Genet. Res., Camb. (1997), **69**, *pp.* 117–125. *With 7 figures.* Printed in the United Kingdom

Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study

VALÉRIE LE CORRE[1], NATHALIE MACHON[2], RÉMY J. PETIT[1]
AND ANTOINE KREMER[1]*

[1] *Laboratoire de Génétique et Amélioration des Arbres Forestiers, Institut National de la Recherche Agronomique, B.P. 45 F-33611 Gazinet cedex, France*

[2] *Laboratoire d'Evolution et Systématique des Végétaux, Batiment 362, Université Paris-Sud, F-91405 Orsay cedex, France*

(*Received 10 October 1996 and in revised form 16 December 1996*)

Summary

Since the last glacial period forest trees have expanded to their present range very rapidly, with rates up to 500 m yr^{-1} for oaks in Europe, which can be explained only by the dispersion of acorns over long distances. We used a stratified dispersal model, including both diffusive and long-distance dispersal of seeds, to simulate the colonization of a 100 km × 300 km grid by populations of oak trees. An appropriate rate of spread is obtained with rare dispersal at distances of the order of tens of kilometres. We simulated the effect of stratified versus diffusive dispersal of seeds on the spatial genetic structure at a maternally inherited locus. Founding events associated with stratified dispersal generate a high amount of genetic differentiation among populations, which is likely to persist for a long time after colonization. Using autocorrelation methods, we show that diffusive and stratified dispersals create quite different spatial patterns of variation for the maternally inherited locus. Stratified dispersal creates patchy patterns that are concordant with a previous experimental investigation of chloroplast DNA variation at a regional scale in the oaks *Quercus petraea* and *Quercus robur*. For plant populations that have passed through recent episodes of range expansion, long-distance dispersal events are probably the most important factors of spatial genetic structuring of maternally inherited genes at small or medium geographic scales.

1. Introduction

Modes of colonization and their consequences are often considered in association with exceptional events such as the rapid invasion of new territories by an introduced species (Hengeveld, 1989). However, at the scale of geological times, most animal and plant species in the temperate zones have passed through dramatic range reductions followed by recolonization phases. These phenomena were the consequences of successive glacial periods, the last of which culminated 18000 years ago in Europe. For forest trees, the timing and directions of recolonization during the last climatic warming are relatively well known from analysis of fossil pollen data and from the construction of isopollen maps (see, e.g., Huntley & Birks, 1983 for Europe; Webb, 1985 for eastern North America). One of the most striking observations is the rapidity of the spread, particularly for oaks, for which estimated rates are up to 500 m yr^{-1} (Huntley & Birks, 1983).

The first attempt to construct models for such expansions was made by Skellam (1951), who determined a diffusion law close to Fisher's (1937) equation for the spread of an advantageous gene in a population. This model is characterized by a 'wave front' progressing at a constant rate. Skellam demonstrated numerically that this mode of propagation could not explain the rate of spread observed for oaks, unless acorns were dispersed over long distances, possibly by birds. The likelihood of long-distance transport of acorns was further confirmed by ethological studies on European jays (Bossema, 1979) and American blue jays (Darley-Hill & Johnson, 1981). Similarly, long-distance transportation of nuts by passenger pigeons was proposed to explain palynological evidence of great leaps during the range expansion of beech in North America (Webb, 1987).

Because forest trees are very long-lived species, their present genetic variability is likely to reflect ancient colonization events. Chloroplast genomes are non-recombinant and transmitted by seeds only in most species of angiosperms (including oaks: Dumolin

---

* Corresponding author. Tel: (+33) 5 57 97 90 74. Fax: (+33) 5 57 97 90 88. e-mail: Antoine.Kremer@pierroton.inra.fr.

et al., 1995). They are therefore particularly appropriate for the study of colonization dynamics in plants, which are governed by seed dispersal. It was recently shown (Petit *et al.*, 1996) that the chloroplast DNA variability of oaks is highly spatially structured at a regional scale, with haplotypes clustered into patches of about 50 km wide. This patchy structure was interpreted as being a probable consequence of rare long-distance dispersal of seeds during colonization, each patch resulting from a founding event beyond the colonizing front.

The use of simulations rather than analytical methods to study genetic structures is particularly appropriate for recent populations that have not yet attained equilibrium values. This is probably the case for temperate forest species that reached their present ranges no more than a few hundreds of generations ago. In our study, simulations were used first to determine to what extent long-distance dispersal of seeds is needed to allow the high rates of spread observed for the postglacial expansion of forest trees. Second, simulations were performed to investigate the effect of long-distance dispersal events on the genetic differentiation and spatial structure of populations for a maternally inherited marker. Palynological and biological data on oaks were used to obtain realistic simulation parameters. We measured genetic differentiation and carried out an autocorrelation analysis of original and trend residual data (Bocquet-Appel & Sokal, 1989) to describe the spatial genetic structures generated by the simulations. Results are compared with molecular data on European oak populations.

2. Materials and methods

(i) *Simulations*

We considered a stochastic, spatially explicit model of a plant metapopulation, in which demography, migration and genetic drift were simulated, but in which no extinctions occurred. A simplified flowchart of the program is shown in Fig. 1. Selection and mutation were assumed to be absent. Demography was simulated through the use of a logistic growth curve. Migration was modelled by using gene flow matrices: one for pollen flow, giving the composition of the pollen cloud which fertilizes the individuals of a given population, and one for seed dispersal, giving the number of seeds received by one population from the others. These numbers are proportional to the sizes of the source populations. The choice of a particular pollen flow matrix had no effect on the evolution of a maternally inherited locus (as long as it is not limiting). The spatial configuration of the populations was reported through the characteristics of the seed flow matrix, distant populations exchanging fewer seeds than proximal populations. The choice of the parent trees, the sampling of gametes and their combination to form zygotes are all performed stochastically in order to simulate the effect of random genetic drift in panmictic populations.

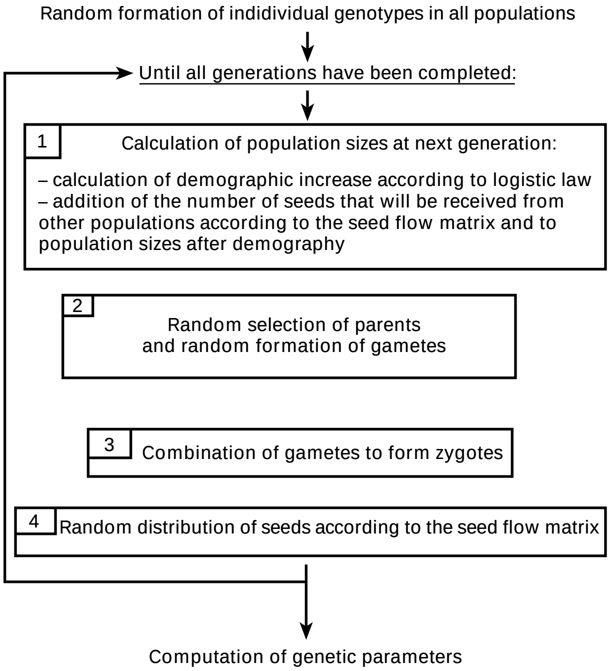

Fig. 1. Simplified flowchart for the program ‘Metapop’ used in this study.

(*a*) *Input data*

One thousand and two hundred populations were regularly distributed on a 300 km × 100 km grid, i.e. one population occurred in each 5 km × 5 km quadrat. The parameters for the logistic growth were a carrying capacity of $K = 1000$ and an initial growth rate $r =$ 1·5. The value of K is a middle course between the need to adequately represent the large population size of forest trees and the size limitations of the computer program. The value of r is in agreement with experimental data on acorn production and survival and on thinning rates in oaks (Harper × White, 1974; Sork, 1984; Iida, 1996). Perenniality of trees was modelled by specifying that individuals existing at any nth generation had a probability c of surviving to the $(n+1)$th generation. The mean longevity of an individual tree was then equal to $c/(1-c)$ simulated generations, the length of one simulated generation being equal to the time between germination and first reproduction. c was given a value of 0·9, which is compatible with the life-span of oaks. However, the results of the simulations (rate of spread and final genetic structure) were not affected by c values varying between 0·8 and 0·95 (data not shown).

The matrix of seed flow was generated from a probability distribution representing the probability of seed dispersal as a function of the distance travelled. We used a weighted sum of two normal distributions (Nichols & Hewitt, 1994):

$$F = (1-a)N[0, sd1] + aN[0, sd2]. \quad (1)$$

The first normal distribution represented short-distance dispersal as performed by simple gravity or by small mammals, and its standard deviation *sd1* was given a value of 250 m. The second normal distribution accounted for long-distance events. Five values were tested for *sd2* (20, 30, 40, 50 and 60 km). The value of the parameter *a* was either 0, or chosen from a set of values allowing rare long-distance events (10^{-4}, 5×10^{-5}, 10^{-5}, 5×10^{-6}, 10^{-6}, 10^{-7}). When *a* is equal to 0, seed dispersal is purely diffusive as in Skellam's (1951) model. When *a* is different from 0, seeds are dispersed both by diffusion and by long-distance movements, and the spread of the populations conforms to a stratified model as defined by Hengeveld (1989). To compute the probabilities of seed migration between any two populations, their areas were approximated to circles, and the dispersal function was integrated over all distances between points in the two areas.

Simulations were initialized with the 20 populations located furthest left on the grid containing 1000 individuals each, all other populations being empty. Each individual is represented by one cytoplasmic (haploid) locus with three alleles, which are strictly maternally inherited. Two different possible initial genetic structures were tested. First, each allele was present with equal probability in each initial population, so that no genetic differentiation was present. Secondly, each population was fixed for a given allele, so that genetic differentiation was maximum. Moreover, in this case, a spatial structure was created by arranging the populations in three zones specific to each allele.

(*b*) *Number of replicated simulations*

We first examined the effects of modifying the probability distribution for seed dispersal on the rate of spread. Four replicated simulations were run for each combination of values of the parameters *a* and *sd2*. Two contrasting dispersal models, one purely diffusive and one stratified, were then chosen to simulate colonization for the two contrasting initial genetic structures. Twenty replicate simulations were performed for each of these four cases. Each simulation lasted 300 generations. The four sets of results obtained are thereafter designated as model 1 (diffusion, initial genetic differentiation is 0), model 2 (diffusion, initial genetic differentiation is 1), model 3 (stratified dispersal, initial genetic differentiation is 0) and model 4 (stratified dispersal, initial genetic differentiation is 1).

(ii) *Analysis of results*

The length of the colonization phase was evaluated as the number of generations that elapsed until all the populations in the grid contained at least one individual. Genetic differentiation among populations containing more than 10 individuals was calculated at each generation using Nei's parameter G_{st} (Nei, 1973).

The spatial genetic structure obtained at the end of each simulation was first visualized by maps showing the predominant allele in each population. Spatial structures were then evaluated by autocorrelation analysis (Sokal & Oden, 1978), after having discarded the populations located furthest left on the grid, up to 50 km on the abscissa, in order to avoid a confusion between the initial and post-colonization patterns. We analysed the two alleles which were the most frequent at the end of each simulation. We used Moran's *I* coefficient as a measure of spatial autocorrelation for allele frequencies:

$$I = n\sum_i \sum_j w_{ij} Z_i Z_j / W \sum_i Z_i^2 \qquad (2)$$

where n is the number of sample points, $w_{ij}=1$ if localities i and j are considered to be spatially connected, otherwise $w_{ij}=0$. W is the sum of all w_{ij}, and Z_i is the difference between allele frequency at locality i and the overall mean allele frequency. The spatial connectedness of populations may be specified either by a graph of connections (Gabriel & Sokal, 1969), or by defining intervals of geographical distances. Moran's *I* may be calculated for different such distance intervals between populations and plotted as a function of distance to draw a correlogram. Moran's *I* is asymptotically normally distributed as n increases. Its statistical significance can be assessed by comparing its value with the expected value $\mu = -1/(n-1)$ under the null hypothesis of random distribution of allele frequencies (Cliff & Ord, 1981), and using the expression of variance given by Sokal & Oden (1978). Significant positive values indicate that allele frequencies at two localities are more similar than expected by chance alone.

Founder events occurring with a colonization process are expected to result in clines following the direction of the spread (Sokal & Menozzi, 1982; Cwynar & McDonald, 1987). This was tested by comparing autocorrelations between neighbour populations for four different graphs of connections, representing four directions of spread in space, conventionally named 'west–east', 'south–north', 'southwest–northeast' and 'northwest–southeast' (Fig. 2). Highest autocorrelations will be found in the direction of the main trend, thus allowing identification *a posteriori* of the connectivity graph that best accounts for the colonization process under study.

Trend surface analysis was thereafter used to separate spatial trends due to drift during colonization from local variations due either to local founder effects after long-distance dispersal or to isolation by distance (Bocquet-Appel & Sokal, 1989). This was done by fitting observed surfaces of allele frequencies to first-order polynomials of the type $z = a_0 + a_1 x + a_2 y + \epsilon$, which represents a two-dimensional cline (z is the allele frequency, x and y are the

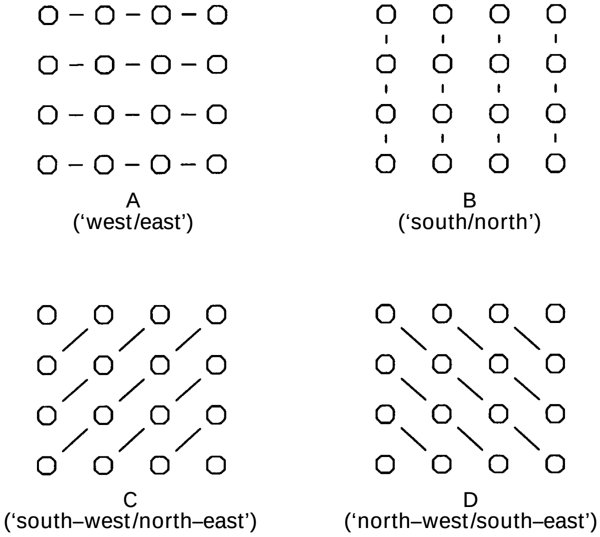
A
('west/east')
B
('south/north')
C
('south–west/north–east')
D
('north–west/south–east')

Fig. 2. Graphs of connections between populations used to represent different directions of spread in space.

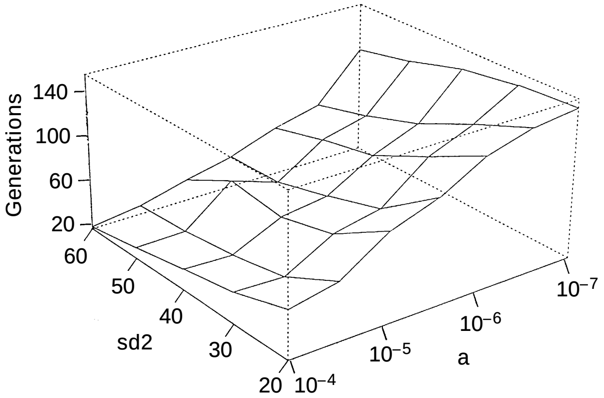

Fig. 4. Length of the colonization phase (in number of generations) for different values of the two parameters *a* and *sd2* of the dispersal function modelled as a weighted sum of two normal laws (see text). The relative weight of the second normal law *a* is taken in the interval 10^{-7} to 10^{-4}. The standard deviation of the second normal law *sd2* varies between 20 and 60 km. Each point on the graph is an average value over four replicated simulations.

coordinates of the grid surface and ϵ is the residue). Correlograms were drawn for both the original and residual allele frequencies, using the graph of connections selected in the earlier step. The overall significance of correlograms was evaluated by the Bonferroni method, as suggested by Oden (1984), with a 0·05 experimentwise error rate.

3. Results

(i) *Colonization dynamics*

When seed dispersal was modelled as a single normal law [$a = 0$ in eqn (1)], populations spread as an advancing front and colonisation was achieved after an average of 156·1 generations (standard error over the 40 replicates: 1·5). For a mixing of diffusion and long-distance seed dispersal, populations spread alternatively as an advancing wave and as long-distance foundations of new foci which expanded into patches, then coalesced to form a new front (Fig. 3). The length of the colonization phase was very sensitive to variations of the two parameters *a* and *sd2* which control long-distance dispersal. Mean values over each set of four replicated simulations ranged from 151·2 generations for $a = 10^{-7}$ and $sd2 = 20$ km, to 21·7 generations for $a = 10^{-4}$ and $sd2 = 60$ km (Fig. 4). An average time of 49·7 generations (6 km per generation) was obtained for $a = 5 \times 10^{-6}$ and $sd2 = 50$ km. This last combination of parameter values was taken as our reference model for stratified dispersal.

It is worth noting that stratified dispersal was far more rapid than pure diffusion, even if long-distance dispersals were very rare events. In the case of our reference model, one of those events occurred on average each 1·2 generations.

(ii) *Genetic differentiation among populations*

When initial populations were undifferentiated, colonization with stratified dispersal resulted in a very rapid increase in G_{st} values in the early generations

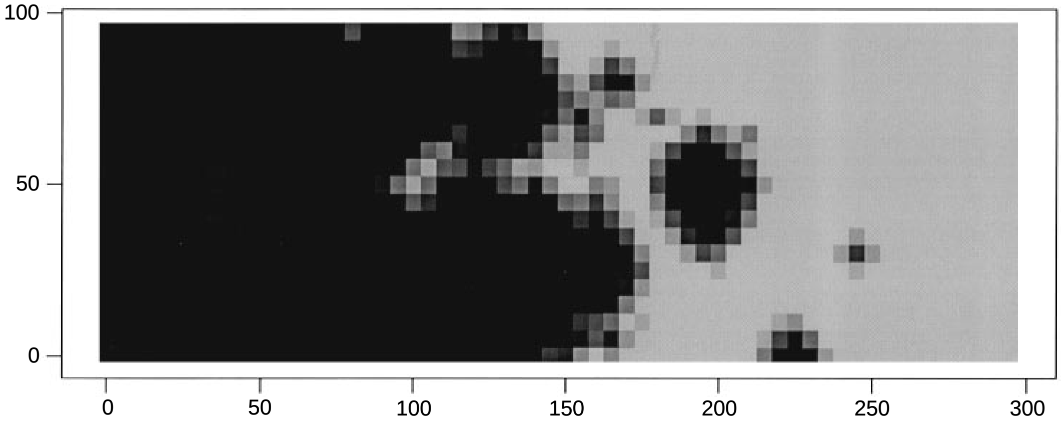

Fig. 3. Map of one simulated grid at generation 30 when colonization occurs by stratified dispersal with the parameters $a = 5 \times 10^{-6}$ and $sd2 = 50$ km. Non-empty populations are represented as black squares.

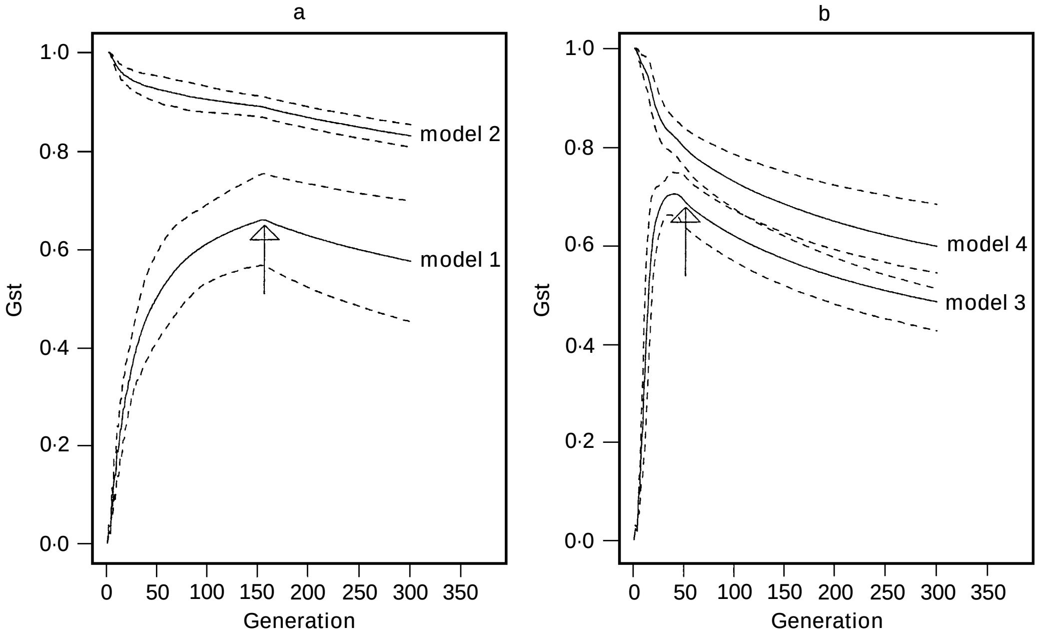

Fig. 5. Evolution of the genetic differentiation among populations during simulations, measured as Nei's (1973) G_{st}. Two different initial conditions are considered: $G_{st} = 0$ at the first generation and $G_{st} = 1$ at the first generation. (*a*) Diffusive dispersal (models 1 and 2). (*b*) Stratified dispersal (models 3 and 4). Continuous lines represent mean values of G_{st} over 20 replicated simulations. Dotted lines show 95% confidence intervals for the mean values, calculated with a normal assumption from the distribution of the 20 replicated values. Arrows indicate the mean time when colonization is achieved.

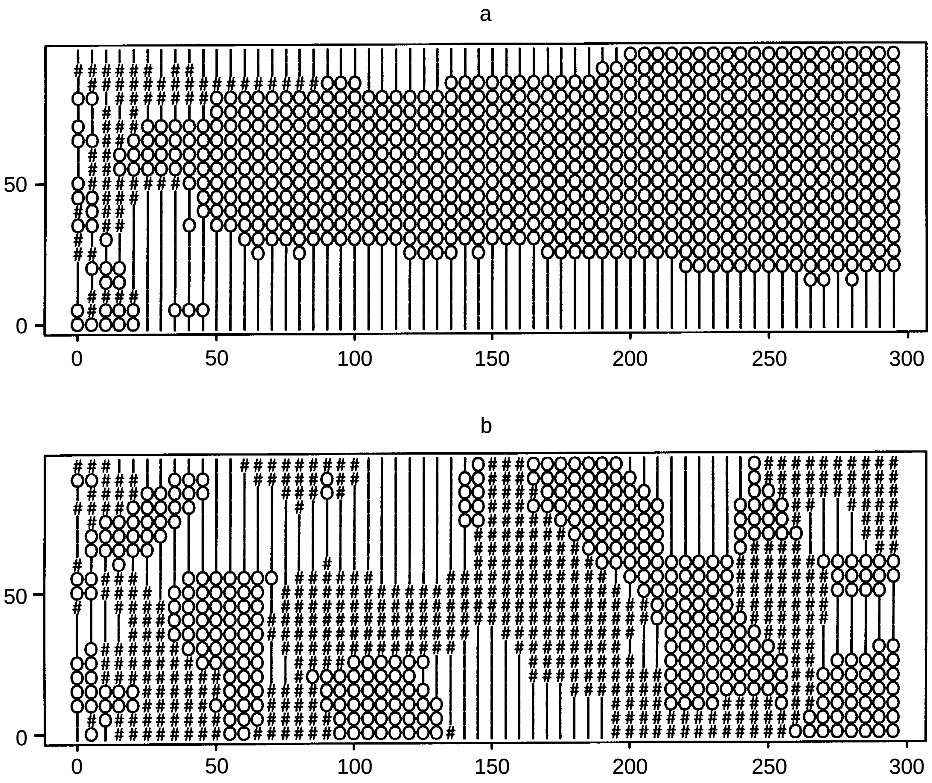

Fig. 6. Maps of predominant allele frequencies obtained at generation 300 after (a) colonization by diffusive dispersal and (b) colonization by stratified dispersal.

(Fig. 5). Diffusion led to a much more gradual increase in differentiation. When initial populations were completely differentiated, G_{st} decreased slowly during the simulations, whichever dispersal process was simulated.

Presence or absence of initial genetic differentiation

Table 1. *Autocorrelations of allele frequencies between neighbour populations calculated for different graphs of connection representing different directions of spread*

	Graph of connections			
Model	A (W–E)	B (S–N)	C (SW–NE)	D (NW–SE)
1	0·992^{a}	0·878^{b}	0·873^{b}	0·878^{b}
2	0·997^{a}	0·936^{b}	0·936^{b}	0·934^{b}
3	0·927^{a}	0·917^{a}	0·858^{b}	0·859^{b}
4	0·917^{a}	0·917^{a}	0·848^{a,b}	0·856^{a,b}

The values given are means over 20 replicated simulations. Values with the same letter are not statistically different (two-tailed test with a 5% error rate).

had almost no effect on the level of differentiation attained after colonization by stratified dispersal: from generation 112, the mean values of G_{st} for models 3 and 4 were no longer significantly different, since the 95% confidence intervals calculated from the two set of simulations overlapped (Fig. 5*b*). By contrast, at the end of the simulations, the value of genetic differentiation for model 2 (diffusion, initial $G_{st} = 1$) was still significantly higher than for model 1 (diffusion, initial $G_{st} = 0$).

(iii) *Spatial genetic structure*

(*a*) *Maps of allelic predominance*

Colonization by diffusion constantly led to the formation of large elongated areas parallel to the direction of the spread. A typical pattern is shown in Fig. 6*a*. The association of diffusion and long-distance dispersal led to more complex spatial patterns, often with an apparent mixing between a clinal trend and a patchy structure composed of more or less circular areas of allelic predominance. One typical map is shown in Fig. 6*b*.

(*b*) *Detection of directional trends*

All Moran's *I* values calculated for pairs of neighbouring populations using different graphs of connection were significant at the 5% probability level. For both diffusion processes (models 1 and 2), the autocorrelations were significantly higher for graph A (colonization following the west–east direction) than for the other three graphs (Table 1). Similar values of autocorrelation were obtained among these latter three graphs. These results clearly indicate the presence of one continuous trend in the west–east direction. By contrast, in the case of stratified processes, none of the graphs could be selected based on significantly higher autocorrelation values, although the higher values obtained for graph A seemed to indicate a preferred trend in the west–east direction.

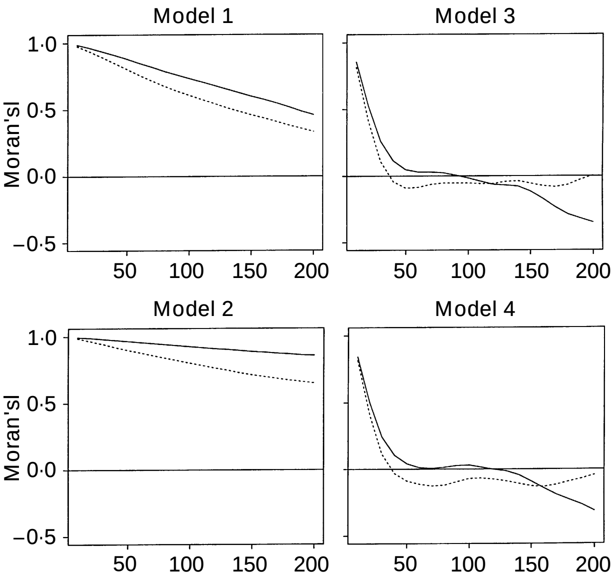

Fig. 7. Mean correlograms over 20 replicated simulations for each colonization model (model 1: diffusion, initial genetic differentiation is 0; model 2: diffusion, initial genetic differentiation is 1; model 3: stratified dispersal, initial genetic differentiation is 0; model 4: stratified dispersal, initial genetic differentiation is 1). Autocorrelations were calculated for both the original (continuous lines) or trend residual (dotted lines) allele frequency surfaces obtained at generation 300. Populations were connected according to the graph A in Fig. 2.

(*b*) *Directional correlograms for original and residual allele frequency data*

Correlograms were drawn by computing Moran's *I* values for pairs of populations connected according to graph A and separated by distances ranging from 10 to 200 km, with 10 km classes (Fig. 7). All correlograms were significant at the 5% level. For the two diffusion models, the original data and trend surface residuals showed identical patterns. Significant positive autocorrelations were found for all classes of distance and all correlograms were monotonically decreasing. These results indicate the presence of a long-range cline in the west–east direction, and the absence of local variations in this direction. For the two stratified models, original data showed a decrease in autocorrelations from significant positive values at low distances to significant negative ones at high distances, values at medium distances being non-significant. After elimination of the trend, correlograms typical of an isolation-by-distance pattern (Sokal & Wartemberg, 1983; Barbujani, 1987) were found, with significant positive autocorrelations at low distances, followed by values close to zero. The distance at the *x*-intercept of the correlogram drawn from trend residuals gives an approximation of the mean size of spatial patches (Sokal & Wartemberg, 1983; Bocquet-Appel and Sokal, 1989). For both stratified models it was near 40 km, which is also approximately the distance at which long-distance dispersal occurred.

4. Discussion

(i) *Colonization dynamics*

The first part of this work clearly confirmed Skellam's inference that a simple diffusion process cannot lead to rates of expansion such as those observed for forest trees, particularly for oaks. High spreading rates were obtained with rare long-distance dispersal events. This is in agreement with previous simulations of an invasion with satellite foci being formed (Moody & Mack, 1988) as well as with the predictions of analytical models (Shigesada *et al*., 1995). Dyer (1995) also showed by simulation that rare (one per generation) long-distance dispersal by birds could allow rapid 'jumps' of oaks to favourable habitats. As predicted by Mollison (1977), the stratified model led to an alternation of steady advances by diffusion and 'great leaps forwards'. The speed was, however, constant on average, and can be compared to palynological estimates, which range from 150 to 500 m yr^{-1} in Europe (Huntley & Birks, 1983). Estimated rates per generation are obtained by multiplying these numbers by the minimum age at first reproduction. In an established forest, oaks do not produce acorns until the age of 30–45 years (Jones, 1959), but the situation is different in open habitats: in a heath–oak wood transition, first reproductions occurred at about 20 years (Jensen & Nielsen, 1986). This leads to rates of spread of between 3 and 10 km per generation. The rate obtained with our stratified model (5·8 km per generation) is therefore consistent with palynological data, whereas the rate obtained with the diffusion model (1·9 km) is too low.

The standard deviation we used for long-distance events (50 km) seems rather excessive when compared with known dispersal abilities of biological vectors. Ethological studies on jays have underlined their great efficiency in propagating oaks, up to seven acorns being transported at a time, so that several thousands may be disseminated by a single jay each season (Schuster, 1950). Acorns are cached separately in the ground, with a preference for open habitats or transition zones (Bossema, 1979). Flight distances vary from tens of metres to kilometres, but the maximum observed was 8 km (Schuster, 1950). Two other birds seem to disperse nuts at greater distances: rooks (Ritchards, 1958), and migrating birds, passenger pigeons in America, or wood pigeons in Europe (Webb, 1986). Other more hypothetical vectors may be considered (for a review see Birks, 1989): human populations that established in Europe during the Neolithic, or physical vectors such as rivers. However, it is worth noting that long-distance dispersal may be very infrequent, and thus not easily observable. The maintenance of relict populations in favourable habitats could also have acted as foci to enhance range expansion. Palynological evidence for early patches of forest trees (including oaks) has been found in the northern Iberian Peninsula (Peñalba, 1994). But this cannot account for the rapid spread of forest trees in the north-western part of Europe, where permafrost was present during the glacial period (Dawson, 1992).

(ii) *Genetic differentiation among populations*

For both diffusive and stratified models, simulations show that colonization results in an increase in genetic differentiation when initial populations were undifferentiated. The effect of founding events has been described for a metapopulation undergoing extinctions and recolonizations (Slatkin 1977; Wade & McCauley, 1988; Whitlock & McCauley, 1990). It was shown that a new genetic equilibrium is attained, characterized by an increased genetic differentiation, except if gene flow is small and colonization occurs via high numbers of founding individuals. Compared with the nuclear genome, plant cytoplasmic genomes are particularly sensitive to founder effects, due to a lower effective number of genes and a lesser amount of gene flow by seeds than by pollen (Birky *et al*., 1983; Birky, 1991; Petit *et al*., 1993*b*; Wade *et al*., 1994). In our simulations we considered founding events but not extinctions. The cumulative effect of successive founding events caused a continuous increase in genetic differentiation until colonization was achieved. This increase was particularly rapid in the early generations of the stratified dispersal models, because long-distance colonization events could involve only one seed, leading to a total loss of genetic diversity until gene flow with other populations was re-established. By contrast, populations founded by diffusion maintain a higher diversity by exchanging seeds with the neighbouring populations. Values of genetic differentiation attained after stratified dispersal were moreover independent of initial values, because founding of populations from a single or a few seeds obliterated the previous genetic structure.

In all cases, the decay in G_{st} values after colonization was very slow. Because long-distance dispersals are very rare events, gene flow is essentially restricted to neighbour populations – a situation which may be compared with a two-dimensional stepping-stone model (Kimura & Weiss, 1964). Equilibrium values of differentiation for this model can be derived from the calculation of probabilities of identity by descent of two alleles within and among demes (Kimura & Weiss, 1964; Malécot, 1968), as described by Slatkin & Barton (1989, eqns. 5, 6). Following their approximations and assuming a low mutation rate of the order of 10^{-9}, we found that, in the case of our stratified dispersal model, the equilibrium value of differentiation should be of the order of 0·10. Some long-term simulations indicated that about 10000 generations are needed to attain this equilibrium (data not shown). In the case of perennial plants, colonization is thus likely to create non-equilibrium genetic structures which can persist over very long periods of time. For annual plants, the same dispersal model would give a migration rate much greater because all individuals would be replaced at each reproductive generation, and thus there would be a much more rapid return to equilibrium.

The actual values of genetic differentiation for the chloroplast genome over the whole range in the two main deciduous oak species in Europe, *Q. petraea* and *Q. robur*, are very high, being respectively 0·905 and 0·925 (Petit *et al.*, 1993*a*). Our simulated values are lower, indicating that less homogenization by gene flow occurs in reality than in our models. The main dispersal agent, the jay, is known to disperse fewer acorns, and less distantly, in dense woodlands than in open areas (Bossema, 1979). This discrepancy between dispersion in the colonization and post-colonization phases may result in a 'printing' of genetic structures after colonization, which could not be taken into account in our simulations.

(iii) *Spatial genetic structure of populations*

Diffusion resulted in long-range clines in allelic frequencies. No isolation-by-distance pattern was detectable, probably because it occurred at a scale beneath the scale of our analyses. The main effect of long-distance dispersal was to create local founding events which blurred the clinal pattern induced by the spread, so that the resulting pattern might be confounded with isolation-by-distance or regional differential selection. The size of the local patches due to founder events seems indicative of the distance at which long-distance colonization events occurred.

A field analysis of the spatial structure of choroplast DNA variation in *Q. petraea* and *Q. robur* was carried out in a 200 km × 270 km area in the north-west of France, with two to three individuals sampled every 20 km (Petit *et al.*, 1996). The three most frequent cytotypes detected were found to be patchily distributed. No predominant directional trend could be detected by directional correlograms. The *x*-intercepts of correlograms ranged from 35 to 45 km. According to our simulations, these results are consistent with the occurrence of long-distance dispersal of acorns during post-glacial recolonization. Alternative hypotheses such as the action of differential selection or large extinction/recolonization processes (due to fires or anthropic deforestation) seem unlikely (Petit *et al.*, 1996).

It is possible that other forest tree species whose seeds may have been dispersed over long distances during postglacial colonization, such as beech or chestnut tree, will share similar regional spatial patterns of variation for maternally inherited markers. The same effects of stratified colonization processes may also hold true for other plant species that have undergone recent geographical expansion. The invasion of *Bromus tectorum* into western North America is a well-studied example of a stratified colonization process that is certainly not unique among herbaceous plants: this species first appeared as scattered nuclei, probably generated by long-distance human transportation (Mack, 1981), then expanded to its present range in about 30 years. For plant populations that have recently passed through such colonization phases and have not yet reached their genetic equilibrium state, modes of dispersal during colonization seem to be relevant factors for explaining spatial structures at small and medium geographical scales. In contrast, assumptions based on gene flow in present populations are generally erroneous because they assume genetic equilibrium (Boileau *et al.*, 1992). This was demonstrated here for a maternally inherited marker and undoubtedly applies also to nuclear genes in the case of autogamous plants. Further experimental studies on colonizing plant species could lead to a better understanding of the consequences of various dispersal processes on spatial genetic structures at medium geographical scales.

References

Barbujani, G. (1987). Autocorrelations of gene frequencies under isolation by distance. *Genetics* **117**, 777–782.

Birks, H. J. B. (1989). Holocene isochrone maps and patterns of tree-spreading in the British Isles. *Journal of Biogeography* **16**, 503–540.

Birky, C. W., Jr (1991). Evolution and population genetics of organelle genes: mechanism and models. In: *Evolution at the Molecular Level* (ed. R. K. Selander *et al.*), pp. 112–149. Sunderland, Mass.: Sinauer.

Birky, C. W., Jr, Maruyama, T. & Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**, 513–527.

Bocquet-Appel, J. P. & Sokal, R. R. (1989). Spatial autocorrelation analysis of trend residuals in biological data. *Systematic Zoology* **38**, 333–341.

Boileau, M. G., Hébert, P. D. N. & Schwartz, S. S. (1992). Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *Journal of Evolutionary Biology* **5**, 25–39.

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

Cliff, A. D. & Ord, J. K. (1981). *Spatial Processes. Models and Applications*. London: Pion.

Cwynar, L. C. & McDonald, G. M. (1987). Geographical variation of lodgepole pine in relation to population history. *American Naturalist* **129**, 463–469.

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

Dawson, A. G. (1992). *Ice Age Earth: Late Quaternary Geology and Climate*. London: Routledge.

Dumolin, S., Demesure, B. & Petit, R. J. (1995). Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* **91**, 1253–1256.

Dyer, J. M. (1995). Assessment of climatic warming using a model of forest species migration. *Ecological Modelling* **79**, 199–219.

Fisher, R. A. (1937). The wave of advance of an advantageous gene. *Annals of Eugenics* **7**, 355–369.

Gabriel, K. R. & Sokal, R. R. (1969). A new statistical approach to geographic variation analysis. *Systematic Zoology* **18**, 259–278.

Harper, J. L. & White, J. (1974). The demography of plants. *Annual Review of Ecology and Systematics* **5**, 419–463.

Hengeveld, R. (1989). *Dynamics of Biological Invasions*. London: Chapman and Hall.

Huntley, B. & Birks, H. J. B. (1983). *An Atlas of Past and Present Pollen Maps for Europe: 0–13000 years ago*. Cambridge: Cambridge University Press.

Iida, S. (1996). Quantitative analysis of acorn transportation by rodents using magnetic locator. *Vegetatio* **124**, 39–43.

Jensen, T. S. & Nielsen, O. F. (1986). Rodents as seed dispersers in a heath–oak wood succession. *Oecologia* (Berlin) **70**, 214–221.

Jones, E. W. (1959). Biological flora of the British Isles: *Quercus* L. *Journal of Ecology* **47**, 160–222.

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

Mack, R. N. (1981). Invasion in *Bromus tectorum* L. into western North America: an ecological chronicle. *Agro-Ecosystems* **7**, 145–165.

Malécot, G. (1968). *The Mathematics of Heredity*. San Francisco, Calif.: Freeman.

Mollison, D. (1977). Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society*, *B* **39**, 283–326.

Moody, M. E. & Mack, P. N. (1988). Controlling the spread of nascent foci. *Journal of Applied Ecology* **25**, 1009–1021.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* **70**, 3321–3323.

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

Oden, N. L. (1984). Assessing the significance of a spatial autocorrelogram. *Geographical Analyses* **16**, 1–16.

Peñalba, M. C. (1994). The history of the Holocene vegetation in northern Spain from pollen analysis. *Journal of Ecology* **82**, 815–832.

Petit, R. J., Kremer, A. & Wagner, D. B. (1993*a*). Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics* **87**, 122–128.

Petit, R. J., Kremer, A. & Wagner, D. B. (1993*b*). Finite island model for organelle and nuclear genes in plants. *Heredity* **71**, 630–641.

Petit, R. J., Demesure, B. & Pineau, E. (1996). Genetic differentiation at a local and continental scale in European oaks species: the importance of historical factors. In: *Inter- and Intraspecific Variation in European Oaks: Evolutionary Implications and Practical Consequences* (ed. A. Kremer & H. Muhs), pp. 145–163. Brussels: European Union.

Ritchards, T. J. (1958). Concealment and recovery of food by birds, with some relevant observations on squirrels. *British Birds* **1**, 497–508.

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

Shigesada, N., Kawasaki, K. & Takeda, Y. (1995). Modeling stratified diffusion in biological invasions. *American Naturalist* **146**, 229–251.

Skellam, J. G. (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. *Theoretical Population Biology* **12**, 253–262.

Slatkin, M. & Barton, N. H. (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**, 1349–1368.

Sokal, R. R. & Oden, N. L. (1978). Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society* **10**, 199–228.

Sokal, R. R. & Menozzi P. (1982). Spatial autocorrelation of HLA frequencies in Europe support demic diffusion of early farmers. *American Naturalist* **119**, 1–17.

Sokal, R. R. & Wartemberg, D. E. (1983). A test of spatial autocorrelation using an isolation-by-distance model. *Genetics* **105**, 219–237.

Sork, V. L. (1984). Examination of seed dispersal and survival in red oak, *Quercus rubra* (Fagaceae), using metal-tagged acorns. *Ecology* **65**, 1020–1022.

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

Wade, M. J., McKnight, M. L. & Shaffer, H. B. (1994). The effects of kin-structured colonization on nuclear and cytoplasmic genetic diversity. *Evolution* **48**, 1114–1120.

Webb, S. L. (1986). Potential role of passenger pigeons and other vertebrates in the rapid Holocene migration of nut trees. *Quaternary Research* **26**, 367–375.

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

Webb, T. III (1985). Vegetation changes in eastern North America: an 18000 years history from pollen maps. *Studies in Plant Ecology* **16**, 100–102.

Whitlock, M. C. & McCauley, D. E. (1990). Some population genetic consequences of colony formation and extinction: genetic correlations within founding group. *Evolution* **44**, 1717–1724.