–0.25.

The model including more frequent long-distance dispersal events of small amplitude (7.5 km) could be distinguished from the other two stratified models by both a slower decrease of H_S towards H_{Smin} and a slower increase of F_{ST} towards F_{STmax} (Figures 1 and 2). The founder effects were also smaller for this model, the extreme values reached by either H_{Smin} or F_{STmax} were greater and smaller, respectively, and less different from the end values, than for models 2 and 3. In the three stratified models, the addition of pollen flow decreased F_{STmax} and increased H_{Smin}. However, the effect was not as strong as for the diffusive dispersal model, and the

Table 2 Effect of the four colonisation models on average within-population diversity (H_s) and among-population differentiation (F_{ST}), for five independent nuclear loci, each with five alleles[a]

Model no.	*Dispersal parameters*		$H_{S\,min}$ [b]		$H_{S\,end}$ [c]		$F_{ST\,max}$ [d]		$F_{ST\,end}$ [e]	
	a	σ_2 (km)	σ_p=0	σ_p=1	σ_p=0	σ_p=1	σ_p=0	σ_p=1	σ_p=0	σ_p=1
1	0		0.325 (0.082)	0.673 (0.049)	0.329 (0.081)	0.674 (0.049)	0.278 (0.049)	0.067 (0.027)	0.272 (0.047)	0.065 (0.024)
2	5×10^{-6}	50	0.312 (0.061)	0.374 (0.069)	0.374 (0.076)	0.439 (0.076)	0.434 (0.054)	0.349 (0.062)	0.322 (0.051)	0.231 (0.043)
3	10^{-4}	20	0.290 (0.058)	0.302 (0.069)	0.339 (0.074)	0.340 (0.077)	0.385 (0.046)	0.338 (0.052)	0.281 (0.054)	0.246 (0.046)
4	10^{-2}	7.5	0.386 (0.076)	0.397 (0.095)	0.401 (0.078)	0.413 (0.095)	0.244 (0.058)	0.240 (0.070)	0.212 (0.057)	0.207 (0.067)

[a]In all cases, the short-distance parameter was σ_1=250 m. Results are given for values of the proportion of long-distance seed dispersal events (a) and the variance of these long-distance dispersal events (σ_2), without (σ_p=0) or with pollen flow (σ_p=1 km) among populations. All values are the average of the outcome of 20 simulations. Standard deviations are given in parentheses below the average values. Increasing the number of alleles to 10 hardly changed the values observed and did not change the comparison between the different models. [b]Minimum value reached by H_s during the simulations. [c]Value of H_s at the end of the simulations. [d]Maximum value reached by F_{ST} during the simulations. [e]Value of F_{ST} at the end of the simulations.

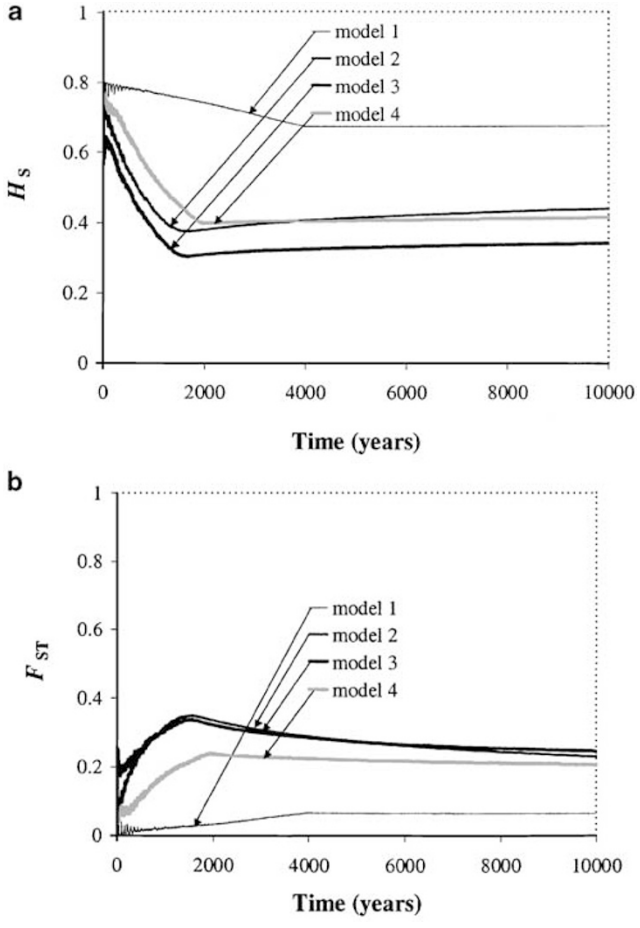

Figure 2 Evolution of (**a**) within-population diversity (H_S) and (**b**) among-population differentiation (F_{ST}) for all models (see Table 1 for the seed dispersal parameters of each model), with pollen flow ($\sigma_p = 1$ km).

maximum value reached by F_{ST} remained high ($F_{STmax} = 0.349$ for model 2 and $F_{STmax} = 0.240$ for model 4).

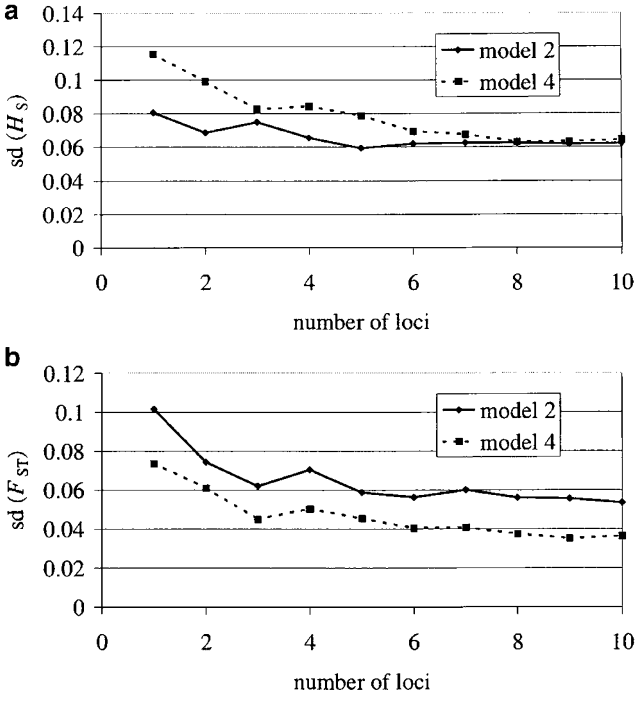

Figure 3 Standard deviation (SD) of the minimum value reached by H_S (**a**) and the maximum value reached by F_{ST} (**b**) during the simulations, plotted against the number of nuclear loci used to compute these quantities.

Standard deviations for each parameter of all models were computed for up to 10 loci and they clearly decreased with the number of loci involved (Figure 3). The decrease was rapid up to five or six loci but became minimal or nonexistent as additional loci were considered, and reached an asymptotic value.

Discussion

The models developed here integrate stochasticity into the process of migration at the level of dispersion (for

both seeds and pollen), at the level of seed establishment in new populations, and at the genetic level. This work therefore represents an improvement for modelling natural processes, compared to previous colonisation models (Austerlitz *et al*, 2000), and also for assessing different colonisation scenarios, while accounting for stochasticity.

As expected in a stochastic process, there is a substantial amount of variation from one simulation to the next for the estimated parameters (diversity and differentiation). Part of this variance is because of the random transmission of alleles, which yields variation in genotypic frequencies from one locus to the next, and explains why standard deviations decrease rapidly with the number of loci used. However, if more than five loci are considered, the standard deviations for any parameter converge to asymptotic values. This is the part of variance that is due to the stochasticity of the demographic process itself. Inspite of this stochasticity, the different colonisation models considered can be clearly distinguished by their effects on the mean patterns of population genetic parameters.

In terms of colonisation rates, in models that account for the overlapping generations, long life cycle and juvenile phase of tree species, a small proportion of long-distance seed dispersal is sufficient to accelerate strongly the colonisation process. This result is similar to previous ones obtained in models designed for plants with a limited number of overlapping generations and no juvenile phase (Le Corre *et al*, 1997b). Importantly, the rates of colonisation obtained with the three stratified dispersal models (models 2–4) were of similar magnitude (153–196 m/year) and consistent with the range of colonisation rates estimated for European tree species on the basis of palynological data (150–500 m/year; Huntley and Birks, 1983). In comparison, the pure diffusive model (model 1) gave a too slow colonisation rate of 75 m/year.

Concerning the parameters of average within-population diversity (H_S) and among-population genetic differentiation (F_{ST}), our results show that in a stochastic diffusive dispersal model, the genetic structure of tree populations is only slightly affected by founder effects; there is only a minor reduction in diversity in each newly founded population, thus H_S decreases to a minimum that is only slightly lower than the end value, while the maximum F_{ST} value is only slightly above the end value. This result is consistent with the conclusion reached by Austerlitz *et al* (2000) for a deterministic diffusive model.

However, founder effects play an important role in stratified dispersal models since new colonies are established at long distances and are thus isolated from the colonisation front. In a model without pollen flow, when the proportion of long-distance migration is low ($a = 5 \times 10^{-6}$), but the standard deviation of these dispersal events is large ($\sigma = 50$ km), F_{ST} evolved towards its highest value (> 0.4) during $\lesssim 20\%$ of the colonisation time. The subsequent homogenisation was limited, and the final value of F_{ST} was still fairly high ($F_{STend} = 0.322$). In this model, long-distance events disperse individuals very far in advance of the colonisation front. When an individual successfully establishes in a far away population, it may reach adulthood before any other migrant establishes in the same population. Even with some pollen flow, in a model with very large long-distance dispersal variance, the founding populations will consist mostly of individuals produced by the seeds of the first founder. The juvenile phase has been shown earlier to reduce dramatically the founder effect in the diffusive dispersal model by allowing founders to arrive during several years (Austerlitz *et al*, 2000). However, we observe here that it does not allow enough seed flow towards the founding populations to compensate the rare occurrence of long-distance dispersal events.

The clearest differences were observed for F_{ST} maximum values and H_S minimum values, between the model with the lowest long-distance dispersal ($\sigma_2 = 7.5$ km) and the other two stratified models. This was true except for H_S in the models with pollen flow, where the lowest minimum was reached with the model with intermediate long-distance dispersal.

While pollen flow reduces much F_{ST} in the diffusive dispersal model, it has a much more limited impact in the stratified models. For the diffusive model, because a newly founded population is always adjacent to an existing one, homogenisation by pollen flow can begin immediately after foundation. In contrast, for stratified dispersal models, the rare events of long-distance dispersal produce clusters of population that remain separated by gaps, long enough for substantial differentiation to occur, without any possible homogenisation by pollen flow.

Comparing the likelihood of different colonisation scenarios, we have seen that the pure diffusive model can be excluded since the colonisation times are too unrealistic compared to the stratified colonisation models. A general pattern of high diversity within populations and low differentiation between populations has been observed experimentally for outcrossing tree species (Hamrick *et al*, 1992), with average allozyme H_S values of 0.253 for woody plants and an average G_{ST} value of 0.084 (range of 0.041–0.206 between different genera). One may argue that it is difficult to compare our F_{ST} values with experimental G_{ST} values because of differences in estimation methods; however, it was shown that these differences are small in practical cases (Culley *et al*, 2002). Among the stratified models when pollen flow is included, the values closest to Hamrick *et al*'s experimental G_{ST} values that we obtained were the final F_{ST} values of ~ 0.21 for the model characterised by long-distance migration events with lower dispersal variance (7.5 km). However, the H_S values of ~ 0.34 obtained with the model of intermediate long-distance dispersal variance (20 km) is the closest to the average observed value of 0.253 for tree species. The Hamrick *et al* (1992) survey does not give a range of values for H_S, but recent studies on populations from different oak species (Le Corre *et al*, 1997a; Streiff *et al*, 1998) give allozyme H_S values > 0.3, which are comparable to the ones obtained in our models. Therefore, both H_S and F_{ST} values obtained from our simulations fall at the upper range of the ones observed experimentally.

Two main reasons could be invoked to explain the relatively high H_S values that we obtained. The first one is the maintenance of a high initial variability at the colonisation front in our models, while this variability could have been reduced by the reduction in the size of refugia populations during the glacial period. The second reason is that the mutation rate of 10^{-6} which

we used in the simulations, might be higher than the mutation rate of allozymes that were the markers considered in Hamrick *et al*'s (1992) review. This, in conjunction with initial equal allelic frequencies and a population effective size of 1000, is likely to maintain more diversity than in the case of allozymes with two or three major alleles only.

The relatively high F_{ST} averages obtained in our model could be explained in different ways. Effective population sizes might have been greater than 1000, which would have resulted in lower F_{ST}. Pollen flow could also be higher than what was assumed in the models: a normal distribution of standard deviation $\sigma_p = 1$ km was used here for pollen dispersal, which means an average dispersal distance of ~ 800 m. This falls in the range of observed values for insect-pollinated species (Dick, 2001; Dutech *et al*, 2002). Concerning wind-pollinated species, Streiff *et al*'s (1999) estimated an average pollen dispersal distance of 286 m for *Q. robur* and 333 m for *Q. petraea*. So the pollen flow used here might again seem realistic, but Streiff *et al* (1999) estimate of pollen dispersal was indirect, and the real value could be higher. Moreover, in some coniferous species, it was estimated that pollen could potentially disperse at a distance of up to 60 km, without any certainty about its viability (Di-Giovanni *et al*, 1996). Only little data is available on the viability of long-distance dispersing pollen. One study showed 75% viability after a 72 km travel (Andersson, 1965), while the germination rate of pollen that had been transported over long distances was very low in two other cases (Solomon, 1979; Pulkkinen and Rantio-Lehtimäki, 1995). In our simulations, the assumed pollen dispersal value might be too low for coniferous species, explaining in part why these species have lower F_{ST} values in Hamrick *et al*'s (1992) review compared with the ones we obtained.

Computational intensiveness meant that it was necessary to use a 1D model of limited scale (300 km), so one may wonder about the extension of these results to a 2D model on a larger scale (typically areas of more than 1000 km $\times$ 1000 km for forest trees in Europe). During the colonisation period in a 2D model, it is expected, as in the 1D model, that rare long-distance dispersal events, by dispersing limited quantities of individuals ahead of the colonisation front, will yield a high founder effect, increasing differentiation and reducing average within-population diversity. Since convergence towards equilibrium is usually faster in 2D models than in 1D models (Maruyama, 1971; Slatkin, 1991), the decrease of differentiation and increase of diversity after the colonisation period in 2D models should be faster. The problem is how 2D modelling would affect the end values and their comparison in the case of the three types of stratified models. After the colonisation, the homogenisation of the whole set of populations is likely to be faster in a 2D model, compared to a 1D model, for medium-distance dispersal, which allows a sufficient number of genes to move between adjacent populations. This difference between 2D and 1D models should however be more limited for the case of rare, long-distance dispersal events, which are expected to have less impact after the colonisation because of the very small probability of a few, rare migrants establishing in a population that is already full.

Several features could still be integrated in the model to increase realism. For example, we could assume that several seeds disperse jointly. This has been observed in the case of oaks, for which jays carry up to seven acorns at a time (Schuster, 1950). This practice should increase the number of founders and therefore reduce initial differentiation (Slatkin, 1977; Wade and McCauley, 1988; Whitlock and McCauley, 1990; Le Corre and Kremer, 1998), provided that these founders are not too related. This might be the case for trees because they are highly outcrossing species, and females are usually fertilized by many males.

This work supports the importance of relatively frequent medium-distance dispersal events over rarer events of much longer distance dispersal, during the process of colonisation of outcrossing forest tree species. We have predicted the evolution of tree species' genetic diversity and differentiation, under various conditions, during the complete process of colonisation using a realistic demographic model. The usefulness of this approach is that it can be contrasted with real data, giving an indication of what the best demographic scenario might be, not only from equilibrium expectations but also from predictions during the different phases of colonisation. This shows how the simulation of population genetic parameters can help to distinguish among different possible demographic hypotheses. The key factor is to have a realistic demographic model for the species under consideration. Parameters such as juvenile phase or overlapping generations can, at equilibrium, be reduced to a statement about effective population size (Charlesworth, 1994), but different combinations yield different outcomes during the transient colonisation period.

Acknowledgements

We thank PE Smouse, J-B Ferdy, S Mariette, RJ Petit, the designated editor and two anonymous reviewers for their helpful comments and suggestions. PE Smouse's help on English usage was also very useful. FA was supported by a Formation Complémentaire par la Recherche grant from French Ministère de l'Agriculture and a complementary grant from NATO. This work was funded in part by 'Programme National Dynamique de la Biodiversité et Environnement' of the French Centre National de la Recherche Scientifique. Some simulations were performed on the SP2 and SGI computers of the Centre des Ressources Informatiques of the Université de Paris-Sud (Orsay) and on the UNIX machines of the Centres de Recherches INRA of Bordeaux and Jouy-en-Josas.

References

Andersson E (1965). Flowering and seed setting in conifers. A short review. Cone and seed studies in Norway spruce (*Picea abies* L. Karst.). *Stud For Suec* **23**: 5–214.

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997). Evolution of coalescence times, genetic diversity and structure during colonisation. *Theor Popul Biol* **51**: 148–164.

Austerlitz F, Mariette S, Machon N, Gouyon P-H, Godelle B (2000). Effects of colonisation processes on genetic diversity: differences between annual plants and tree species. *Genetics* **154**: 1309–1321.

Bossema I (1979). Jays and oaks: an eco-ethological study of a symbiosis. *Behaviour* **70**: 1–117.

Charlesworth B (1994). *Evolution in Age-Structured Populations*. Cambridge University Press: Cambridge.

Clark JS (1998). Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *Am Nat* **152**: 204–224.

Culley TM, Wallace LE, Gengler-Nowak KM, Crawford DJ (2002). A comparison of two methods of calculating G_{ST}, a genetic measure of population differentiation. *Am J Bot* **89**: 460–465.

Darley-Hill S, Johnson WC (1981). Dispersal of acorns by blue jays (*Cyanocitta cristata*). *Oecologia* **50**: 231–232.

Dick CW (2001). Genetic rescue of remnant tropical trees by an alien pollinator. *Proc R Soc London B* **268**: 2391–2396.

Di-Giovanni F, Kevan PG, Arnold J (1996). Lower planetary boundary layer profiles of atmospheric conifer pollen above a seed orchad in northern Ontario, Canada. *Forest Ecol Manage* **83**: 87–97.

Dutech C, Seiter J, Petronelli P, Joly HI, Jarne P (2002). Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Mol Ecol* **11**: 725–738.

Ennos RA (1994). Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250–259.

Feller W (1957). *An Introduction to Probability Theory and its Application*. John Wiley and Sons: New York. Vol. 1.

Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity in plants. *Philos Trans R Soc London B* **351**: 1291–1298.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992). Factors influencing levels of genetic diversity in woody plant species. *New Forests* **6**: 95–124.

Huntley B (1990). European vegetation history: paleovegetation maps from pollen data – 13 000 yr BP to present. *J Quat Sci* **5**: 103–122.

Huntley B, Birks HJB (1983). *An Atlas of Past and Present Pollen Maps of Europe: 0–13000 Years Ago*. Cambridge University Press: Cambridge.

Le Corre V, Dumolin-Lapegue S, Kremer A (1997a). Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. *Mol Ecol* **6**: 519–529.

Le Corre V, Kremer A (1998). Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping stone model. *J Evol Biol* **11**: 495–512.

Le Corre V, Machon N, Petit RJ, Kremer A (1997b). Colonisation with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genet Res* **69**: 117–125.

Maruyama T (1971). Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. *Ann Hum Genet* **35**: 179–196.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321–3323.

Nichols RA, Hewitt GM (1994). The genetic consequences of long distance dispersal during colonisation. *Heredity* **72**: 312–317.

Orive ME (1993). Effective population size in organisms with complex life-histories. *Theor Popul Biol* **44**: 316–340.

Petit RJ, Kremer A, Wagner DB (1993). Finite island model for organelle and nuclear genes in plants. *Heredity* **71**: 630–641.

Pulkkinen P, Rantio-Lehtimäki A (1995). Viability and seasonal distribution patterns of Scots pine pollen in Finland. *Tree Physiol* **15**: 515–518.

Schuster L (1950). Über den Sammelbetrieb des Eichelhähers (*Garrulus glandarius*). *Vogelwelt* **71**: 9–17.

Skellam JG (1951). Random dispersal in theoretical populations. *Biometrika* **38**: 196–218.

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

Slatkin M (1991). Inbreeding coefficients and coalescence times. *Genet Res* **58**: 457–462.

Solomon AM (1979). Sources and characteristics of airborne materials: pollen. In: Edmonds RL (ed) *Aerobiology: The Ecological Systems Approach*. Dowden, Hutchinson and Ross, Inc.: Stroudsberg, PA, USA.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Mol Ecol* **8**: 831–841.

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Gloessl J, Kremer A (1998). Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Mol Ecol* **7**: 317–328.

Thebauld C, Debussche M (1991). Rapid invasion of *Fraxinus ornus* along the Herault River system in southern France: the importance of seed dispersal by water. *J Biogeogr* **18**: 7–12.

Wade MJ, McCauley DE (1988). Extinction and recolonisation: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995–1005.

Webb SL (1987). Beech range extension and vegetation history pollen stratigraphy of two Wisconsin (USA) lakes. *Ecology* **68**: 1993–2005.

Webb III T (1985). Vegetation changes in eastern North America: an 18 000 years history from pollen maps. *Stud Plant Ecol* **16**: 100–102.

Whitlock MC, McCauley DE (1990). Some population genetics consequence of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717–1724.

Wright S (1951). The genetical structure of populations. *Ann Eugen* **15**: 323–354.

Appendix : Values chosen for the coefficients of the transition matrix **A** and the stand basal area of each class

The following parameters were chosen for the transition matrix **A**: $f=50$, $P_{11}=0$, $P_{ii}=0.54$ for $2\leqslant i\leqslant 24$, $P_{25,25}=0.98$, $P_{12}=0.5$, $P_{i,i+1}=0.4$ for $2\leqslant i\leqslant 24$. The total standard basal area at equilibrium was set to $G_e=89\,955$, and the standard basal areas (G_i) of each class i were set at $G_1=0.01$, $G_i=0.5$ for $2\leqslant i\leqslant 8$, $G_i=1$ for $9\leqslant i\leqslant 16$, $G_i=2$ for $17\leqslant i\leqslant 24$, $G_{25}=3$. As in our previous study (Austerlitz *et al*, 2000), these values are somewhat arbitrary, but were adjusted in order to fulfil the main characteristics of tree species, as indicated in the main text.