Molecular Ecology (2003) **12**, 2483–2495 doi: 10.1046/j.1365-294X.2003.01923.x

Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash

M. HEUERTZ,*† X. VEKEMANS,†‡ J.-F. HAUSMAN,* M. PALADA§ and O. J. HARDY†
Centre de Recherche Public-Gabriel Lippmann, CREBS Research Unit, avenue de la Faïencerie 162a, L-1511 Luxembourg, Luxembourg, †Université Libre de Bruxelles, Laboratoire de Génétique et Ecologie Végétales, chaussée de Wavre 1850, B-1160 Bruxelles, Belgium, ‡Université de Lille 1, Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016 — FR CNRS 1818, F-59655 Villeneuve d'Ascq cedex, France, §ICAS — Statiunea Simeria & Arboretumul Simeria, RO-2625 Simeria, Romania

Abstract

Spatial genetic structure was analysed with five highly polymorphic microsatellite loci in a Romanian population of common ash (*Fraxinus excelsior* L.), a wind-pollinated and wind-dispersed tree species occurring in mixed deciduous forests over almost all of Europe. Contributions of seed and pollen dispersal to total gene flow were investigated by analysing the pattern of decrease in kinship coefficients among pairs of individuals with geographical distance and comparing it with simulation results. Plots of kinship against the logarithm of distance were decomposed into a slope and a shape component. Simulations showed that the slope is informative about the global level of gene flow, in agreement with theoretical expectations, whereas the shape component was correlated with the relative importance of seed vs. pollen dispersal. Hence, our results indicate that insights into the relative contributions of seed and pollen dispersal to overall gene flow can be gained from details of the pattern of spatial genetic structure at biparentally inherited loci. In common ash, the slope provided an estimate of total gene dispersal in terms of Wright's neighbourhood size of $Nb = 519$ individuals. No precise estimate of seed vs. pollen flow could be obtained from the shape because of the stochasticity inherent to the data, but the parameter combinations that best fitted the data indicated restricted seed flow, $\sigma_s \leq 14$ m, and moderate pollen flow, $70 \text{ m} \leq \sigma_p \leq 140$ m.

Keywords: *Fraxinus excelsior*, microsatellites, neighbourhood size, pollen dispersal, seed dispersal, spatial genetic structure

Received 11 December 2002; revision received 3 April 2003; accepted 2 June 2003

Introduction

In natural plant populations, spatial genetic structure of sexually reproducing individuals is determined by the combined effects of gene flow, local genetic drift, various forms of natural selection and the spatial arrangement of individuals (Wright 1943; Epperson 1993; Doligez *et al*. 1998). Gene flow through pollen and seed dispersal is a key determinant in the establishment of genetic structure: when overall gene flow is restricted, a decrease in genetic relatedness among pairs of individuals is expected with increasing geographical distance between them, in agreement with models of isolation by distance (Wright 1943; Malécot 1950).

Many studies on forest tree species have tested for local spatial genetic structure using genetic markers and the tools of spatial autocorrelation analysis. In some species, spatial distribution of genotypes was close to random (Epperson & Allard 1989; Knowles 1991; Xie & Knowles 1991; Chung *et al*. 2000). These species featured either wide-ranging dispersal of both pollen and seed, as it is the case for wind-pollinated and wind-dispersed conifers (*Pinus contorta* ssp. *latifolia*, Epperson & Allard 1989; *P. banksiana*, Xie & Knowles 1991; *Picea mariana*, Knowles 1991) or, alternatively, they were insect-pollinated, but had very efficient seed dispersal through frugivorous birds (*Neolitsea sericea*, Chung *et al*. 2000). In most tree species on

Correspondence and present address: Myriam Heuertz, Uppsala University, Department of Conservation Biology and Genetics, Norbyvägen 18d, SE–752 36 Uppsala, Sweden. Fax: 0046 18 471 64 24; E-mail: myriamheuertz@gmx.net

the other hand, significant spatial genetic structure was detected and, concomitantly, either seed dispersal or both pollen and seed dispersal are restricted spatially. For example, different oak species (*Quercus rubra*, Sork *et al*. 1993; *Q. laevis*, Berg & Hamrick 1995; *Q. robur* and *Q. petraea*, Streiff *et al*. 1998) exhibit spatial genetic structure at a short spatial scale, which was attributed to localized, gravity-mediated seed dispersal, as pollen dispersal by wind is extensive in those species. Strong spatial genetic structure was found in tree species featuring substantial restrictions to both pollen and seed dispersal, i.e. pollination by small insects and seed dispersal by gravity (*Gleditsia triacanthos*, Schnabel *et al*. 1991; *Eurya emarginata*, Chung & Epperson 2000).

Relative pollen and seed contributions to gene flow can be estimated readily from genetic structure when it is analysed concomitantly at markers with different modes of inheritance (Latta *et al*. 1998; Asmussen & Orive 2000; Orive & Asmussen 2000; Latta *et al*. 2001). However, it remains unclear whether the relative pollen and seed contributions could be deduced from the observed patterns of spatial genetic structure at nuclear loci alone. Methods for estimating the extent of local gene flow from the observed spatial genetic structure within populations have been proposed recently (Hardy & Vekemans 1999; Rousset 2000) and applied to several plant species (Heuertz *et al*. 2001; Dutech *et al*. 2002; Fenster *et al*. 2003). These methods are based on a kinship–distance curve expressing the rate of decrease of genetic relatedness between individuals with geographical distance (Hardy & Vekemans 1999; Rousset 2000). Theoretical models of isolation by distance in a two-dimensional space predict that relatedness decreases linearly with the logarithm of the distance at a rate inversely proportional to $Nb \equiv 4\pi D\sigma_g^2$, where D is the effective population density, σ_g^2 is half the mean squared gene dispersal distance, and Nb may be interpreted as a neighbourhood size (Wright 1943, although Wright's interpretation of Nb as the size of 'panmictic' subpopulation units has little theoretical support; Rousset 1997). This relationship is actually valid only as long as kinship is observed between σ_g and $\sigma_g/(2\mu)^{1/2}$, where μ is the mutation rate (Rousset 1997, 2000). Gene dispersal is contributed by pollen and seed ($\sigma_g^2 = \sigma_s^2 + \sigma_p^2/2$, where σ_s^2, σ_p^2 are half the mean squared dispersal distances of seed and pollen, respectively; Crawford 1984), but their relative contributions cannot be inferred using this approach. However, because this method focuses on the kinship–distance curve within only a limited distance range, it does not exploit the whole information available. This study attempts to determine if it is possible to recover information on the seed and pollen contributions to total gene dispersal from spatial distributions of biparental genetic markers. The working hypothesis is that, outside the $\sigma_g - \sigma_g/(2\mu)^{1/2}$ distance range, the shape of the kinship–distance curve will change with the relative contributions of seed and pollen.

Common ash (*Fraxinus excelsior* L., Oleaceae) is a temperate tree species which occurs in mixed deciduous forests over almost all of Europe. It is wind-pollinated and its single-seeded fruits, the winged samaras, are wind-dispersed. Its mating system is polygamous: there is a continuum from pure male to pure female individuals with hermaphroditic intermediates (Picard 1982; A. Lamb and D. Boshier unpublished; Wallander 2001). In a previous study, we have identified strong genetic structure at short geographical scale (less than 1 ha) in four of 10 common ash populations from Bulgaria (Heuertz *et al*. 2001). This pattern was attributed tentatively to restricted seed dispersal (heavy wind-dispersed seed) and possibly restricted pollen dispersal. Our objective in the present study was to investigate whether a pattern of isolation by distance could be observed at a larger geographical scale in a common ash population from southeastern Romania and whether some inference could be drawn on the relative contributions of pollen and seed to total gene flow from the observed spatial genetic structure. For this purpose, we used intensive computer simulations of a theoretical population that matched closely features of the study population. We simulated different combinations of pollen and seed dispersal parameters and examined the resulting plots of relatedness against distance in order to (1) investigate how the relative contributions of seed vs. pollen dispersal affect the patterns of decrease of relatedness with distance and (2) infer levels of pollen and seed dispersal in the study population by comparing observed and simulated data.

Materials and methods

Plant material

The study site is a putatively autochthonous continuous mixed deciduous forest (approximately 710 ha) located in southeastern Romania (44°50′ N, 26°04′ E) near the city of Ploesti (Fig. 1). *F. excelsior* occurs there at high density, on average 200 mature trees (trees with diameter at breast height, d.b.h. > 15 cm) per ha, in a mixture with *Q. robur*, *Acer platanoides*, *A. campestre*, *A. tataricum*, *Ulmus minor* and other broadleaved tree species. An average of 30 non-adjacent common ash trees were sampled and mapped in each of five sampling locations, called subpopulations hereafter, covering an area of 0.71–3.69 ha (average 1.68 ha). Distances between subpopulations ranged from 514 to 3043 m. In total, 152 trees were sampled. Samples consisted of buds, which were shipped to the laboratory on their twigs wrapped in wet paper. They were then dissected out, separated from their scales and kept at –70 °C until DNA extraction.

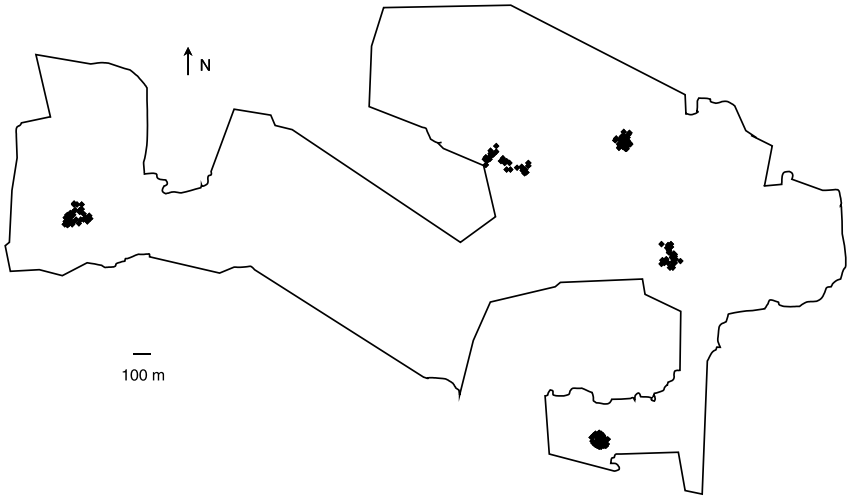

Fig. 1 Sampling locations of common ash trees in the Ploesti forest.

Table 1 Statistics of genetic diversity of microsatellite loci in the overall sample

Locus	*K*	*H*	F_{IS}[a]
M2-30	42	0.918	0.043
FEMSATL4	37	0.880	0.007
FEMSATL11	32	0.918	0.003
FEMSATL16	10	0.663	0.165
FEMSATL19	27	0.568	−0.066

K, total number of alleles; *H*, gene diversity; F_{IS}, Wright's inbreeding coefficient.
[a]Exact tests for departure from Hardy–Weinberg genotypic proportions were nonsignificant for all loci.

Microsatellite analysis

Total DNA was extracted with the CTAB procedure of the NucleoSpin Plant kit (Macherey Nagel) from 50 to 70 mg of buds ground in an automatic grinding mill (Retsch MM200). Microsatellite analysis was performed as described in Heuertz *et al*. (2001). Five highly polymorphic microsatellite loci were amplified with polymerase chain reaction (PCR) (Table 1). The quantity of template DNA in the PCR was increased to 3–10 ng compared with the previously published protocol, probably because metabolites susceptible to interfere with PCR were better eliminated with the DNA extraction kit used here. Fluorescent labelling of the forward primers allowed detection of amplification products on an automated DNA sequencer (ABI PRISM® 377 DNA sequencer). Sizing of fragments was performed with the software programs GENESCAN® 3.1 and GENOTYPER® 2.5 from Applied Biosystems by comparison with an internal sizing standard (GENESCAN-350 Rox). Another modification brought to the previously published protocol was that polyacrylamide gels were produced with 10% LongRanger® gel solution (Sanver Tech).

Microsatellite data analysis

The following genetic variation statistics were computed for each locus, and as averages over loci, for each of the five subpopulations and for the entire population with the program GEN-SURVEY (Vekemans & Lefèbvre 1997): (1) the average number of observed alleles *A*, (2) the average expected heterozygosity or gene diversity *H* corrected for small sample size (Nei 1978) and (3) Wright's inbreeding coefficient F_{IS} corrected for small sample size (Kirby 1975). Deviation of genotypic frequencies from Hardy–Weinberg proportions was tested with the program GENEPOP version 3.3d (Raymond & Rousset 1995) in order to check whether inbreeding occurred or microsatellites had null alleles.

To assess whether stepwise mutations affected genetic differentiation in the population, we applied the new test developed by Hardy *et al*. (2003) on microsatellite data. The principle of this test is to compute an R_{ST} (an analogue of F_{ST} based on allele size rather than allele identity; Slatkin 1995; Rousset 1996) among subpopulations and to compare its value with the distribution of R_{ST} values (called pR_{ST}) obtained after permuting randomly allele sizes among allelic states (1000 permutations). A significant test (i.e. observed R_{ST} larger than the 95% pR_{ST}) indicates that mutations contributed to the genetic differentiation among subpopulations.

Spatial genetic structure was analysed using kinship coefficients estimated relative to a sample of genotyped individuals: $F_{ij} \equiv (Q_{ij} - \bar{Q})/(1 - \bar{Q})$, where Q_{ij} is the probability that a gene drawn at random from individual *i* and a gene drawn at random from individual *j* are identical in state and $\bar{Q}$ is the mean Q_{ij} over all pairs of sampled individuals. Under the low mutation limit, F_{ij} approximates the ratio $(\Phi_{ij} - \bar{\Phi})/(1 - \bar{\Phi})$ where Φ_{ij} represents probabilities of identity by descent (Rousset 2002). Kinship coefficients were computed for all pairs of individuals in each subpopulation and in the entire population using the

statistic of Loiselle *et al*. (1995), which is defined for each allele k and each pair of individuals, i and j, as $F_{ij} = (p_i - \bar{p}_k)(p_j - \bar{p}_k)/(\bar{p}_k(1 - \bar{p}_k)) + 1/(2n - 1)$ where p_i, p_j are the frequencies of allele k in individuals i and j (taking the values 0, 0.5 or 1) and $\bar{p}_k$ is the average allele frequency of allele k in the reference population with sample size n. Average multiallelic (multilocus) estimates were computed by weighting the F_{ij} for each allele k by its polymorphism index $\bar{p}_k(1 - \bar{p}_k)$.

In order to test for isolation by distance, the multilocus kinship coefficient for each pair of individuals was plotted against the logarithm of the geographical distance separating them (kinship–distance plot). The significance of the linear regression slope, b_{ro}, was tested by Mantel tests (Manly 1997) with 10 000 permutations. The extent of gene dispersal was estimated from this slope as

$$\hat{N}b = -(1 - F_0)/b_{ro} \quad (1)$$

where F_0 is the kinship coefficient between adjacent individuals, i.e. in distance class zero. Note that the inference of Nb is most reliable within a distance range comprised between σ_g and $\sigma_g/(2\mu)^{1/2}$ (Rousset 1997, 2000) so that, assuming that the mutation rate μ approaches 10^{-3} for microsatellites, the upper bound of this range is about $20\sigma_g$. All computations were performed with the program SPAGeDi (Hardy & Vekemans 2002; http://www.ulb.ac.be/sciences/lagev). For graphical representation of kinship in the total population, average kinship coefficients were computed for 17 distance classes, equivalent to those used for the analysis of simulated data (see below).

Computer simulation procedure

In order to determine whether inference on the relative contributions of pollen and seeds to total gene flow could be drawn from spatial genetic structure, we compared the kinship–distance plot of the common ash population to kinship–distance plots from simulated data sets produced with various combinations of pollen and seed dispersal parameters. Computer simulations were carried out on a theoretical population that matched closely features of the real population. In the real population, the density was on average 200 mature trees/ha and the maximum distances between samples were 3.43 km and 1.75 km on the abscissa and the ordinate of an orthogonal grid. Hence, the population could be represented by 490×250 trees positioned on a rectangular grid where adjacent trees were separated by 7 m. To minimize edge effects, a larger theoretical population of 600×300 regularly spaced mature hermaphrodite diploid individuals was simulated considering overlapping generations. Simulations proceeded by the following steps: (1) an initial population is created by defining individual genotypes, choosing alleles stochastically at five loci according to allele numbers and frequencies observed in the common ash population; (2) each individual on the grid is given a 80% chance of surviving (assuming one simulation cycle corresponds to 1/5 of average lifetime); (3) each empty cell (death event) is replaced by an offspring whose maternal parent is chosen stochastically around the empty cell according to a predefined seed dispersal curve, and the paternal parent is chosen stochastically around the maternal parent according to a predefined pollen dispersal curve (individuals present at the previous step were considered when selecting parents); (4) the genotype of each new offspring is defined by choosing, at each locus, one allele of each parent at random; (5) once all individuals have been considered, a new population is defined by keeping previously surviving individuals and replacing dead ones by the offspring. Stages 2–5 were repeated 1000 times. This corresponds to 200 generations, which is a realistic figure for postglacial recolonized populations (Kremer 1994). Mutations were neglected as they did not affect differentiation in the real population according to the allele size permutation test (see Results). Seed and pollen dispersal followed an isotropic bivariate normal distribution characterized by parameters σ_s and σ_p, respectively. We ran 100 independent simulation replicates in each of 56 combinations of dispersal parameters, where $\sigma_s = 0.75$, 1, 2, 3, 4, 5, 7, 10 and $\sigma_p = 3$, 5, 7, 10, 20, 50, 100 (in lattice units; these values are to be multiplied by seven to obtain σ in metres). After 1000 simulation cycles, equilibrium spatial genetic structure was reached for most parameter combinations, except when both seed and pollen dispersal were very narrow. In each replicate, 152 individuals were then sampled at the grid positions corresponding to those closest to the position of sampled trees in the reference population if the latter had been overlaid by a 7×7 m grid. Average multilocus kinship coefficients per distance interval, F_k, were computed for the following 17 distance classes (upper boundary distance in grid units; $\times$ 7 to obtain metres): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 40, 200, 300, 600. We used short intervals for the first distance classes to obtain a detailed picture of how kinship varies with distance at a small spatial scale (i.e. within subpopulations, ≤ 40 lattice units). At larger spatial scales, wider intervals were used because kinship is then expected to vary little and estimates suffer less stochastic variance.

Analysis of simulated data sets

Simulation results were first used to investigate how different combinations of pollen and seed dispersal distances affected (i) the slope and (ii) the shape of the kinship–distance plots.

First, to verify the expected linear decrease of kinship–distance plots, we plotted $\bar{F}_k$ (the average F_k over 100

replicates for a given set of simulation parameters) against the logarithm of distance (measured in m), and computed the slope b_r of the linear regression (i) over all distances and (ii) over distances ranging from σ_g to 20 σ_g. To compute b_r, $\bar{F}_k$ values were weighted by the number of pairs of individuals belonging to class k, n_k:

$$b_r = \frac{\sum_k n_k(x_k - \bar{x})(\bar{F}_k - \bar{F})}{\sum_k n_k(x_k - \bar{x})^2}, \quad (2)$$

where x_k is the average of ln(distance between individuals), with distance in metres, for distance class k; $\bar{x}$ is the weighted average of x_k, $\bar{x} = \sum_k n_k x_k \Big/ \sum_k n_k$; and $\bar{F}$ is the weighted average kinship coefficient $\bar{F} = \sum_k n_k \bar{F}_k \Big/ \sum_k n_k$. Equation 2 ensures that b_r is essentially independent of distance class designations because it is nearly equivalent to regressing all pairwise F_{ij} values on the logarithm of distance. The fit of the linear regression for each parameter combination was evaluated with the coefficient of determination R^2.

The extent of local gene flow, σ_g, was estimated from the regression slope b_r: $\hat{\sigma}_g = \sqrt{\hat{N}b/(4\pi D_e)}$, where $\hat{N}b = -(1 - F_{k=1})/b_r$. Assuming that density was constant over time, the effective density D_e was computed from $D_e = D[4/(2 + V)][1/(1 + F_{IS})]$ (Crawford 1984), where D corresponds to the density of individuals, V is the variance of the lifetime reproductive success among individuals and F_{IS} is Wright's inbreeding coefficient. V and F_{IS} were recorded in the course of the simulations and we found $V \approx 5$ and $F_{IS} \approx 0.01$ for all parameter combinations, resulting in $D_e \approx 0.56D$. Similarly, as seed and pollen dispersal events occurring in the simulations were recorded, we also computed the realized pollen and seed dispersal standard deviations and neighbourhood size, σ_{pr}, σ_{sr} and Nb_r, respectively $\left(\sigma_r = \frac{1}{2}\sqrt{\sum r_i^2 \Big/ N}\right.$, where r_i is the distance crossed by the ith dispersal event, N being the total number of dispersal events recorded; $\left. Nb_r = 4\pi(\sigma_{sr}^2 + \frac{1}{2}\sigma_{pr}^2)\frac{4}{(2 + V)(1 + F_{IS})}\right)$. Realized values were slightly smaller than their corresponding parameter values because of edge effects and the use of a lattice rather than a continuous space. The estimates $\hat{\sigma}_g$ were compared to their expected values computed as $\sigma_{g\,exp} = \sqrt{\sigma_{sr}^2 + 0.5\sigma_{pr}^2}$.

The shape of dispersal curves was investigated further by examining plots of residuals

$$d_k = \bar{F}_k - F_{ke} \quad (3)$$

with F_{ke} being the kinship coefficient for class k expected from linear regression over the entire distance range. Residuals d_k were divided by

$$S = s\sqrt{1 - h_k}, \quad (4)$$

to obtain plots in standard deviation scale, where $s = \sqrt{\sum_k n_k d_k^2 \Big/ (n - 2)}$ is the unexplained standard deviation (i.e. not due to linear regression) and $h_k = 1/K + n_k(x_k - \bar{x})^2 \Big/ \sum_k n_k(x_k - \bar{x})^2$ the leverage coefficient of distance class k, as recommended by Sokal & Rohlf (1995: 531), weighted by the number of pairs of individuals in class k. The total number of classes K was 17. The visual appearance of plots of standardized residuals against distance suggested that they might be explained by polynomial regression. We fitted polynomial functions up to the third degree of the logarithm of distance to the observed standardized residuals plots following Sokal & Rohlf (1995: 616) and weighting residuals in each distance class by its respective number of pairs. Coefficients of these functions were compared to the ratio of seed vs. pollen contributions to total dispersal.

Fitting the observed spatial genetic structure to simulation results

Simulation results were also used to find out which combinations of seed and pollen dispersal distances best fitted the data. After ensuring that distance classes for the simulated and observed data were comparable, the fit of the observed kinship plot was investigated (i) using a χ^2 statistic describing the global departure between observed and simulated kinship coefficients over all distance classes and (ii) using two statistics based on the regression of kinship coefficients on the logarithm of distances, one being the regression slope, the other being a shape index based on the residuals. These statistics are defined as follows.

χ^2-like test. For each independent simulation replicate in each parameter combination, we computed:

$$\chi^2 = \sum_k \left[\frac{(F_k - \bar{F}_k)^2}{\mathrm{var}(F_k)}\right] \quad (5)$$

where F_k is the average multilocus kinship coefficient in distance class k, and $\bar{F}_k$ and var(F_k) are, respectively, the average and variance of F_k over the 100 independent simulation replicates of the investigated parameter combination. For each parameter combination, the observed χ_o^2 was obtained by replacing F_k values in eqn 5 by their respective values observed in the real population. The smaller χ_o^2, the better the simulations fitted the observation. Parameter combinations for which χ_o^2 was larger than the 95% lowest χ^2 values from independent replicates were considered incompatible with the common ash data, and therefore rejected (one-sided test).

Linear regression. For each independent simulation replicate in each parameter combination and for the real population, kinship coefficients F_k in each distance class k were plotted against the logarithm of the distance (measured in metres). The slope b_r of the linear regression was computed following eqn 2, with F_k instead of $\bar{F}_k$ (the whole distance range was considered). The slope b_{ro} of the observed kinship plot was compared to the 100 ranked b_r values from independent simulation replicates for each parameter combination. A parameter combination was considered incompatible with the observation when b_{ro} was smaller than the third or larger than the 98th ranked b_r value, i.e. lying outside a 96% confidence interval (two-sided test).

Polynomial fit of residuals. Residuals from the linear regression described above were computed for each independent simulation replicate in each parameter combination and for the real population as in eqn 3, with F_k replacing $\bar{F}_k$. After standardizing the scale following eqn 4, a polynomial function involving the third power of the logarithm of distance,

$$y = a + bx + cx^2 + dx^3, \quad (6)$$

was fitted to the plot of residuals against distance, weighting the residuals in each distance class by its number of observations. As we observed that the relative contribution of seed vs. pollen dispersal was related closely to the coefficient d of the polynomial (see Results), the coefficient d_o of the term in x^3 in the real population was compared to the 100 ranked values of d of independent replicates in each parameter combination. A parameter combination was considered incompatible with the observation when d was smaller than the third or larger than the 98th ranked d-value (two-sided test).

To represent graphically the quality of the fit to the simulation results according to each of the three statistics described above, we attributed P-values for each parameter combination according to the ranking of the statistic obtained for the common ash data with those of the simulation replicates. P-values were plotted on a grid representing all 56 parameter combinations and contours of $P = 0.05$, $P = 0.2$ and $P = 0.5$ were interpolated automatically with Microsoft Excel. Parameter combinations that had not been rejected in any of the three tests defined the space of parameters compatible with the common ash data.

Results

Microsatellite analysis of the common ash population

Allelic diversity of microsatellite loci. Between 10 and 42 alleles per microsatellite locus (148 alleles total for all loci) were observed in a total sample of 152 individuals. Gene diversity H ranged from 0.568 to 0.918 per locus, and Wright's inbreeding coefficient F_{IS} from −0.066 to 0.165 (Table 1). In the total sample, Hardy–Weinberg genotypic proportions were not rejected for any locus, indicating that inbreeding is at most weak and null alleles are rare or absent.

Table 2 Genetic diversity statistics within subpopulations and the total population

Population	n	A	H	F_{IS}[a]
Subpopulation 1	30	13.2	0.799	−0.011
Subpopulation 2	30	14.8	0.765	0.044
Subpopulation 3	30	13.8	0.773	0.045
Subpopulation 4	32	16.6	0.807	0.060
Subpopulation 5	30	14.8	0.794	−0.022
Mean subpopulations	30.4	14.6	0.788	0.023
Total population	152	29.6	0.792	0.029

n, sample size; A, average number of alleles per locus; H, gene diversity; F_{IS}, Wright's inbreeding coefficient.
[a]Exact tests for departure from Hardy–Weinberg genotypic proportions were nonsignificant for all subpopulations and for the total population.

Genetic diversity within subpopulations and the total population. Genetic diversity statistics for subpopulations and the total population are given in Table 2. Average number of alleles per locus A and gene diversity H were similar in all subpopulations; they ranged, respectively, from 13.2 to 16.6 and from 0.765 to 0.807. For the total population, $A = 29.6$ and $H = 0.792$. Wright's inbreeding coefficient F_{IS} was comprised between −0.022 and 0.06 within subpopulations ($F_{IS} = 0.029$ for the total population) and no departures from Hardy–Weinberg genotypic proportions were detected.

Impact of mutations on subpopulation differentiation. The R_{ST} among subpopulations ($R_{ST} = -1.5 \times 10^{-6}$) was not higher than the values after allele size permutations [mean pR_{ST} = 0.0037, 95% CI: (–0.0079, 0.0200)]. Hence, mutations did not contribute to genetic differentiation at this scale.

Spatial genetic structure. In agreement with models of isolation by distance, a significant linear decrease of estimated pairwise kinship coefficients F_{ij} with the logarithm of increasing geographical distance was detected in subpopulations 1, 2 and 4 ($P < 0.05$) and in the total population ($P < 0.01$; Table 3), although only few average kinship values per distance class lay outside the 95% confidence interval obtained from 10 000 random permutations of locations among individuals (Fig. 2). The neighbourhood size $\hat{N}b$ was estimated from eqn 1, using the regression

Table 3 Spatial genetic structure and estimates of gene dispersal

Population	Pairs of individuals	b_{ro}	*P*-value	F_0	$\hat{Nb}$	Sampling area (ha)
Subpopulation 1	435	−0.0177	0.0201	−0.012	57.3	0.83
Subpopulation 2	435	−0.0165	0.0217	0.054	57.3	1.38
Subpopulation 3	435	−0.0039	0.2475	0.047	nc	3.69
Subpopulation 4	496	−0.0170	0.0241	0.064	55.1	1.76
Subpopulation 5	435	0.0005	0.5001	−0.022	nc	0.71
Mean subpopulations	2236	−0.0106	0.0019	0.027	92.2	1.68
Total population	11476	−0.0019	0.0038	0.031	518.5	~710

b_{ro}, slope of the regression of pairwise kinship coefficients on the logarithm of geographical distance; *P*-value of the one-sided Mantel test with H_0: observed $b_{ro} = 0$; F_0, average kinship coefficient between adjacent individuals; $\hat{Nb}$, estimate of neighbourhood size; nc, not computed because the Mantel test was not significant.

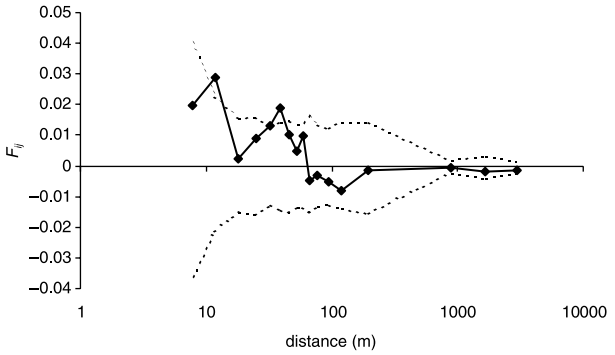

Fig. 2 Average kinship coefficients F_{ij} between pairs of individuals plotted against the logarithm of geographical distance in the whole population. Dashed lines represent 95% confidence intervals for F_{ij} under the null hypothesis that genotypes are randomly distributed.

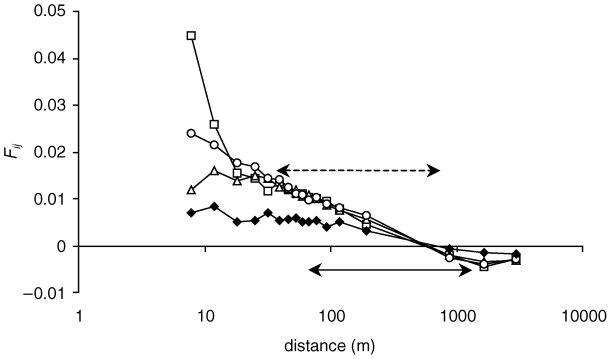

Fig. 3 Average kinship coefficients F_{ij} over 100 independent replicates of four simulated parameter combinations, plotted against the logarithm of distance. For squares, $\sigma_s = 5$ m, $\sigma_p = 49$ m; for circles, $\sigma_s = 14$ m, $\sigma_p = 49$ m; for triangles, $\sigma_s = 28$ m, $\sigma_p = 35$ m; for diamonds, $\sigma_s = 49$ m, $\sigma_p = 49$ m. Open symbols stand for parameter combinations with $\sigma_g \approx 35$–37 m, whereas filled symbols feature $\sigma_g \approx 60$ m. Arrows represent the distance range between σ_g and 20 σ_g, where a linear decrease of F_{ij} is expected, the dashed arrow corresponding to $\sigma_g = 37$ m, the solid arrow to $\sigma_g = 60$ m.

slopes observed for the different subpopulations and the whole population. In the three subpopulations with significant isolation by distance, $\hat{Nb}$ ranged from 55.1 to 57.3. These values are in agreement with those obtained previously at a similar spatial scale in Bulgaria ($\hat{Nb} = 37.9$–66.3, Heuertz *et al.* 2001). The average regression slope observed within all five subpopulations resulted in $\hat{Nb} = 92.2$. The regression applied to the whole population gave $\hat{Nb} = 518.5$, i.e. five times larger than at short spatial scale (Table 3). Note that regressions were applied over the whole available distance ranges, which did not necessarily correspond to the adequate distance range for *Nb* inference based on eqn 1.

Computer simulations

Analysis of simulated data sets. Within the distance range between σ_g and $20\sigma_g$, average estimated kinship coefficients decreased linearly with the logarithm of geographical distance (Fig. 3), as verified by the average value of the coefficient of determination, $\bar{R}^2 = 0.968 \pm 0.044$ (SD), for parameter combinations with $\sigma_p \leq 140$ m. For combinations with larger σ_p, the regression could not be computed because the upper boundary, $20\sigma_g$, lay outside the simulated distance range. Wide-ranging gene dispersal ($\sigma_g \approx$ 60 m, filled symbols in Fig. 3) produced shallower slopes than narrower dispersal ($\sigma_g \approx 37$ m, open symbols). Linear regression over the total distance range resulted in $\bar{R}^2 = 0.737 \pm 0.272$ (SD) for all parameter combinations and $\bar{R}^2 = 0.875 \pm 0.100$ (SD) for those with $\sigma_p \leq 140$ m, showing that deviation from linearity is substantially more pronounced than within the σ_g to $20\sigma_g$ distance range.

Interestingly, deviations from the linear relationship at distances shorter than σ_g depended strongly on the relative contributions of pollen and seed dispersal, as illustrated by three parameter combinations with similar total gene dispersal ($\sigma_g \approx 37$ m; Fig. 3): narrower dispersal of seeds than pollen produced an upward concave initial curving (squares in Fig. 3); and, as the relative seed dispersal

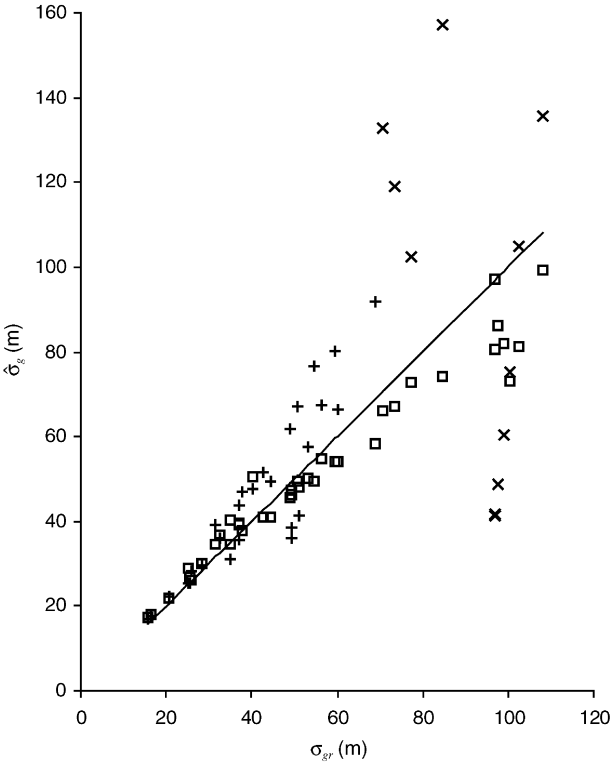

Fig. 4 Plot of the estimates of gene dispersal $\hat{\sigma}_g$ computed from the regression slopes b_r of kinship–distance plots from simulated data sets against the realized σ_{gr}. The line shows the expected relationship for an unbiased $\hat{\sigma}_g$ estimate. When regression is performed over an adequate range (σ_g to $20\sigma_g$, Rousset 1997, 2000), $\hat{\sigma}_g$ estimates are little biased (squares). When regression is performed within subpopulations, $\hat{\sigma}_g$ estimates are little biased under narrow dispersal ($\sigma_g < 70$ m, + symbols), but can become substantially biased under wide-ranging dispersal ($\sigma_g > 70$ m, × symbols).

contribution increased, the shape of the curve flattened (circles) before becoming downward concave (triangles). It is noteworthy that, in agreement with theoretical expectations, the three parameter combinations with contrasting σ_s and σ_p but very similar σ_g values show the same pattern of decrease of kinship coefficients within the σ_g to $20\sigma_g$ distance range (Fig. 3).

Estimates of standard deviations of gene dispersal distances $\hat{\sigma}_g$ using slopes b_r computed over the adequate distance range (σ_g to $20\sigma_g$, Rousset 1997, 2000) were close to the expected values (Fig. 4). When the regression was applied on too short a distance range, such as those within common ash subpopulations, substantial errors appeared in the estimates of σ_g (Fig. 4): dispersal was underestimated when the seed vs. pollen contribution was small (i.e. steep negative slope at short distances, Fig. 3); and it was overestimated when the relative seed contribution was high (shallower slope at short than at large distance, Fig. 3).

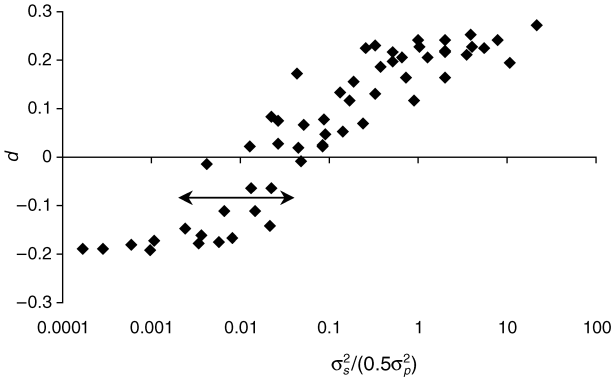

Fig. 5 Shape parameter d of the kinship–distance plots according to the ratio of seed vs. pollen contributions to total gene dispersal, $\sigma_s^2/(0.5\sigma_p^2)$. The quantity d corresponds to the coefficient of the term of third power of the polynomial functions ($y = a + bx + cx^2 + dx^3$) that were fitted to the standardized residuals of kinship coefficients (see text for details). The arrow stands for the observed d-value in the common ash population and the corresponding range of estimates of $\sigma_s^2/(0.5\sigma_p^2)$.

After subtraction of the linear regression slope for the whole distance range, polynomial functions of the third degree of the logarithm of distance were fitted to the standardized residuals, giving a mean $\bar{R}^2$ over all parameter combinations equal to 0.700 (± 0.201, SD). Cubic regression functions are indicators of the shape of plots of kinship against distance, independently of their slope. The form of these equations was $y = a + bx + cx^2 + dx^3$. All coefficients, a, b, c and d, were correlated with the ratios of contributions of seed and pollen to total gene dispersal, $\sigma_s^2/(0.5\sigma_p^2)$: R^2 values obtained from linear regression of each coefficient against the logarithm of $\sigma_s^2/(0.5\sigma_p^2)$ were between 0.791 and 0.846, and approximations with sigmoid curves with four parameters yielded R^2 values between 0.822 and 0.889. The narrowest relationships, i.e. the highest R^2 values, were observed for d, which corresponds to 1/6 times the third derivative of these functions (Fig. 5). This suggests that d might be a good indicator of the relative contributions of seed and pollen to gene dispersal.

Analysis of the fit of the observed spatial genetic structure to that obtained from simulations

χ^2-like test. A large proportion of simulated parameter combinations (36 of 56 combinations) were compatible with the observed data (Fig. 6a, shaded area); the fit was rejected ($\alpha = 0.05$) only when both pollen and seed dispersal were either narrow (approximately $\sigma_s \leq 21$ m and $\sigma_p \leq 35$ m), or wide-ranging (approximately $\sigma_s \geq 35$ m and $\sigma_p \geq 350$ m); or, alternatively, when seed dispersal was narrow and pollen dispersal wide-ranging (Fig. 6a, white area). A relatively good fit ($P > 0.2$) was obtained for $\sigma_s \leq 35$ m and $\sigma_p = 70$ m on one hand, and $\sigma_s = 49$ m and

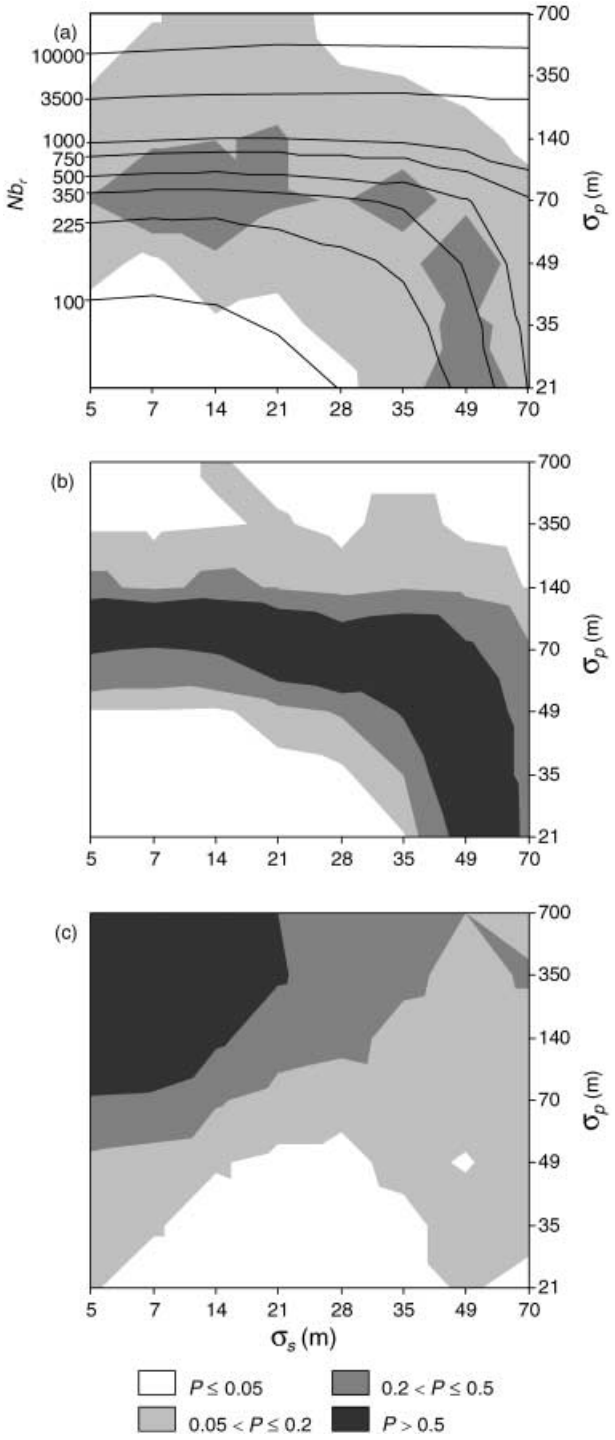

Fig. 6 Fit of kinship–distance plots from simulated data sets to the observed kinship–distance plot for three statistics: (a) χ^2; (b) slope, b_r statistic; (c) shape, d statistic. The axes represent the seed and pollen dispersal standard deviations used for simulations (scales not linear). Parameter combinations in white areas were rejected; darker shades represent an increasingly better fit of simulations with the observed data (see legend). The contours of shaded areas were interpolated. Lines of equal realized neighbourhood size Nb_r were also plotted in (a).

$\sigma_p \leq 49$ m on the other hand. These conditions correspond to a realized neighbourhood size approximately comprised between 200 and 1000.

Slope b_r. Wide-ranging pollen dispersal and a combination of narrow pollen and narrow seed dispersal were incompatible with the observed data (Fig. 6b). Overall, 31 of the 56 simulated parameter combinations remained compatible at $\alpha = 0.04$. A zone of relatively good fit (here: $P > 0.5$) was observed for roughly the same parameter combinations as in the χ^2-like test (compare Fig. 6a,b).

Shape d. Generally, narrow pollen dispersal ($\sigma_p \leq 49$ m) produced shapes of kinship plots differing from the observed data: although only 13 parameter combinations with $\sigma_p \leq 49$ m could be rejected at $\alpha = 0.04$ (Fig. 6c), several others were at the verge of significance. Good fit ($P > 0.5$) was observed for narrow seed dispersal ($\sigma_s \leq 14$ (21) m) and moderate to wide-ranging pollen dispersal ($\sigma_p \geq (70)$ 140 m).

The direct comparison of the shape parameter in the common ash population, $d = -0.084$, with the relationship of d to seed vs. pollen dispersal contributions (Fig. 5) produced an estimate of $\sigma_s^2/(0.5\sigma_p^2)$ comprised between approximately 0.002 and 0.04. The resulting ratio of seed vs. pollen dispersal standard deviations was $7.1 \leq \sigma_p/\sigma_s \leq 31.6$.

The comparison of the results from the three tests revealed that 24 of 56 simulated parameter combinations remained compatible with the observed data. Regarding pollen, very narrow as well as very wide-ranging dispersal conditions were effectively rejected, intermediate values of roughly 70 m $\leq \sigma_p \leq$ 140 m being in good agreement with the observed data. The picture was less clear for seed dispersal; no type of dispersal could be rejected efficiently. However, a good fit for both slope b_r and shape d was observed for intermediate pollen and narrow seed dispersal (Fig. 6b,c), delimiting an area of parameter combinations (70 m $\leq \sigma_p \leq$ 140 m and $\sigma_s \leq 14$ m) that were best compatible with the observed data.

Discussion

In the Ploesti forest in southeastern Romania, we detected spatial genetic structure in common ash in three of the five subpopulations and over the area of the total population. In general, kinship coefficients among pairs of individuals decreased with increasing logarithm of geographical distance between individuals. However, this decrease was not uniformly linear over the whole range of distances, as reflected by steep regression slopes at short distance, and a shallower slope at large distance. Resulting estimates of the neighbourhood size were more than five times larger at the scale of the total population than within subpopulations.

In order to find explanations for this observed pattern, possibly in terms of seed vs. pollen dispersal, we thoroughly examined the shape of kinship to distance plots from simulated data sets.

Estimating gene dispersal from spatial genetic structure by the regression approach

Simulations have verified that the decrease of kinship coefficients against the logarithm of distance is linear (Rousset 1997, 2000) when regression is performed over an adequate distance range (Fig. 3). The resulting estimates of gene dispersal are close to their expectations (Fig. 4), confirming the models of Hardy & Vekemans (1999) and Rousset (2000).

However, substantial errors occurred in the estimates of gene dispersal when regression was performed over too short a distance range (Fig. 4). Using the proper interval for regression is thus essential for a reliable estimate of *Nb*. As σ_g is usually unknown, the following approach can be attempted to determine the right interval: (i) estimating σ_g from the first *Nb* estimate, knowing D_e; (ii) estimating *Nb* again by performing linear regression over an interval of $\hat{\sigma}_g$ to $20\hat{\sigma}_g$; and (iii) repeating the procedure iteratively until the estimates $\hat{\sigma}_g$ and $\hat{N}b$ stabilize (e.g. Fenster *et al.* 2003). When this method was initialized on the entire data set using within-subpopulation dispersal estimates, successive $\hat{\sigma}_g$ values shifted to larger values, suggesting that the within-subpopulation scale is too narrow to estimate *Nb*. However, even at the level of the whole population, $\hat{\sigma}_g$ did not stabilize because the a posteriori regression slope became positive for $\hat{\sigma}_g \geq 53$ m or $\hat{N}b \geq 395$ (using $D_e = 0.56D$). This was attributed to stochastic variation of the data (see below). Hence, the suggested approach to determine the relevant interval for estimating *Nb* could not be applied to the current data set.

Impact of seed vs. pollen dispersal on the observed spatial genetic structure at nuclear loci

One possible explanation for the observed steeper decrease of kinship at short than at long distances is restricted seed vs. pollen dispersal, as revealed by our simulations (Fig. 3). More generally, simulations showed that a steep decrease of kinship at short distances occurs when gene dispersal follows a highly leptokurtic distribution (O.J. Hardy, unpublished data). The latter situation occurs when seeds are substantially less dispersed than pollen, because half the genes move over short distances (the maternal ones), and the other half can move over large distances (the paternal ones), producing a leptokurtic composite distribution. Highly leptokurtic gene dispersal distributions also occur when both seed and pollen dispersal distributions are (i) equal and (ii) highly leptokurtic themselves. In that case, our method would produce a wrong estimate of seed vs. pollen dispersal, as it assumes normality. Wind-dispersed pollen and seed, which are both characteristic for the common ash, may follow leptokurtic dispersal functions, as suggested in the literature. For instance, pollen deposition patterns in anemophilous plants were explained through leptokurtic dispersal functions (e.g. Levin & Kerster 1974; Ellstrand 1992; Richards 1997), and paternity analyses have often left high proportions of offspring with no potential father within the study population, demonstrating a strong long-distance component in wind-mediated pollen dispersal (Burczyk & Prat 1997; Streiff *et al.* 1999; Vassiliadis *et al.* 2002). Indications for leptokurtic dispersal of wind-borne winged seeds are weaker, referring mainly to an important deposition close to the source detected in direct studies (e.g. Levin & Kerster 1974; Johnson 1988). Wind-dispersed seed of tree species are not expected to move as far as pollen, because of their generally much greater mass. More restricted dispersal of wind-borne seed compared to pollen was also confirmed by a stronger spatial genetic structure at maternally than paternally inherited markers in conifer species (Latta *et al.* 1998; Liepelt *et al.* 2002). Hence, although it is likely that dispersal functions of pollen and seed in common ash may contain a leptokurtic component, potentially biasing our estimate of relative dispersal levels, we think that most of the leptokurtosis of gene dispersal is due to the relative difference between mean seed and pollen dispersal distances.

Implications of the shape of the kinship–distance plot on the estimate of dispersal

In the case of the common ash data, the steep negative initial slope of the kinship–distance plot and the demonstration that the within-subpopulation scale is inadequate for the regression approach indicate that we most certainly underestimated *Nb* at short distance.

Our results showed that the departure from linearity of the kinship–distance plot (its shape) depends on the relative contributions of seed vs. pollen dispersal, and can be characterized by fitting a cubic function (eqn 6) to the residuals of the kinship coefficients from the linear regression. The coefficient *d* of this function can provide an estimate of the relative magnitudes of seed and pollen dispersal, independently of the slope of the kinship–distance plot. The effect of seed and pollen on *d* is verified especially for seed contributions 1000 times smaller up to equalling pollen contributions (Fig. 5). Larger seed than pollen contributions probably do not reflect a biologically realistic situation in wind-pollinated and wind-dispersed species such as common ash, and seed contributions more than 1000 times smaller than pollen contributions represent most probably rather isolated cases of long-distance pollination.

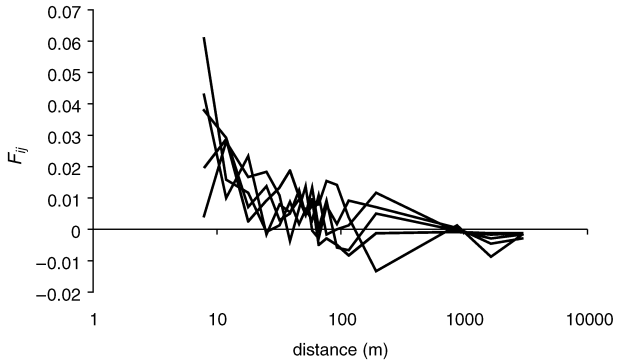

Fig. 7 Stochasticity in plots of average kinship coefficients F_{ij} against the logarithm of distance: observed values in the common ash population (bold line) and four independent simulation replicates for $\sigma_s = 7$ m and $\sigma_p = 70$ m (plain lines).

Power of methods and stochastic variation

The analysis of the fit of the observed spatial genetic structure to that obtained from simulations revealed that the decomposition of the dispersal function into a slope and a shape component allows one to extract more information than simply using the overall departure by a χ^2 statistic: the combined tests on the slope b_r and the shape d rejected 31 parameter combinations at $\alpha = 0.04$, whereas the χ^2-like test rejected only 20 at $\alpha = 0.05$. However, a large range of parameter combinations remained compatible with the observation. Although more simulation replicates (200 or 500 instead of 100) might have reduced the number of compatible combinations by rejecting those at the limit of significance, we argue that the main cause for weak rejection power is the stochasticity inherent in the data. Large fluctuations of kinship values were indeed found in the observed and the simulated data, despite the analysis of five highly polymorphic microsatellites in 152 individuals (Fig. 7). Such fluctuations could be related to the sampling scheme. However, simulations with other sampling locations but the same distance classes showed similar variation. The precision that can be reached on an *Nb* estimate based on spatial genetic structure is very dependent on the true *Nb* value: the stronger the spatial structure (i.e. the smaller the *Nb*), the smaller the coefficient of variation of the regression slope, hence the more precise the *Nb* estimate. The present study suggests that a larger sample size would have been necessary to infer gene dispersal with sufficient precision in common ash, where large gene flow occurs.

Estimates of dispersal in common ash

For the common ash, the results from our investigations suggest a neighbourhood size $\hat{N}b = 519$, relative contributions of pollen vs. seed dispersal of $7.1 \leq \sigma_p/\sigma_s \leq 31.6$ and restricted seed and moderate pollen dispersal, with most probable parameters of, respectively, $\sigma_s \leq 14$ m and $70 \text{ m} \leq \sigma_p \leq 140$ m. The latter would translate into a *Nb* estimate between 363 and 1468 when computed as $Nb = 4\pi D_e(\sigma_s^2 + 0.5\sigma_p^2)$.

These estimates of *Nb* are much higher than our previous results from Bulgaria ($\hat{N}b = 37.9-66.3$, Heuertz *et al*. 2001); but those estimates are probably biased downwards because the samples were taken at too-narrow distances. The Ploesti common ash population is an unusually dense one, as this species occurs mainly scattered in European mixed deciduous forests. The very restricted seed dispersal suggested here probably reflects the high vertical terminal velocities of ash seeds (1.2–1.7 m/s, Johnson 1988; Greene & Johnson 1995) dispersing under a relatively closed forest canopy where wind velocity and turbulence are reduced (Levin & Kerster 1974). Our estimate is smaller than that of Morand-Prieur *et al*. (pers. comm.), who found a mean seed dispersal distance of 88.4 m among a small sample of mother trees in a parentage study in a French common ash population (which corresponds to $\hat{\sigma}_s = 70.5$ m in case of a normal dispersal function, using distance $\hat{d} = \hat{\sigma}\sqrt{\pi/2}$). Various factors might contribute to explain this discrepancy, for instance a lower density and an elongated shape following a brook, with a preferential direction of seed dispersal in the latter population. Our own estimates of σ_s and σ_p might also be underestimated if the effective density in the real population was much lower than we assumed (i.e. $D_e/D = 0.56$) because of a high variance in lifetime reproductive success among adult trees. Estimates of pollen dispersal from the present study are slightly lower, but of the same order of magnitude than in the widely occurring European white oaks, where paternity analysis revealed $\sigma_p = 140$ m for *Q. petraea* and $\sigma_p = 176$ m for *Q. robur* (Streiff *et al*. 1999); and pollen sedimentation velocities in oaks and in ashes are similar (~3 cm/s, Jackson & Lyford 1999).

Concluding remarks

We have shown that, using nuclear loci alone, the shape of a kinship–distance curve bears information on the relative contributions of seed and pollen to gene flow. This is an important message because it may help interpret spatial genetic structure of plant populations in future studies. In particular, when seed disperses much less than pollen, kinship coefficients decrease faster with the logarithm of the distance at short distances than at large distances. Our data analysis suggests that the exploitation of this information requires a large sample size and/or a large number of highly variable loci to overcome the stochasticity of the genetic structure, particularly when gene dispersal is extensive (low level of genetic structure). Additional work

will be necessary to assess more effectively the conditions required for obtaining sufficiently precise estimates of the pollen vs. seed dispersal components. Moreover, the robustness of the method against different types of dispersal functions (e.g. leptokurtic ones) must be tested.

Acknowledgements

We are grateful to C. Muller from Laboratoire National de Santé, Department of Immunology, Luxembourg, for allowing us to use the automated sequencer, to M. E. Morand-Prieur and N. Frascaria-Lacoste for sharing unpublished results, and to three anonymous reviewers for helpful comments on an earlier draft of the manuscript. M. Heuertz acknowledges a scholarship (BFR 99/046) from MCESR (Ministry of Culture, Higher Education and Research) Luxembourg. This work was carried out with financial support from IPGRI (International Plant Genetic Resources Institute), project REGECON.

References

Asmussen MA, Orive ME (2000) The effects of pollen and seed migration on nuclear-dicytoplasmic systems. I. Nonrandom associations and equilibrium structure with both maternal and paternal cytoplasmic inheritance. *Genetics*, **155**, 813–831.

Berg EE, Hamrick JL (1995) Fine-scale genetic structure of a turkey oak forest. *Evolution*, **49**, 110–120.

Burczyk J, Prat D (1997) Male reproductive success in *Pseudotsuga menziesii* (Mirb.) Franco: the effects of spatial structure and flowering characteristics. *Heredity*, **79**, 638–647.

Chung MG, Chung MY, Oh GS, Epperson BK (2000) Spatial genetic structure in a *Neolitsea sericea* population (Lauraceae). *Heredity*, **85**, 490–497.

Chung MG, Epperson BK (2000) Clonal and spatial genetic structure in *Eurya emarginata* (Theaceae). *Heredity*, **84**, 170–177.

Crawford TJ (1984) The estimation of neighbourhood parameters for plant populations. *Heredity*, **53**, 273–283.

Doligez A, Baril C, Joly HI (1998) Fine-scale spatial genetic structure with non-uniform distribution of individuals. *Genetics*, **148**, 905–919.

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. *Molecular Ecology*, **11**, 725–738.

Ellstrand NC (1992) Gene flow by pollen: implications for plant conservation genetics. *Oikos*, **63**, 77–86.

Epperson BK (1993) Recent advances in correlation analysis of spatial patterns of genetic variation. *Evolutionary Biology*, **27**, 95–155.

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of lodge-pole pine. *Genetics*, **121**, 369–377.

Fenster CB, Vekemans X, Hardy OJ (2003) Comparison of direct and indirect estimates of gene flow in *Chamaecrista fasciculata* (Leguminosae). *Evolution*, **57**, 995–1007.

Greene DF, Johnson EA (1995) Long-distance wind dispersal of tree seeds. *Canadian Journal of Botany*, **73**, 1036–1045.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation and population genetics models. *Heredity*, **83**, 145–154.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618.

Heuertz M, Hausman JF, Tsvetkov I, Frascaria-Lacoste N, Vekemans X (2001) Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). *Molecular Ecology*, **10**, 1615–1623.

Jackson ST, Lyford ME (1999) Pollen dispersal models in quaternary plant ecology: assumptions, parameters and prescriptions. *Botanical Review*, **65**, 39–75.

Johnson WC (1988) Estimating dispersibility of Acer, Fraxinus and Tilia in fragmented landscapes from patterns of seedling establishment. *Landscape Ecology*, **1**, 175–187.

Kirby GC (1975) Heterozygote frequencies in small subpopulations. *Theoretical Population Biology*, **8**, 31–48.

Knowles P (1991) Spatial genetic structure within two natural stands of black spruce (*Picea mariana* (Mill.) B.S.P.). *Silvae Genetica*, **40**, 13–19.

Kremer A (1994) Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. *Genetics Selection Evolution*, **26** (Suppl. 1), 105–123.

Latta RG, Linhart YB, Fleck D, Elliot M (1998) Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. *Evolution*, **52**, 61–67.

Latta RG, Linhart YB, Mitton JB (2001) Cytonuclear disequilibrium and genetic drift in a natural population of ponderosa pine. *Genetics*, **158**, 843–850.

Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evolutionary Biology*, **7**, 139–220.

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial geneflow among refugia. *Proceedings of the National Academy of Sciences USA*, **99**, 14590–14594.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Malécot G (1950) Quelques schémas probabilistes sur la variabilité des populations naturelles. *Annales de l'Université de Lyon A*, **13**, 37–60.

Manly BFJ (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edn. Chapman & Hall, London.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Orive ME, Asmussen MA (2000) The effects of pollen and seed migration on nuclear-dicytoplasmic systems. II. A method for estimating plant gene flow from joint nuclear-cytoplasmic data. *Genetics*, **155**, 833–854.

Picard JF (1982) Contribution à l'étude de la biologie florale et de la fructification du Frêne commun (*Fraxinus excelsior* L.). *Revue Forestière Française*, **XXXIV**, 97–107.

Raymond M, Rousset F (1995) GENEPOP version 12. Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Richards AD (1997) *Plant Breeding Systems*, 2nd edn. Chapman & Hall, London.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Rousset F (2002) Inbreeding and relatedness coefficients: what do they measure? *Heredity*, **88**, 371–380.

Schnabel A, Laushman RH, Hamrick JL (1991) Comparative analysis of population genetic structure of two co-occurring tree species, *Maclura pomifera* (Moraceae) and *Gleditsia triacanthos* (Leguminosae). *Heredity*, **67**, 357–364.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 1463–1463.

Sokal RR, Rohlf FJ (1995) *Biometry: the Principles and Practice of Statistics in Biological Research*, 3rd edn. W.H. Freeman & Co, New York.

Sork VL, Huang S, Wiener E (1993) Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. *Annales des Sciences Forestières*, **50** (Supplement), 261s–270s.

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. & *Quercus petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.

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

Vassiliadis C, Saumitou-Laprade P, Lepart J, Viard F (2002) High male reproductive success of hermaphrodites in the androdioecious *Phillyrea angustifolia*. *Evolution*, **56**, 1262–1373.

Vekemans X, Lefèbvre C (1997) On the evolution of heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology*, **10**, 175–191.

Wallander E (2001) Evolution of wind-pollination in *Fraxinus* (Oleaceae) — an ecophylogenetic approach. PhD Thesis, Göteborg University.

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

Xie CY, Knowles P (1991) Spatial genetic substructure within natural populations of jack pine (*Pinus banksiana*). *Canadian Journal of Botany*, **69**, 547–551.

This study represents a portion of the doctoral research of Myriam Heuertz on population genetics of common ash from mainly southeastern Europe, which was carried out at CRP-Gabriel Lippmann and ULB under the supervision of Jean-François Hausman and Xavier Vekemans. Olivier Hardy is a research associate at ULB interested in describing population genetic processes. Jean-François Hausman is senior researcher at CRP-Gabriel Lippmann, working on the physiological response of woody species to environmental constraints. Magdalena Palada is senior researcher at ICAS, interested in micropropagation and regeneration of tree species. Xavier Vekemans is a lecturer at Lille University working in plant population genetics.
