

Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species

SYLVIE ODDOU-MURATORIO* and ETIENNE K. KLEIN†

*Ecologie des Forêts Méditerranéennes, UR 629, INRA, F-84914 Avignon cedex 9, France, †Biostatistique et Processus Spatiaux, UR 546, INRA, F-84914 Avignon cedex 9, France

Abstract

The comparison between historical estimates of gene flow, using variance in allelic frequencies, and contemporary estimates of gene flow, using parentage assignment, is expected to provide insights into ecological and evolutionary processes at work within and among populations. Genetic variation at six microsatellite loci was used to quantify genetic structure in the insect-pollinated, animal-dispersed, low-density tree *Sorbus torminalis* L. Crantz, and to derive historical estimates of gene flow. The neighbourhood size and root-mean-squared dispersal distance inferred from seedling genotypes ($N_b = 70$ individuals, $\sigma_e = 417$ m) were similar to those inferred from adult genotypes ($N_b = 114$ individuals, $\sigma_e = 472$ m). We also used parentage analyses and a neighbourhood model approach after an evaluation of the statistical properties of this method on simulated data. From our data, we estimated even contributions of seed- and pollen-mediated dispersal to the genetic composition of established seedlings, with both fat-tailed pollen and seed dispersal kernels, and slightly higher mean distance of pollen dispersal (248 m) as compared to seed dispersal (135 m). The resulting contemporary estimate of gene dispersal distance ($\sigma_e = 211$ m) was ~twofold smaller than the historical estimates. Besides different assumptions and statistical nuances of both approaches, this discrepancy is likely to reflect a recent restriction in the scale of gene flow which requires manager's attention in a context of increasing forest fragmentation.

Keywords: isolation by distance, neighbourhood model, parentage analyses, *Sorbus torminalis* L. Crantz, spatial genetic structure

Received 17 September 2007; revision accepted 19 March 2008

Introduction

Gene flow is a major evolutionary force, homogenizing allelic frequencies across populations through gene exchange, and shaping the effective size within local neighbourhood (*sensu* Wright 1943). Traditionally, standard methods for estimating gene flow from genetic data rely on measures of genetic differentiation among populations (Wright 1951) or among individuals within population (Rousset 2000). Assuming that the population obeys an infinite island model of population structure at evolutionary equilibrium, different variants of Wright (1951)'s F_{ST} have been used to estimate $N_e m_e$, the effective number of migrants per generation (see references in Whitlock & McCauley 1999). Alternatively, under isolation by distance among discrete

populations, linear regression of pairwise F_{ST} against distance can yield estimate of $N_e \sigma_e^2$, where σ_e^2 is the mean-squared axial dispersal distance (Rousset 1997). The increasing realization that the equilibrium hypothesis is critical for the reliability of gene dispersal inferences based on F_{ST} (Whitlock & McCauley 1999) has led to focus the estimation process on a local scale, where this equilibrium is more likely to be quickly reached and occurrence of mutations can be neglected (Leblois *et al.* 2003). Considering thus a continuous population exhibiting isolation by distance, the decay rate of genetic relatedness between individuals with distance has been shown to be proportional to $1/d_e \sigma_e^2$, d_e being the effective density of individuals (Rousset 2000; Vekemans & Hardy 2004). Intuitively, the product $d_e \sigma_e^2$ expresses the degree of overlap between individual 'gene shadows' (the spatial distribution of gene dispersal events around each parent). It implies that the intensity of genetic structuring decreases both with increasing dispersal and increasing

Correspondence: Sylvie Oddou-Muratorio, Fax: +33 490135 959; E-mail: oddou@avignon.inra.fr

individual density. Assuming that the genetic markers used in these analyses are not affected by natural selection, the summary parameter $d_e\sigma_e^2$ reflects the importance of both gene flow and genetic drift in shaping local differentiation at historical timescale.

More recently, contemporary estimates of gene flow have become available through assignment methods, which use individual multilocus genotypes instead of allele frequencies to ascertain population membership or parental origin of individuals (Manel *et al.* 2005). These approaches have benefited from the development of highly variable molecular markers (e.g. microsatellite loci) that provide unequivocal individual fingerprint, even with 5–10 loci and a large number of individuals analysed. Within continuous populations, patterns of gene flow between parental and offspring cohorts are classically addressed through parentage analyses (Meagher 1986; Morgan & Conner 2001). In the case of plant populations, parentage analysis consists in genotyping a sample of dispersed seeds or established seedlings and all the reproductive plants within a circumscribed area for a set of shared polymorphic markers, in order to detect the parent pair of each seedling (Meagher 1986). Then, a main problem is to discriminate between male and female parentage of seed/seedlings: a solution is to genotype maternally inherited tissues collected on dispersed seeds (Jones *et al.* 2005; Jordano *et al.* 2007). When dealing with established seedlings, where purely maternal tissues are no longer available, average effective pollen/seed dispersal distance can be directly estimated from parent–offspring genotype data using model fitting, such as the neighbourhood model (Burczyk *et al.* 2006). These approaches have gained a broad acceptance among population geneticists and ecologists because they allow first characterizing the seed and pollen dispersal process and the heterogeneity in male/female fertility at ecological timescale (Morgan & Conner 2001; Burczyk *et al.* 2006), and also investigating ecological factors that are likely to influence these patterns, such as parental phenotypic traits (Gonzalez-Martinez *et al.* 2006), seed dispersers behaviour (Jordano *et al.* 2007), or spatial environmental heterogeneity (Jones *et al.* 2005).

Contemporary estimates of gene flow, based on individual genotypes and parentage assignment and historical estimates, based on spatial genetic structure (SGS) are complementary approaches to gauge the respective importance of gene flow and genetic drift, with specific assumptions and advantages. A first assumption of historical estimates is that SGS is representative of an isolation-by-distance pattern at equilibrium. Second, SGS depends on the balance between gene dispersal and genetic drift (the product $d_e\sigma_e^2$), so that the estimation of gene dispersal per se requires an independent estimate of effective population density. By contrast, contemporary estimates can jointly provide detailed descriptions regarding the dispersal distribution and effective population density (e.g. Austerlitz

et al. 2004). Moreover, they rely on fewer assumptions, and can thus be used in more realistic situations, such as departure from equilibrium or temporal variance in demographic parameters (Manel *et al.* 2005). Therefore, contemporary estimates appear more relevant to address events occurring on short (ecological) timescales, that are typically considered as noise in equilibrium models (Manel *et al.* 2005). However, they are costly in terms of marker polymorphism and sampling effort, as they theoretically require exhaustive sampling of parental population at least within the circumscribed area under study. The risk of missing some candidate parents is clearly a strong constraint in parentage-based estimates of dispersal, in particular in long-lived organisms, where genotyped established seedlings can originate from parents that have disappeared from the population. However, only very few studies attempt to estimate the potential biases in contemporary estimates because of the subsampling of the parental population.

The comparison of historical vs. contemporary estimates of gene flow is expected to provide interesting insights on the balance and the timescale between evolutionary forces at work within and among populations. In undisturbed populations, consistency among historical and contemporary estimates has been argued to support the drift–migration equilibrium hypothesis (Dunphy & Hamrick 2005; Otero-Arnaiz *et al.* 2005), while divergence among real-time and effective migration rate was used to highlight the consequence of sex-biased dispersal and social structure in small mammal species (Schweizer *et al.* 2007). Alternatively, in a very recently disturbed population, only contemporary estimates should be affected because of temporal inertia of historical estimates for few (< 10) generations (Leblois *et al.* 2004) and the discrepancy among these both estimates would reflect the magnitude of the perturbation (Dutech *et al.* 2005). By contrast, consistent contemporary and historical estimates of gene flow in a recently perturbed area may indicate that compensatory mechanisms allow the maintenance of gene movements (Bacles *et al.* 2005). However, there are still too few thorough comparisons of genetic direct–indirect dispersal estimates across a variety of taxa to generalize their conclusions.

We focus here on the effects of forest fragmentation on a low-density, animal-dispersed, temperate forest tree: *Sorbus torminalis* (L Crantz). With a scattered distribution (usually less than one individual/ha) and a preference for open woodland, *S. torminalis* appears very sensitive to habitat fragmentation induced by silvicultural practices (Angelone *et al.* 2007). Indeed, changes in silvicultural practices in Central Europe have been creating for ~150 years forests with closed canopies, thereby favouring dominant and competitive species (oak, beech). Meanwhile, early successional species, encountered less favourable conditions and became scarce (Wohlgemuth *et al.* 2002). The decline

of scattered trees such as *S. torminalis* and other trees from *Sorbus*, *Malus* or *Pyrus* genera may also result from intense logging for their valuable wood. Finally, these insect-pollinated, animal-dispersed species may also be sensitive to indirect effects of habitat fragmentation, as the demography of pollinators and seed dispersers is also affected by silvicultural practices.

To study forest fragmentation consequences, we consider here both adult/young seedlings established, respectively, before and after change in silvicultural practices within a French managed forest. First, we use SGS to get a historical estimate of gene flow both in seedlings ($\sigma_{e\text{-seedlings}}$) and in adults ($\sigma_{e\text{-adult}}$) following standard isolation-by-distance theory for populations in a two-dimensional habitat (Rousset 2000; Vekemans & Hardy 2004). This stage-specific analysis is a first way to investigate the changes in the causal factors of SGS that may have resulted from forest fragmentation (Kalisz *et al.* 2001).

Second, we extend the pollen dispersal estimation method described in Oddou-Muratorio *et al.* (2005) to jointly estimate the seed and pollen dispersal curves, and the heterogeneity in male and female fertilities. Our method is related to the seedling neighbourhood model of Burczyk *et al.* (2006), which accounts for various reproductive factors (i.e. pollen and seed dispersal and male/female selection gradients) influencing the genetic composition of naturally established seedlings. However, whereas (Burczyk *et al.* 2006) explicitly model effective pollen/seed dispersal through an exponential distribution, our approach consists in directly estimating the shape and the scale of these dispersal curves. Additionally, we use computer simulations to assess the bias of direct gene flow estimates in the case of subsampling of the parental population, which may be a common flaw when assessing parentage of established seedlings of long-lived organisms such as trees.

Finally, we test the null hypothesis of no recent reduction in the scale of gene flow by comparing historical $d_e\sigma_e^2$ estimates for the adult and seedlings cohorts vs. contemporary estimates of the variance in effective gene dispersal (σ_{rt}). The observed discrepancy in σ_e vs. σ_{rt} is discussed in regard of the statistical nuances of both estimates, and of the expected effects of diverse ecological factors, including silvicultural practise on the building of SGS in this scattered species.

Materials and methods

Data

Rambouillet intensive study plot (ISP) covers 472 ha of mixed stands (oak, beech and secondary species), with 185 reproductive *Sorbus torminalis* exhaustively mapped and genotyped at six hypervariable microsatellite loci (Oddou-Muratorio *et al.* 2004). As in many forests in Central Europe, the traditional silvicultural system (an irregular mixture

of coppice under reserve trees) was changed in the 1960s for 'high forest' (a regular system where all trees on a same management unit have roughly the same age, and form a closed canopy; see Angelone *et al.* 2007 for details).

Past studies. We conducted a series of studies in Rambouillet ISP. The genetic structure within the adult population of *S. torminalis* (mostly established before this silvicultural change) indicated rather high levels of historical gene flow, with $\sigma_e \sim 360$ m (Oddou-Muratorio *et al.* 2004). The study population also showed high levels of contemporary pollen flow, with the real-time root-mean-squared axial pollen dispersal distance $\sigma_{p\text{-rt}} \sim 470$ m (Oddou-Muratorio *et al.* 2005). This contemporary estimate was achieved assuming an exponential power model for pollen dispersal and using the neighbourhood model applied to genotype data from a sample of fruiting mother trees, from a sample of seeds harvested on each of these mother tree, and from the blooming males within the ISP (Oddou-Muratorio *et al.* 2006).

Sampling design. This study was based on a sample of 439 young established seedlings (nonreproductive individuals, height < 5 m, expected to be 1–40 years old), grouped in 18 patches (9–40 seedlings/patch). Although isolated seedlings can be found, *S. torminalis* regeneration occurs in patches, which may correspond to favourable sites for seed deposition (e.g. under big old oak trees where birds perch or rest) and/or seedling development (shade/competition conditions). These patches were chosen in the central part of the ISP (Fig. 1), therefore a large range of distances between seedlings and adult trees were represented.

Genetic markers. For each sample, DNA was isolated from one to two frozen leaves per tree following the DNeasy procedure (QIAGEN). The genotypes of all individuals were scored at the same six microsatellite loci as used for the adult and seed samples of the previous studies and assessed through the same polymerase chain reaction (PCR) and electrophoresis procedure (Oddou-Muratorio *et al.* 2004, 2005).

Microsatellite markers were used to detect the occurrence of vegetative reproduction. As in Oddou-Muratorio *et al.* (2004), seedlings sharing exactly the same multilocus genotype were suspected to be clones. To validate this decision, we computed the probability $P_A(k)$ of observing a group of k or more individuals sharing the same multilocus genotype A without vegetative reproduction, following Oddou-Muratorio *et al.* (2004).

Indirect estimates of gene dispersal

Spatial genetic structure (SGS). SGS was characterized by the relationship between spatial distance and genetic relatedness as measured by the kinship (or co-ancestry) estimator F_{ij}

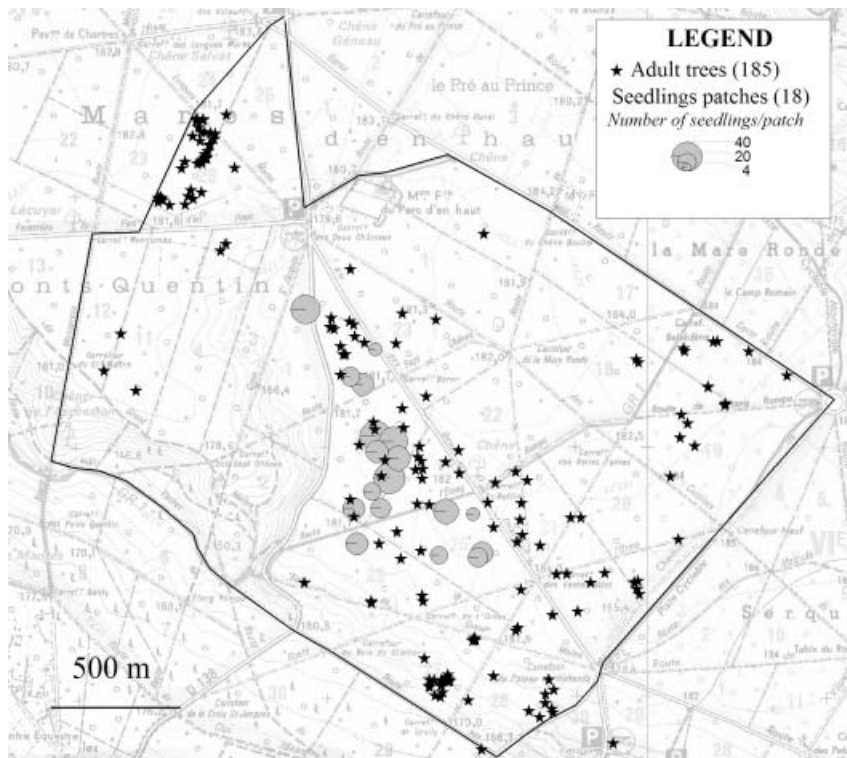


Fig. 1 Location within the study site of adult trees (◆) and seedlings patches (with size of the circle proportional to the number of seedlings/patch).

of Loiselle *et al.* (1995). For adult trees or seedlings arising from vegetative reproduction, only one ramet per genet was kept for the analyses. Using the computer software SPAGeDi (Hardy & Vekemans 2002), we successively computed F_{ij} between (i) all pairs of seedlings, and (ii) all pairs of reproducing individuals, using in both cases the pooled sample of adults and seedlings to estimate allelic frequencies.

To visualize SGS, F_{ij} values were averaged over a set of distance classes (d), giving, respectively, $F_S(d)$ and $F_A(d)$, and plotted against the logarithm of the distance. Approximate standard errors for the multilocus estimates of F_{ij} within each distance class were obtained through a jackknife procedure that consisted in deleting each locus one at a time. This assumes that the different loci provide independent replicates of the genetic structure process. To test SGS, the F_{ij} values were regressed on $\ln(d_{ij})$, where d_{ij} is the spatial distance between individuals i and j , to provide the regression slope b , respectively, denoted b_s and b_a for seedlings and adults. Then, the spatial positions of the individuals were permuted 10 000 times in order to get the frequency distribution of b_s and b_a under the null hypothesis that F_{ij} and d_{ij} were uncorrelated. Following Vekemans & Hardy (2004), the SGS intensity was quantified by $Sp = b/(F_1 - 1)$, where F_1 is the average kinship coefficient between individuals of the first distance class (< 50 m).

Gene flow estimates (σ_e^2). An indirect estimate of Wright's neighbourhood size ($N_b = 4\pi d_e \sigma_e^2$) can be obtained through:

$$N_b = -\frac{(1 - F_N)}{b_r} \quad (1)$$

where b_r is the restricted regression slope of F_{ij} on $\ln(d_{ij})$ in the range $\sigma_e > d_{ij} > 20\sigma_e$ (Rousset 2000; Vekemans & Hardy 2004), and where F_N is the kinship coefficient between neighbouring (competing) individuals (here F_N was approximated by F_1).

We used the iterative procedure detailed in Hardy *et al.* (2006) to estimate gene dispersal parameters for seedling and adult cohorts (respectively $\sigma_{e,S}^2$ and $\sigma_{e,A}^2$) assuming a range of values for the unknown d_e . The regression slopes (respectively b_{rs} and b_{ra}) of F_{ij} on $\ln(d_{ij})$ were computed over a restricted distance range (determined by convergence). For the extraction of σ_e estimates from the regression slopes, we assumed three d_e values (d_{obs} , $d_{obs}/2$ and $d_{obs}/4$), d_{obs} being the observed density of flowering trees ($d_{obs} = 0.39$ trees/ha). For a fixed d_e , approximate lower and upper bounds for the 95% confidence interval (CI) of N_b and σ_e were computed as detailed in Hardy *et al.* (2006).

Direct estimates of seedling dispersal

Mating model. Our mating model is a modified version of the seedling neighbourhood model of Burczyk *et al.* (2006). The main differences are that (i) we considered that all sampled adults are a single neighbourhood N (the whole ISP) for all seedlings; (ii) within the neighbourhood N , we

assumed a two-parameter dispersal kernel for seed and pollen dispersal.

We considered that each seedling i can be mothered either (i) by a mother tree located outside the neighbourhood N due to seed immigration (with probability m_s) or (ii) by a local mother tree located within the neighbourhood N [with probability $(1 - m_s)$]. In the latter case, we considered that offspring i may be the result either from self-pollination (with probability s), or from pollen flow from outside the neighbourhood N (with probability m_p), or from pollen from a sampled male (with probability $1 - s - m_p$). The probability of observing a multilocus diploid genotype G_i among the seedlings is thus:

$$P(G_i) = m_s T(G_i | BAF) + (1 - m_s) \sum_{j \in ISP} \Psi_{ij} \left[s T(G_i | M_j, M_j) + m_p T(G_i | M_j, BAF) + (1 - s - m_p) \sum_{k \in ISP} \Phi_{j,k} T(G_i | M_j, F_k) \right] \tag{2}$$

where $T(G_i | BAF)$ is the transition probability that an offspring immigrating from a mother tree located outside of the ISP has genotype G_i knowing the background allelic frequencies (BAF); $T(G_i | M_j, M_j)$, $T(G_i | M_j, BAF)$, $T(G_i | M_j, F_k)$ are the transition probabilities that an offspring has diploid genotype G_i when its mother tree of genotype M_j is, respectively (i) self-pollinated, (ii) pollinated by a father drawn at random in a population with allelic frequencies BAF, or (iii) pollinated by a father of genotype F_k . Parameter Ψ_{ij} is the proportion of seeds from mother tree j at the location of seedling i among the virtual seed pool originating from all known mother trees from the neighbourhood N :

$$\Psi_{ij} = \frac{\tau_{ij}}{\sum_{l \in N} \tau_{il}} \tag{3}$$

where τ_{ij} is a function of one or more factors influencing female reproductive success, including here the seed dispersal kernel (see below).

Parameter Φ_{jk} is the proportion of pollen from tree k in the pollen pool of mother tree j among the pollen pool from the known father trees from the neighbourhood N :

$$\Phi_{jk} = \frac{\pi_{jk}}{\sum_{l \in N} \pi_{jl}} \tag{4}$$

where π_{jk} is a function of one or more factors influencing male reproductive success, including here the pollen dispersal kernel (see below).

Models for pollen and seed dispersal kernel. As in Oddou-Muratorio *et al.* (2005), pollen dispersal was modelled using a dispersal kernel $p_p(\cdot; x, y)$ describing the probability for a pollen grain emitted at position $(0, 0)$ to participate to the

pollen cloud at position (x, y) . Following this scheme, we modelled seed dispersal using a dispersal kernel $p_s(\cdot; x, y)$ describing the probability for a seed emitted at position $(0, 0)$ to establish as a seedling at position (x, y) . We considered here the family of exponential power functions:

$$p(a, b; x, y) = \frac{b}{2 \sqrt{a^2 \Gamma(2/b)}} \exp \left\{ - \left(\frac{\sqrt{x^2 + y^2}}{a} \right)^b \right\} \tag{5}$$

where Γ is the classically defined gamma function (see, e.g. Abramowitz & Stegun 1964). The parameter b is the shape parameter affecting the tail of the dispersal function and a is a scale parameter homogeneous to a distance. The mean distance (δ) travelled by a pollen grain/seed under the kernel $p(a, b; \cdot)$ and the root-mean-square axial distance are given by Austerlitz *et al.* (2004) and Rousset (2004):

$$\delta = a \frac{\Gamma(3/b)}{\Gamma(2/b)} \quad \text{and} \quad \sigma = a \sqrt{\frac{\Gamma(4/b)}{2\Gamma(2/b)}} \tag{6}$$

Both increase with a and decrease with b . When $b < 1$, the dispersal kernel is fat tailed (Clark 1998), that is, the long-range decrease is slow, increasing the probability of long-distance dispersal events. Conversely, when $b > 1$ (for instance the Gaussian model, for which $b = 2$), the dispersal is thin tailed, with a rapid decrease of the dispersal function, implying much less long-distance dispersal events than when $b < 1$.

Model for female/male reproductive success (the τ_{ij} 's and the π_{jk} 's). Individual female reproductive success τ_{ij} of female j on seedlings location i was considered to be determined by two independent factors: (i) diameter of tree j as a surrogate for female fecundity, and (ii) the distance of seedling i from mother j and the seed dispersal parameters. Here τ_{ij} was expressed as:

$$\tau_{ij} = e^{\gamma_f D_j} \times P_s(a_s, b_s; x_i - x_j, y_i - y_j) \tag{7}$$

where D_j is mother j diameter, and a_s and b_s the parameters of the seed dispersal kernel p_s described by equation 5. Following Morgan & Conner (2001), we chose an exponential function of parameter γ_f to relate fecundity surrogate (here D_j) to reproductive success because it assures positivity of the reproductive success parameter. Additionally, this model of selection gradient is well-suited for continuously distributed variable such as diameter.

Similarly, individual male reproductive success π_{jk} of tree k in the local pollen cloud of mother j depended on two independent factors: (i) the diameter of tree k as a surrogate for male fecundity, and (ii) the distance of mother j to father k and pollen dispersal parameters. Therefore, π_{jk} was expressed as:

$$\pi_{jk} = e^{\gamma_m D_k} \times p_p(a_p, b_p x_j - x_k, y_j - y_k) \quad (8)$$

where D_k is the diameter of father k , γ_m is the male fecundity parameter, and a_p and b_p are the parameters of the pollen dispersal kernel p_p . Note that we could not include flowering intensity as a covariate in the model of male fecundity because we had measurements for this variable only for 1 year, whereas the analysed seedlings correspond to several years of recruitment.

Joint estimation of seed and pollen dispersal kernel. We obtained a maximum-likelihood estimate of the dispersal parameters (a_p, b_p and a_s, b_s), of the fecundity parameters (γ_f and γ_m), of the selfing rate (s) and of the pollen and seed immigration rates (m_p and m_s) by maximizing the log-likelihood function:

$$\log L(a_p, b_p, a_s, b_s, m_p, m_s, \gamma_f, \gamma_m) = \sum_{i=1}^S \log[P(G_i)] \quad (9)$$

where S is the total number of genotyped seedlings. For convenience, we estimated directly δ -parameter (see equation 6) rather than α .

The fits were achieved following a quasi-Newton algorithm, using MATHEMATICA 5.2 (Wolfram Research) and classical statistical procedures associated with likelihood. Also, parametric bootstrap confidence intervals were computed for the complete model (but without selfing parameter s), as described in online Appendix 2 of Oddou-Muratorio *et al.* (2005). Briefly, we used the estimated values of the parameters to simulate 100 genotypic data sets similar to the experimental data set (see below for the detailed simulation algorithm), and we re-estimated the parameters from the simulated data sets. Then, the empirical distributions for the parameters were used to compute confidence intervals.

Simulations of missing parents. As in Burczyk *et al.* (2006), our neighbourhood model uses as input data genotypes for biparentally inherited markers, that cannot separate seed-from pollen-mediated gene flow. To circumvent this problem, we assume that when only one compatible genotype is found within the neighbourhood of a given seedling, it is that of the seed donor. This assumption is reasonable, considering that seed dispersal has been independently shown to be slightly more restricted than pollen dispersal (Oddou-Muratorio *et al.* 2001). Although, it may not hold (i) when the mother tree is just outside the ISP, so that some father trees siring it can still be found within the ISP; (ii) when random mortality or logging suppress the mother tree but not the father tree; or (iii) when genotyping errors randomly affect the maternal contribution to the seedlings but not the paternal one. In these cases, direct estimation procedure is expected to provide underestimates of seed dispersal, and overestimates of pollen dispersal.

We thus used Monte Carlo simulations to assess the bias of the direct estimates due to this conservative hypothesis. First, we randomly sampled a given number of seedlings in the seed rain generated as follows:

- Seed and pollen donor's population.* We used the spatial distribution, genotypes and diameters of the 185 adult trees within the ISP. Additionally, 20 unknown adult trees were simulated within the ISP (NS-W for nonsampled within): their spatial positions were random, their diameter was drawn uniformly between 10 and 60 cm, and their genotypes were drawn in the BAF. Finally, new, nonsampled adult trees were also simulated on a 16 km \times 16 km perimeter around the ISP (NS-O for nonsampled outside). Based on the spatial distribution observed for *Sorbus torminalis* within the study site (Oddou-Muratorio *et al.* 2004), NS-O trees were distributed following a Neyman-Scott clustered point process of the same density (0.35 tree/ha), with an average of 10 trees per cluster, and square cluster of 1.6 km². Their diameter and genotypes were generated as for NS-W trees.
- Seedling mothers.* The seed rain above each seedlings patch (the ψ_{ij} in equation 3) was predicted at the centroid of each patch using the spatial positions and diameters of all adult trees over the 16-km² area, and equation 9, with seed dispersal parameters (a_s and b_s) and selection gradient on female fertility γ_s fixed to previously estimated values. Then, 429 seedlings were randomly sampled from this seed rain (N_i seedlings/patch with N_i following our experimental design), by drawing their mother following the relative contributions of each tree in the seed rain, and the maternal gamete following Mendelian rules of inheritance.
- Seedling fathers.* The pollen cloud above each mother tree actually contributing to the seedlings (the Φ_{jk} in equation 4) were computed using the spatial positions and diameters of all adult trees over the 16-km² area, and equation 9, with pollen dispersal parameters (a_p and b_p), and selection gradient on male fertility γ_m fixed to previously estimated values. Then, we randomly selected 429 fathers for the seedlings, following the relative contributions of each tree in each maternal pollen cloud. Their paternal gametes were drawn following Mendelian rules.

We simulated 50 data sets, and then we retrospectively estimated seed and pollen dispersal and fertility parameters using the same procedure. However, we considered as sampled parents only the 185 adult trees within the investigated area. Thus, NS-W trees correspond to either nonsampled or removed parents from the ISP. We computed the mean of the estimated dispersal parameters (a_p, b_p and a_s, b_s), fertility parameters (γ_f and γ_m), and pollen and seed immigration rates (m_p and m_s), and the biases in these estimates (mean minus true value). Note that the estimated

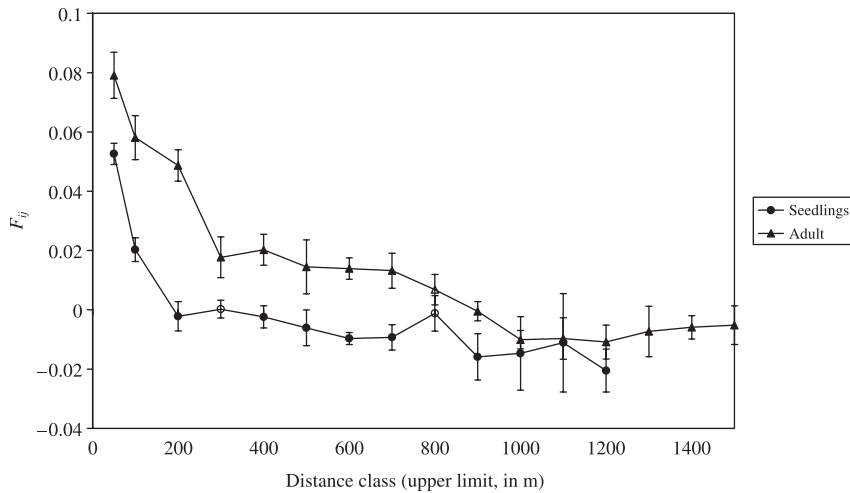


Fig. 2 Correlograms of F_{ij} coefficient across the central part of the ISP for seedlings (●) and adults (▲). Abscise values correspond to the upper limit of the distance intervals. Filled symbols represent values significantly different from the expected value under a random distribution of genotypes (95% confidence level). Confidence intervals around each F_{ij} -value were obtained through a jackknife procedure over loci.

Table 1 Estimates of spatial genetic structure (SGS) and of historical gene flow for seedlings and adult cohorts: inbreeding coefficient (F_i), average kinship coefficient between individuals separated by less than 50 m (F_1), SGS intensity (Sp) and its standard error (SE), neighbourhood size (N_b) and 95% confidence interval, gene dispersal distance (s_e) and 95% confidence interval for three effective densities (d_e) estimated from the observed adult density (d_{obs})

Cohort	SGS parameters				Gene dispersal parameters									
	F_i	F_1	Sp	(SE)	$(d_e = d_{obs})$		$(d_e = d_{obs}/2)$		$(d_e = d_{obs}/4)$					
					N_b	σ_e (m)	N_b	σ_e (m)	N_b	σ_e (m)				
Seedlings	0.0609***	0.0514***	0.0141	0.0008	70	(26; +∞)	417	(231; +inf)	81	(21; +∞)	506	(346; +inf)	—	—
Adult	-0.007NS	0.0794***	0.0187	0.0023	114	(42; +∞)	472	(416; +∞)	—	—	—	—	—	—

immigration rate is composed of a long-distance component (i.e. pollen or seed from NS-O trees), and of a short-distance component (i.e. pollen or seed from NS-W trees).

Results

Importance of vegetative reproduction

None of the 439 analysed seedlings shared exactly the same multilocus genotype with any of the 185 genotyped adult trees. Among these 439 seedlings, 20 seedlings showed non-unique genotypes and were grouped into nine genets. In the first genet, two distant seedlings (separated by > 400 m) shared the same genotype G but with a probability $P_G(2) = 0.04$ (mostly due to missing data at three loci among six). These two seedlings were thus kept in the following analyses. For the remaining eight genets (18 seedlings), the short distance between seedlings within genet, combined with the low probabilities of obtaining these groups of seedlings by chance (from 1.10^{-18} to 2.10^{-8}) strongly supported the hypothesis of vegetative reproduction. For the following geneflow analyses, we thus kept only one seedling per genet, eliminating 10 clonal genotypes. This left us with a data set of 429 nonclonal seedlings.

SGS and indirect geneflow estimates

In agreement with models of isolation by distance, a significant linear decrease of F_{ij} -coefficients with the logarithm of geographical distance was detected both in adult and seedling cohorts (Fig. 2). Broadly stronger SGS was detected in adults as compared to seedlings, as depicted by the Sp -statistics ($Sp = 0.0187$ in adults vs. $Sp = 0.0141$ in seedlings, see Table 1). This was mainly due to higher levels of relatedness between neighbouring adults ($F_1 = 0.0794$) as compared to neighbouring seedlings ($F_1 = 0.0514$). Though, inbreeding coefficient was significant in seedlings ($F_1 = 0.0609$) but not in adults, which could be explained by a progressive purge of inbred individuals throughout lifespan (Bush & Smouse 1992). Finally, the shape of the kinship curves were slightly similar for adults and seedlings, with a steep regression slope at short distance (< 300 m) and a shallower slope at large distance (Fig. 2).

The procedure used to estimate dispersal parameters converged for both adult and seedlings cohorts only when assuming $d_e = d_{obs}$ and also for seedlings when assuming $d_e = d_{obs}/2$. Using equation (1) and assuming $d_e = d_{obs}$, the neighbourhood size was estimated to 70 individuals in the seedlings cohort, vs. 114 individuals in the adult cohort,

Table 2 Direct estimates of contemporary gene flow and mating system parameters and 95% parametric bootstrap confidence interval (CI): mean distances of seed (δ_s) and pollen (δ_p) dispersal, scale and shape parameter of the seed (a_s and b_s) and pollen (a_p and b_p) dispersal kernels, rate of seed (m_s) and pollen (m_p) migration from outside the ISP, selfing rate (s), selection gradient of tree diameter on female (γ_f) and male (γ_m) fertility

Parameter	Seed dispersal and female fertility					Pollen dispersal and male fertility					
	δ_s	a_s	b_s	m_s	γ_f	δ_p	a_p	b_p	m_p	s	γ_m
Estimated value	135	11.7	0.56	0.39	0.06	248	47.9	0.69	0.69	0.02	0.09
CI -	107	1.4	0.4	0.35	0.04	158	9.6	0.52	0.64	—	0.06
CI +	177	93.6	1.03	0.44	0.08	368	312.4	1.43	0.77	—	0.13

while the average gene dispersal distance were, respectively, estimated to 417 m in the seedlings vs. 472 m in the adults, with large confidence intervals in both cases. In the seedlings case, the convergence procedure cycled periodically between two σ_e -values (~ 390 m and ~ 450 m). The estimation procedure failed to converge in most cases with $d_e < d_{\text{obs}}$.

Direct gene flow estimates

Maximum-likelihood estimates of contemporary gene flow and mating system parameters were obtained using the neighbourhood model (see equation 2) and the genotypes and spatial positions of all the 185 adult trees found without the ISP, and of 424 nonclonal seedlings (five seedlings were discarded due to missing data at > 3 loci).

For both pollen and seed dispersal, we observed a general tendency of fat-tailed dispersal kernel (i.e. $b < 1$), with $b_s = 0.56$ and $b_p = 0.69$ (Table 2). Despite rather large confidence intervals around \hat{b}_s - and \hat{b}_p -values, the exponential power kernel provided a better fit than the exponential (i.e. when fixing $b = 1$ in equation 10) both for pollen and seed dispersal (P -value $< 10^{-3}$ for both likelihood ratio tests, fixing in turn $b_p = 1$ and $b_s = 1$). We estimated a slightly larger mean distance for pollen dispersal ($\hat{\delta}_p \approx 248$ m; median = 188 m) than for seed dispersal ($\hat{\delta}_s \approx 135$ m; median = 96 m). Also, the pollen immigration rate ($\hat{m}_p = 69\%$) was significantly higher than the seed immigration rate ($\hat{m}_s = 39\%$).

Diameter had a significant effect both on male and female fecundity (P value $< 10^{-3}$ for all three likelihood-ratio tests, fixing in turn $\gamma_m = 0$; $\gamma_f = 0$; and $\gamma_m = \gamma_f = 0$ in equation 9). Both male and female fecundities increased with diameter ($\hat{\gamma}_m = 0.09$ and $\hat{\gamma}_f = 0.06$), but the selection gradient was steeper for male fecundity than for female fecundity.

Impact of missing parents on direct estimates of gene flow

The seed immigration rate realized in simulations ('true' $m_s = 5.8\%$, Table 1) was contributed at 58% by NS-W trees (nonsampled adult within neighbourhood, not listed as a potential parent in the estimation procedure). This realized

m_s was underestimated on average by 53% ('estimated' $\hat{m}_s = 2.7\%$), whereas the pollen immigration rate (m_p) was overestimated by 30%. These results are consistent with the conservative hypothesis of our model. Indeed, for all the seedlings mothered by one of NS-W/NS-O adults (i.e. not listed as a potential parent) and fathered by one of the 185 'sampled' adults, the estimation procedure systematically considered the father tree as the mother tree (false mother), thereby missing a seed immigration event and adding a false pollen immigration event.

Considering now the seed and pollen dispersal kernels, the estimation biases were low, in particular for the shape parameter b (underestimated by only 1.4% for b_s and 1.7% for b_p). The slight overestimate on δ_s ($\hat{\delta}_s = 150$ m instead of simulated $\delta_s = 135$ m) was consistent with the scenario detailed above: for seedlings mothered by NS-OW/NS-O adults, and fathered by a 'sampled' tree, the distance to their false mother (i.e. their father) was larger than distance to true mother, in particular because $\delta_p > \delta_s$ in simulations.

Finally, the selection gradient of tree diameter on male and female fertility was estimated with low bias. The bias on $\hat{\gamma}_m$ was negligible, while the slightly positive bias on $\hat{\gamma}_f$ (overestimated by 4.1%) was also consistent with the scenario detailed above: since $\hat{\gamma}_f$ was estimated using partially false fathers considered as mothers, and since $\gamma_m > \gamma_f$ in simulations, the variations in female fertilities due to tree size tend to be overestimated.

Discussion

Genetic structure and historical estimates of gene flow

Both seedlings and adult trees showed a marked SGS at the investigated scale, associated with important historical dispersal rates. The extent of gene dispersal varied from $\sigma_e = 417$ m in seedlings to $\sigma_e = 472$ m in adults (Table 1), as estimated with the iterative procedure described in Hardy *et al.* (2006) when assuming effective density d_e equal to observed adult density ($d_{\text{obs}} = 0.39$ trees/ha). Although only this assumption lead to convergence of the iterative procedure for both seedlings and adults, the true value of

d_e/d_{obs} could actually be lower than 1. For long-lived organisms such as trees, it is notoriously difficult to estimate d_e/d_{obs} , or equivalently N_e/N (the ratio of effective vs. census population size), as these quantities depend on the variance in lifetime reproductive success among individuals. However, approximate d_e -value can be obtained from the variance in reproductive success measured over a single or few reproductive events. Here, we used equation 2b in (Smouse & Robledo-Arnuncio 2005) to estimate the probability of paternal/parental identity (the inverse of N_e), respectively, in (i) maternal progenies after categorical paternity assignment (Oddou-Muratorio *et al.* 2005), and (ii) seedlings after categorical parentage assignment (this data set). We obtained, respectively, $N_e/N = 0.33$ and $N_e/N = 0.38$, consistently with the N_e/N range observed in natural populations (i.e. between 0.1 and 0.5; Frankham 1995). Because the estimation of σ_e depends on the square root of d_e , σ_e -values in *Sorbus torminalis* could thus be multiplied by ~ 1.7 , leading up to ~ 800 m.

Our historical estimates were consistent with results in 10 tropical scattered tree species (Hardy *et al.* 2006), where σ_e ranged from 150 to 1200 m, and was always > 500 m for species with as low density as *S. torminalis*. Large confidence intervals and frequent convergence problems were also reported in (Hardy *et al.* 2006), and are likely to result from (i) the typically very large estimation error associated with indirect methods (Leblois *et al.* 2003), (ii) not so large sample size (in particular in adults in our case), and (iii) inappropriate sample scale. Regarding this last point, the goal of the iterative procedure is to reduce estimation bias by measuring the slope of the kinship curve over a restricted distance range (between σ_e and $20\sigma_e$) for which theory provides robust expectations. However, for species showing σ_e values as high as for *S. torminalis*, this restricted distance range will correspond to a poorly informative section of the kinship curve (see Fig. 2), where the relation between pairwise kinship and distance is mostly stochastic. An appropriate sampling scheme for *S. torminalis* in Rambouillet should include individual's separated by distances up to ~ 10 km (20 times σ_e -estimate, ~ 500 m), whereas we had only sampled distances up to 2 km. Only such scale of study will provide σ_e -estimates in a realistic range of effective density (i.e. $d_e < d/2$).

Contemporary estimates of dispersal based on established seedlings

Direct estimates of seed and pollen dispersal curves using the neighbourhood model showed high dispersal abilities in this insect-pollinated, animal-dispersed species. Indeed, seed and pollen dispersal kernels estimated from established seedlings were both fat tailed ($b < 1$), with important mean dispersal distances ($\delta_p \approx 248$ m and median = 188 m; $\delta_s \approx 135$ m and median = 96 m). Additionally high levels of seed and

pollen immigration from outside the study population were detected ($m_p = 69\%$ and $m_s = 39\%$).

Considering first seed dispersal, the shape and the range of our seed dispersal kernel showed that *S. torminalis* combines preferential short-distance dispersal with non-negligible events of long-distance dispersal (LDD). This is consistent with genetic data on seed and seedlings that are beginning to accumulate in other species, and that highlight the importance of LDD in plant dispersal strategies (Jones *et al.* 2005; Hardesty *et al.* 2006; Jordano *et al.* 2007). In another scattered and animal-dispersed tree dispersed by frugivorous animals (*Prunus mahaleb*), Robledo-Arnuncio & Garcia (2007) estimated fat-tailed seed dispersal kernel ($b_s = 0.29$; $\delta_s = 277$ m), using a methodology similar to ours, but based on categorical maternity assignment. Additionally, LDD (> 1500 m) component of *P. mahaleb* seed shadow was shown to be contributed mainly by carnivorous mammals (foxes, badgers, martens) and medium-size birds (thrushes, crows), while small birds (Passerine) dispersed most seeds at shorter distances, rarely beyond 100 m (Jordano *et al.* 2007).

Considering now pollen dispersal, patterns inferred here from established seedlings (ES) can be compared with those previously inferred from maternal progenies (MP) collected on mother trees (Oddou-Muratorio *et al.* 2005). Using a neighbourhood model, we had also estimated a fat-tail pollen dispersal kernel [$b_{p-MP} = 0.33$ (0.27; 0.40)] with higher mean dispersal distance [$\delta_{p-MP} \approx 743$ m (540,1160); median = 400 m] but lower pollen immigration rate ($m_{p-MP} = 43.7\%$). We can not exclude that the higher m_p -value obtained with the ES design resulted in part from logging/natural mortality of some mother trees of sampled seedlings, although we have good evidence of the absence of logging operation from 1970 to the year of study (2001). Indeed, all mistyped seedlings will contribute to upwards bias in estimated m_{pES} , and our simulations showed that the bias can reach 29.3% for 5% of uncensored mother trees.

This study is the first to our knowledge to estimate the bias in contemporary estimate that can result from missing some candidate parents, whereas their exhaustive sampling is theoretically required in this approach. Our simulations highlighted very different behaviours for dispersal and fertility parameters within the ISP (b , δ and γ) on one hand and the seed/pollen immigration parameters (m_s and m_p) on the other hand. Indeed, the dispersal kernels were estimated with low bias (+10% for the δ -parameter and -1.5% for the b -parameter on average), which may result in slightly over-estimation of the mean dispersal rate (σ , equation 6) for pollen and seeds. The selection gradients of diameter on fertility were also only slightly overestimated (+2.3% on average), and our simulations indicate that the bias in female fertility parameters is likely to increase when the trait under consideration (here diameter) has differential quantitative effects on male and female fertilities. By contrast,

Table 3 Impact of missing mother trees on the estimation procedure; see Table 2 for legends

Parameter	Seed dispersal and female fertility				Pollen dispersal and male fertility				
	δ_s	b_s	m_s	γ_f	δ_p	b_p	m_p	s	γ_m
Mean estimated value	150.40	0.246	0.027	0.062	759.64	0.296	0.207	0.018	0.090
True value	135	0.25	0.058	0.06	700	0.3	0.160	0.02	0.09
Bias	15.40	-0.004	-0.031	0.002	59.64	-0.004	0.047	-0.002	0.0003
Percentage of bias	11.4%	-1.7%	-53.5%	4.1%	8.5%	-1.4%	29.3%	-10.4%	0.4%

the absolute bias in immigration parameter was superior to 40%, confirming that this figure should be interpreted with caution, if ever interpreted. This contrasted performance of parameters is likely to be general, as a mechanical consequence of the estimation procedure. Indeed, dispersal kernels are fitted with minimum extrapolation: for pollen, the fit is based on true dispersal distances within the known neighbourhood, and for seeds, there may be a bias because of the incorporation of pollen dispersal events in case of missing mother tree. Meanwhile, all unknown events (LDD, missing parents, genotyping error ...) are aggregated into the seed/pollen immigration parameter (m_s and m_p) instead of incorrectly affecting the shape of the LDD.

More generally, the comparison of results from simulation and real data shows that the contemporary approach may underestimate LDD. Indeed, the simulated migration rates for pollen and seeds actually result from the extrapolation of dispersal kernels to longer distances (up to ~8 km), since the simulated data sets were obtained by simulating individuals outside the ISP following a spatial point process similar to that observed inside the ISP (Table 3). We obtained average migration rates of ~6% for seeds and 16% for pollen (Table 3, True value), largely less than the values directly estimated from observed genotypes (39% for seeds in this study and 44% for pollen in Oddou-Muratorio *et al.* 2005). This discrepancy is partly due to (i) the choice of the background allelic frequencies that bias the estimation of the migration rate (Burczyk & Chybicki 2004), (ii) mistyping of microsatellite alleles leading to seeds/seedlings wrongly assigned to the external gene pool, (iii) simplified modelling of propagule paths through the use of dispersal kernels with two parameters. But these estimation biases are unlikely to fully explain the discrepancy in real and simulated migration rate, and LDD events are probably more frequent than estimated from the dispersal observed inside the ISP. This could be due to dispersal by animal vectors with different spatial range of movement, as observed in *P. mahaleb* (Jordano *et al.* 2007). Consequently, our dispersal kernels are likely to underestimate LDD, implying that mean dispersal distances for pollen and seeds, and thus root-mean-squared axial distances may be higher than reported above.

Comparison of contemporary vs. historical estimates of gene flow

In a two dimensional space, for hermaphrodite, outcrossing species: $\sigma_{rt}^2 = \frac{1}{2}\sigma_{p-rt}^2 + \sigma_{s-rt}^2$ (Crawford 1984), where σ_{p-rt}^2 and σ_{s-rt}^2 are the respective second moments of the pollen and seed dispersal kernels. We computed axial root mean squared σ_{p-rt} and σ_{s-rt} from our contemporary estimates of mean distance of seed and pollen dispersal (δ_s and δ_p , Table 2) using equation 6. This yielded a real-time, total geneflow estimate $\sigma_{rt} = 211$ m (confidence interval CI: 149–299 m). By contrast, the historical estimates ranged between 417 m (CI: 231 – +∞) and 506 m (CI: 346 – +∞) depending on the data set and on assumptions for d_e (Table 1). There was thus a 2.1- to 2.4-fold difference in the range of gene dispersal estimated by contemporary and historical estimates. As effective density may be over-estimated, the historical estimates are likely to be even larger.

It is however, difficult to be certain whether the observed 2.1- to 2.4-fold discrepancy is due to the assumptions and statistical nuances of the two approaches or whether range of gene dispersal has actually changed. First, the geographical scale ensuring the better precision of contemporary and historical estimates of gene flow (respectively σ_{rt} and σ_e) may substantially differ, in particular for species showing large range of seed and/or pollen dispersal. As detailed above, an appropriate sampling scheme for σ_e should include individuals separated by distances up to 10 km, but estimating σ_{rt} at the same scale would require exhaustive sampling of > 2300 potential parents over ~8000 ha, with thus a high risk of missing some of them. Second, our contemporary and historical estimates should be taken with caution because they both have large variances and rely on assumptions that affect the estimates. Contemporary estimates were shown to suffer low bias at the scale of the ISP, but they may underestimate LDD as detailed above. By contrast, historical estimates have minimum precision, and suffer from the general high uncertainty on the value of the effective density d_e . Here, we used census density and variance in mating success observed in year 2000 to estimate d_e . However, census and effective density may have decreased subsequently to sylvicultural change in the

1960s, with intensive logging of *S. torminalis* and reduced opportunities of establishment. Our results highlight the difficulty in documenting temporal variation in density in long-lived species, and suggest that we definitely need more studies comparing contemporary and historical geneflow estimates in species with shorter life cycle, in order to gauge to what extent this comparison highlight evolutionary forces at work within populations.

Based on our results, it can however, not be excluded that the range of gene dispersal in *S. torminalis* population has recently been restricted, partly on account of the recent sylvicultural change between the adults and seedlings generation (in the 1960s). Although extended gene flow abilities were estimated both by historical and contemporary approaches, ongoing restriction in the range of gene dispersal may render *S. torminalis* unable to counteract detrimental effects of sylvicultural practises. In particular, most of the seedlings used in this study were found in old-grown management units, before the wide opening of canopy at the end of the sylvicultural cycle, which may have severe demographic effect in light-demanding species like *S. torminalis*. This study highlights the need for forest managers to carefully monitor recruitment process in secondary species all along the sylvicultural cycle, in order to identify potential limitation in propagules' dispersal, establishment and survival of seedling.

Acknowledgements

This work is part of S.O.-M. PhD thesis, which was funded by the French National Forest Services (ONF). This study was supported by the Bureau des Ressources Génétiques. We thank Nadia Gouet and Benedicte Le Guerroué for help with laboratory work. We thank Cristina Garcia, Olivier Hardy, Pierre Gérard and three anonymous reviewers for helpful comments on successive versions of this manuscript.

References

- Abramowitz M, Stegun IA (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. U.S. Govt. Print. Off., Washington, D.C.
- Angelone S, Hilfiker K, Holderegger R, Bergamini A, Hoebee SE (2007) Regional population dynamics define the local genetic structure in *Sorbus torminalis*. *Molecular Ecology*, **16**, 1291–1301.
- Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.
- Bacles CFE, Burczyk J, Lowe AJ, Ennos RA (2005) Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior*. *Evolution*, **59**, 979–990.
- Burczyk J, Chybicki IJ (2004) Cautions on direct gene flow estimation in plant populations. *Evolution*, **58**, 956–963.
- Burczyk J, Adams WT, Birkes DS, Chybicki IJ (2006) Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics*, **173**, 363–372.
- Bush RM, Smouse PE (1992) Evidence for adaptive significance of allozymes in forest trees. *New Forest*, **6**, 179–196.
- Clark JS (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist*, **152**, 204–224.
- Crawford TJ (1984) The estimation of neighborhood parameters for plant populations. *Heredity*, **52**, 273–283.
- Dunphy B, Hamrick JL (2005) Gene flow among established Puerto Rican populations of the exotic tree species, *Albizia lebbek*. *Heredity*, **94**, 418–425.
- Dutech C, Sork VL, Irwin AJ, Smouse PE, Davis FW (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species, *Quercus lobata* (Fagaceae). *American Journal of Botany*, **92**, 252–261.
- Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.
- Gonzalez-Martinez SC, Burczyk J, Nathan R *et al.* (2006) Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology*, **15**, 4577–4588.
- Hardesty BD, Hubbell SP, Bermingham E (2006) Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. *Ecology Letters*, **9**, 516–525.
- Hardy O, Vekemans X (2002) SPAGEdi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618.
- Hardy OJ, Maggia L, Bandou E *et al.* (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, **15**, 559–571.
- Jones FA, Chen J, Weng G-J, Hubbell SP (2005) A genetic evaluation of seed dispersal in the Neotropical tree *Jacaranda copaia* (Bignoniaceae). *American Naturalist*, **166**, 543–555.
- Jordano P, Garcia C, Godoy JA, Garcia-Castano JL (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences, USA*, **104**, 3278–3282.
- Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trilium grandiflorum*: the roles of dispersal, mating, history and selection. *Evolution*, **55**, 1560–1568.
- Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a 'continuous' population under isolation by distance. *Molecular Biology and Evolution*, **20**, 491–502.
- Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics*, **166**, 1081–1092.
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.
- Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136–142.
- Meagher TR (1986) Analysis of paternity within a natural population of *Chamaelirium luteum*. I. Identification of most-likely male parents. *American Naturalist*, **128**, 199–215.
- Morgan MT, Conner JK (2001) Using genetic markers to directly estimate male selection gradients. *Evolution*, **55**, 272–281.
- Oddou-Muratorio S, Petit RJ, Le Guerroué B, Guesnet D, Demesure B (2001) Pollen- versus seed-mediated gene flow in a scattered woody species. *Evolution*, **55**, 1123–1135.

- Oddou-Muratorio S, Demesure-Musch B, Pélissier R, Gouyon P-H (2004) The roles of colonization and history in the local genetic structure of a scattered temperate tree species. *Molecular Ecology*, **13**, 3689–3702.
- Oddou-Muratorio S, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology*, **14**, 4441–4452.
- Oddou-Muratorio S, Klein EK, Austerlitz F (2006) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. III. Mating patterns and the ecological maternal neighborhood. *American Journal of Botany*, **93**, 1650–1659.
- Otero-Arnaiz A, Casas A, Hamrick JL (2005) Direct and indirect estimates of gene flow among wild and managed populations of *Polaskia chichipe*, an endemic columnar cactus in Central Mexico. *Molecular Ecology*, **14**, 4313–4322.
- Robledo-Arnuncio JJ, Garcia C (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology*, **16**, 5098–5109.
- 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 (2004) *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, New Jersey.
- Schweizer M, Excoffier L, Heckel G (2007) Fine-scale genetic structure and dispersal in the common vole (*Microtus arvalis*). *Molecular Ecology*, **16**, 2463–2473.
- Smouse PE, Robledo-Arnuncio JJ (2005) Measuring the genetic structure of the pollen pool as the probability of paternal identity. *Heredity*, **94**, 640–649.
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.
- Wohlgemuth T, Burgi M, Scheidegger C, Schutz M (2002) Dominance reduction of species through disturbance — a proposed management principle for central European forests. *Forest Ecology and Management*, **166**, 1–15.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Sylvie Oddou-Muratorio is interested in assessing the spatial and temporal components of reproductive success and effective size in expanding tree populations. Etienne Klein is interested in modelling and estimating long-distance dispersal through pollen and seed.
