Molecular Ecology (2004) **13**, 921–935 doi: 10.1046/j.1365-294X.2004.02076.x

# New insights from fine-scale spatial genetic structure analyses in plant populations

X. VEKEMANS* and O. J. HARDY†
*Université de Lille 1, Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016, Bat SN2, F- 59655 Villeneuve d'Ascq cedex, France, †Université Libre de Bruxelles, Laboratoire de Génétique et Ecologie Végétales, Chaussée de Wavre 1850, B-1160 Bruxelles, Belgium

**Abstract**

**Many empirical studies have assessed fine-scale spatial genetic structure (SGS), i.e. the nonrandom spatial distribution of genotypes, within plant populations using genetic markers and spatial autocorrelation techniques. These studies mostly provided qualitative descriptions of SGS, rendering quantitative comparisons among studies difficult. The theory of isolation by distance can predict the pattern of SGS under limited gene dispersal, suggesting new approaches, based on the relationship between pairwise relatedness coefficients and the spatial distance between individuals, to quantify SGS and infer gene dispersal parameters. Here we review the theory underlying such methods and discuss issues about their application to plant populations, such as the choice of the relatedness statistics, the sampling scheme to adopt, the procedure to test SGS, and the interpretation of spatial autocorrelograms. We propose to quantify SGS by an '*Sp*' statistic primarily dependent upon the rate of decrease of pairwise kinship coefficients between individuals with the logarithm of the distance in two dimensions. Under certain conditions, this statistic estimates the reciprocal of the neighbourhood size. Reanalysing data from, mostly, published studies, the *Sp* statistic was assessed for 47 plant species. It was found to be significantly related to the mating system (higher in selfing species) and to the life form (higher in herbs than trees), as well as to the population density (higher under low density). We discuss the necessity for comparing SGS with direct estimates of gene dispersal distances, and show how the approach presented can be extended to assess (i) the level of biparental inbreeding, and (ii) the kurtosis of the gene dispersal distribution.**

*Keywords*: dispersal, isolation by distance, plant breeding systems, spatial autocorrelation, spatial genetic structure

*Received 20 July 2003; revision received 31 October 2003; accepted 31 October 2003*

## Introduction

Spatial genetic structure (SGS) in natural populations, i.e. the nonrandom spatial distribution of genotypes, can result from different processes, including selection pressures or historical events. At a fine spatial scale, however, the most prevalent cause is probably the formation of local pedigree structures as a result of limited gene dispersal. In this context, genetic similarity is higher among neighbouring than among more distant individuals, and the theory of isolation by distance predicts the expected pattern of SGS at drift–dispersal equilibrium.

As compared to most animal species, adults from plant species do not move and plants' propagules, i.e. pollen and seeds, often show moderate to strong spatial restriction in their dispersal. Hence, SGS is expected to occur frequently within plant populations. Many empirical studies have investigated fine-scale SGS within plant populations, often using spatial autocorrelation methods (reviewed in Heywood 1991; Epperson 1993). Most of these studies describe patterns in a qualitative way, limiting the inferences that could be made, and making quantitative comparisons among studies difficult. Moreover, the patterns revealed through spatial autocorrelation often show substantial

Correspondence: X. Vekemans. Fax: 33 3 20 43 69 79; E-mail: xavier.vekemans@University-lille1.fr

stochasticity because of the process of genetic drift and the limited information content available from genetic markers (Slatkin & Arter 1991), especially for the earliest studies based on allozymes. Since then, very polymorphic markers (e.g. microsatellites) or markers providing many loci (e.g. amplified fragment length polymorphisms) have become available, reducing the stochasticity problem. Moreover, new theoretical and methodological advances now permit a more quantitative assessment of SGS, allowing new inferences on isolation-by-distance processes (e.g. Rousset 1997, 2000; Hardy & Vekemans 1999; Hardy 2003).

In this paper we focus on a method based on the relationship between pairwise genetic and spatial distances in continuous populations, more specifically in plants. The paper has three objectives. (i) To provide concise accounts of the theory underlying the approach, the methodological procedure, the choice of the statistics, and the interpretation of the results. (ii) To apply the method to 47 plant species, from a survey of published and unpublished data, and to relate patterns of SGS to life-history traits and ecological conditions such as population density. (iii) To present new methodological perspectives arising from this method that allow new insights into plant population biology. Throughout, we will assume that genetic markers are neutral and the impact of selection will not be addressed.

## Methodological approaches to assess SGS

### *Theoretical background on isolation by distance within continuous populations*

Under isolation by distance (i.e. limited gene dispersal), the probability of identity in state between two neutral genes ($Q$) decreases with the spatial distance separating them ($r$), and this phenomenon can be used to characterize SGS. The analytical modelling of this process was undertaken by Malécot (1950). At drift–dispersal–mutation equilibrium under isotropic dispersal, the function $Q(r)$ depends on the gene dispersal function, the population effective density ($D$), the geometry of the population, and the mutation rate ($\mu$) and process.

In this paper, we focus on SGS within 'continuous populations', meaning areas where individuals are distributed homogeneously (constant density). Most of the theory we will refer to, however, is based on a 'lattice' model (i.e. a regular grid with one individual per node) but its results seem robust to the spatial clustering often displayed in natural populations (e.g. Barton *et al*. 2002). A convenient analytical result for isotropic dispersal is that, for $r$ ranging between $\sigma$ and $0.5\,\sigma/(2\,\mu)^{1/2}$, where $\sigma^2$ is the average squared axial parent–offspring distance (or half this value in a two-dimensional space), $Q(r)$ decreases approximately linearly with $r$ in a one-dimensional space, and $\ln(r)$ in a two-dimensional space, at a rate proportional to $1/D\sigma^2$ (Rousset 1997, 2000; Barton *et al*. 2002). When $r < \sigma$, the shape of $Q(r)$ depends on the details of the dispersal function (e.g. leptokurtosis), and when $r > 0.5\,\sigma/(2\,\mu)^{1/2}$, it depends on the mutation rate. Hence, when looking at an adequate spatial scale, the product $D\sigma^2$ can potentially be inferred from SGS whatever the exact form of the dispersal kernel (as long as $\sigma^2$ is finite, Rousset 2001).

However, a major requirement for successful $D\sigma^2$ inference is that SGS has reached a stationary phase representative of the drift–dispersal equilibrium state with current demographic parameters. This requires a few generations on a spatial scale one order of magnitude larger than $\sigma$, but tens or hundreds of generations on a scale two or three orders of magnitude larger than $\sigma$ (e.g. Hardy & Vekemans 1999). Thus, $D\sigma^2$ estimates based on SGS are more likely to be reliable if based on a scale not larger than say 10–50 times $\sigma$. At such a spatial scale, mutations can be neglected and the model of mutation process (e.g. stepwise or not) is not relevant, unless the mutation rate is very high (e.g. $\mu > 10^{-3}$; Leblois *et al*. 2003).

The function $Q(r)$ depends strongly on the mutation rate and is thus locus specific [only the shape of $Q(r)$ is mutation independent at short distance]. On the contrary, ratios of the form $[Q(r) - Q_R]/(1 - Q_R)$, where $Q_R$ is the probability of identity for a particular class of gene pairs (a reference), are independent of $\mu$ under the low mutation limit, i.e. $\mu \to 0$ (Rousset 2002). SGS is thus best characterized by such ratios which express the mean genetic similarity among individuals separated by a distance $r$ relative to the genetic similarity between genes of a reference class of gene pairs. Two kinds of reference ($Q_R$) have been proposed: genes within individuals ($Q_0$) and random genes from a sample of individuals ($\bar{Q}$).

In the first case, Rousset (2000) devised the parameter $a_r \equiv [Q_0 - Q(r)]/(1 - Q_0)$. Note that the terms in the numerator are reversed so that $a_r$ expresses a distance rather than a similarity measure. For diploids, assuming that male and female gametes disperse independently, the regression slope, $b_a$, of $a_r$ on $r$ in one-dimensional space is $b_a = (4D\sigma^2)^{-1}$, and of $a_r$ on $\ln(r)$ in two-dimensional space is $b_a = (4\pi D\sigma^2)^{-1}$ (Rousset 2000). In the second case, one can use pairwise 'kinship' coefficients between individuals separated by distance $r$, $F(r) \equiv [Q(r) - \bar{Q}]/(1 - \bar{Q})$. As $a_r = [F_I - F(r)]/(1 - F_I)$ where $F_I \equiv (Q_0 - \bar{Q}]/(1 - \bar{Q})$ is Wright's inbreeding coefficient expressing the deficit of heterozygotes relative to Hardy–Weinberg proportions, the regression slope of $F(r)$ on $r$ in one-dimensional space is $b_F = -(1 - F_I)/4D\sigma^2$, and of $F(r)$ on $\ln(r)$ in two-dimensional space is $b_F = -(1 - F_I)/4\pi D\sigma^2$. Interestingly, in two dimensions the quantity $4\pi D\sigma^2$ matches Wright's 'neighbourhood size', $Nb$ (Wright 1943, 1946), assuming Gaussian dispersal functions. Although the common interpretation of $Nb$ as the size of a 'panmictic unit' is misleading and the biological significance of $Nb$ should not be overestimated (e.g. Rousset

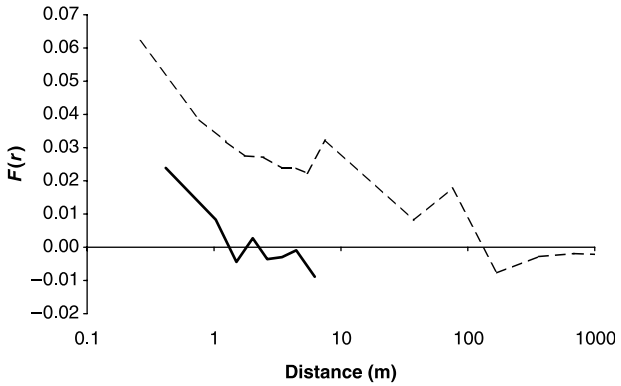


**Fig. 1** Impact of the sample scheme on plots of $F(r)$ against physical distance between individuals, $r$. Average kinship coefficients between pairs of individuals in a given class of distances, $F(r)$, were computed from samples taken on a short spatial scale (plain line) or on a wide spatial scale (broken line), using data from Fenster *et al*. (2003) in *Chamaecrista fasciculata*. It can be seen that the overall rates of decrease in $F(r)$ with distance are similar, but the actual values of $F(r)$ and of the $x$-axis intercept do differ substantially.

1997; Fenster *et al*. 2003), $Nb$ remains a convenient synthetic way of expressing the balance between local genetic drift and gene dispersal within continuous populations, predicting SGS in the range of $\sigma$ to one to three orders of magnitude $\sigma$.

The advantage of $a_r$ for characterizing SGS is that it does not depend on a particular sample, unlike $F(r)$ which involves the $\bar{Q}$ term, so that the sampling scheme influences expected $F(r)$ values (Fig. 1). However, as shown below, estimators of $a_r$ typically suffer higher sampling variance than estimators of $F(r)$. A second difficulty with $a_r$, in particular with plants, is that its relationship with $D\sigma^2$ given above does not hold when substantial selfing occurs (with highly selfed individuals, the homozygosity $Q_0 \to 1$, so that $a_r \to \infty$). This problem can be overcome with $F(r)$ because, even when selfing occurs, the regression slope of $F(r)$ on $r$ or $\ln(r)$ is still approximately equal to $-(1 - F_N)/4D\sigma^2$ in one-dimensional space, or $-(1 - F_N)/4\pi D\sigma^2$ in two-dimensional space, where $F_N$ represents the kinship coefficient between neighbouring individuals (note that $F_I$ was replaced by $F_N$). Strictly speaking, in a lattice model, $F_N$ should be the kinship between individuals competing for the same location before one is selected at random (F. Rousset, personal communication) but, in practice, the kinship between neighbours is the best approximation we can get for this quantity. We can likewise imagine an alternative definition for $a_r$ solving the problem of selfing: $a'_r \equiv [Q_N - Q(r)]/(1 - Q_N)$, where $Q_N$ is the probability of identity in state between the genes of neighbours. Here, it is the genetic similarity between neighbours that is used as reference. In the following, we will concentrate on $F(r)$ because $a'_r$ can be derived directly from $F(r)$ values, as $a'_r = [F_N - F(r)]/(1 - F_N)$, and $a'_r$ has never been used to our knowledge in the empirical literature in plants. Nevertheless, we should note that $a'_r$ would be an interesting way to compare graphically SGS among studies that differ in their sampling scheme.

### *Outline of the procedures to assess SGS within continuous populations*

Studies devoted to analysing SGS within populations of plants generally have one or several of the following objectives: (i) to describe the pattern of SGS; (ii) to test the pattern of SGS against that of a random spatial distribution of genotypes; (iii) to quantify the amount of SGS; and (iv) to infer values of biological parameters (e.g. dispersal distances) from the pattern of SGS. These objectives are generally achieved by computing pairwise relatedness coefficients between individuals in the sample and analysing their relationship with the spatial distance separating individuals.

### *Computing pairwise relatedness coefficients between individuals*

There are several types of relatedness coefficients between individuals, and several estimators have been proposed for each of them. We can distinguish two types of 'two-genes' (i.e. based on the probability of identity of single pairs of genes) relatedness coefficient between two individuals $i$ and $j$. First, the 'kinship' or 'coancestry' coefficient, $F_{ij}$, based on the probability that a random gene from $i$ is identical to a random gene from $j$, and defined as $F_{ij} \equiv (Q_{ij} - \bar{Q})/(1 - \bar{Q})$. Second, the 'relationship' coefficients, $R_{ij}$, based on the probability that a random gene from $i$ is identical to one of the genes from $j$. The two coefficients are closely related and provide the same information: in the case of diploids $R_{ij} = 2F_{ij}/(1 + F_I)$ (Hardy & Vekemans 1999; see also Hardy 2003 for a more detailed account on terminology and definitions).

For codominant genetic markers, estimators of $F_{ij}$ are given in Loiselle *et al*. (1995) and Ritland (1996), and estimators of $R_{ij}$ are given in Queller & Goodnight (1989), Li *et al*. (1993), Hardy & Vekemans (1999), Lynch & Ritland (1999), and Wang (2002) (see Van de Casteele *et al*. 2001 for variants of the estimators of Queller & Goodnight 1989 and Li *et al*. 1993). For dominant markers, Lynch & Milligan (1994) proposed an estimator of $R_{ij}$ and Hardy (2003) proposed estimators of $F_{ij}$ and $R_{ij}$.

These estimators differ in their bias (which is generally small) and their sampling variance (which is always huge), and the statistical properties of some of them have been compared (e.g. Lynch & Ritland 1999; Van de Casteele *et al*. 2001; Wang 2002; Hardy 2003). It must be noted that many of these estimators (Li *et al*. 1993; Lynch & Milligan 1994;

Lynch & Ritland 1999; Wang 2002) assume Hardy–Weinberg genotypic proportions and are thus *a priori* inadequate to study SGS where heterozygote deficits are likely to occur, especially in plant species subject to selfing. The following estimators can be used irrespective of the mating system of the species because the logic behind their construction makes no assumption regarding Wright's inbreeding coefficient: for codominant markers, statistics defined in Loiselle *et al*. (1995), Ritland (1996), Queller & Goodnight (1989), and Hardy & Vekemans (1999); for dominant markers, statistics from Hardy (2003).

### *Regressing relatedness coefficients on distance*

To visualize and describe SGS assuming isotropic dispersal processes, mean $F_{ij}$ (or $R_{ij}$) estimates over pairs of individuals in a given distance interval $r$, $\hat{F}(r)$, can be plotted against distance as in a spatial autocorrelogram (note that Moran's $I$ statistic is an estimator of $R(r)$; Hardy & Vekemans 1999). If $\hat{F}(r)$ tends to decrease linearly with $r$ or $\ln(r)$, the extent of SGS can be quantified by the regression slope ($\hat{b}_F$) of $\hat{F}_{ij}$ on $r_{ij}$ or $\ln(r_{ij})$. However, $\hat{b}_F$ depends somewhat on the sampling scheme used (see above) and is negative, so that a better way to quantify SGS is by the ratio $-\hat{b}_F/(1-\hat{F}_{(1)})$, where $\hat{F}_{(1)}$ is the mean $\hat{F}_{ij}$ between individuals belonging to a first distance interval that should include all pairs of neighbours (i.e. $\hat{F}_{(1)}$ estimates $F_N$). This ratio can be used to compare the extent of SGS among populations or species, and we shall refer to it as the '*Sp*' statistic. It is better than using for example $\hat{F}_{(1)}$, which strongly depends on the sampling scheme (Fig. 1; Fenster *et al*. 2003). Nevertheless, if $\hat{F}(r)$ does not decrease linearly with $r$ or $\ln(r)$, the *Sp* statistic will depend on the distance range implicit in the sample.

Depending on the sampling scheme, there may be a lack of pairs of neighbours in real data sets to properly estimate $F_N$. However, as $F_N$ is typically much closer to zero than to unity and it is $1-F_N$ that matters, a small error in $F_N$ estimation is negligible. Hence, using a first distance class containing enough pairs of individuals to get a reasonable precise $\hat{F}_{(1)}$ value to estimate $F_N$ should be adequate in practice, even if these pairs are not strictly speaking neighbours.

When SGS is truly representative of an isolation-by-distance pattern at dispersal–drift equilibrium, the dispersal parameters can be estimated from the *Sp* statistic, which is expected to be equal to $1/4D\sigma^2$ in one-dimensional space, or $1/4\pi D\sigma^2$ in two-dimensional space (i.e. $1/Nb$), provided that the regression is restricted to an appropriate distance range ($\sigma$ to 10–50 $\sigma$ in two-dimensional space, see above). As $\sigma$ is unknown, an iterative approach can be applied, assuming that the effective density, $D$, is known (Heuertz *et al*. 2003). Therefore, a first estimate of $D\sigma^2$ is based on a global regression (i.e. over the full distance range available), then $\hat{\sigma}$ is extracted, and a new estimate of $D\sigma^2$ is based on a restricted regression considering only pairs separated by a distance between $\hat{\sigma}$ and, say, $20\hat{\sigma}$. The process is repeated until successive $\hat{\sigma}$ estimates stabilize (procedure implemented in the software SPAGEDI; Hardy and Vekemans 2002; http://*http://www.ulb.ac.be/sciences/lagev*). It may well happen that the procedure does not converge, which suggests either that the spatial scale of the sample was not adequate (e.g. Fenster *et al*. 2003), that SGS is not representative of an isolation-by-distance process, and/or that the information from the genetic markers is insufficient to get reliable estimates.

The assumed value for $D$ in this procedure is critical because it is an effective density, depending on the variance in reproductive success among individuals and through time. As a first approximation, $D$ is the product of the census density and the $N_e/N$ ratio (effective over census population sizes), and computer simulations suggest that this approximation is accurate at least when the density is fairly constant in space (O. J. Hardy, unpublished results). If an estimate of the variance in lifetime reproductive success among individuals, $V$, is available, one can approximate $N_e/N \cong 4/[2(1-F_I)+(1+F_I)V)]$ assuming stable population density over time (Kimura & Crow 1963). In natural plant populations, $N_e/N$ typically ranges between 0.5 and 0.1 (Husband & Barrett 1992; Frankham 1995) so that, in the absence of data on $V$, one might set $D$ as one-half to one-tenth the density of adults.

Although the *Sp* statistic may be used both to quantify SGS and to estimate $D\sigma^2$, the two purposes should be distinguished. The pattern of SGS is synthetically quantified by the *Sp* statistic, whatever its cause [it is not ideal when $\hat{F}(r)$ strongly departs from linearity with respect to $r$ or $\ln(r)$, but such departures are often difficult to assess]. However, estimating $D\sigma^2$ from *Sp* implicitly assumes that: SGS results solely from isotropic limited gene dispersal, SGS has reached a stationary phase, the sampling scale is adequate with respect to $\sigma$, and the geometry of the population is also adequate (one- or two-dimensional across the $\sigma - 20\sigma$ range). The distribution of individuals does not need to be perfectly homogeneous (constant density), but an aggregated distribution affects $D$ (Barton *et al*. 2002).

### *Interpreting correlograms*

The interpretation of autocorrelograms has not always been performed properly. A common misinterpretation is to say that SGS extends to a distance $x$, where $x$ is the distance at which $\hat{F}(r)$ (or Moran's $I$) reaches zero (or an expected value in the absence of SGS, which is somewhat negative for Moran's $I$). This distance is, however, not a characteristic of the populations studied, as it depends strongly on the sampling scheme (it is the distance at which individuals are, on average, as similar as two random individuals from the sample). This is well illustrated in

Fenster *et al*. (2003) where analyses over samples taken at different spatial scales within the same population were compared, $x$ increasing substantially when larger scales were analysed (Fig. 1). Hence, $\hat{F}(r)$ values are arbitrary, and only the way they change with distance is relevant. Paying attention to the distance above which $\hat{F}(r)$ estimates are no longer significantly positive according to a randomization test is not a better option. Indeed, the results still depend on the sampling scheme and, moreover, depend on the arbitrarily set distance intervals (the testing power increases with the number of pairwise comparisons falling in the interval) and on the information content of genetic markers. There is one instance where a critical distance can be defined: if $\hat{F}(r)$ decreases steadily until some distance $x$, showing no further trend, one may say that SGS occurs until $x$. But note that to be sure that $F(r)$ remains stable above $x$, precise estimates are required (i.e. large sample size and/or very informative genetic markers).

Quantifying SGS by the $\hat{F}(r)$ value at short distances ($\hat{F}_{(1)}$), as has been suggested to estimate dispersal parameters (e.g. Heywood 1991; Epperson & Li 1996), is also not ideal because it is dependent on the sampling scheme and the setting of the first distance interval. The *Sp* statistic, $-\hat{b}_F / (1 - \hat{F}_{(1)})$, however, is robust to the sampling scheme, at least as long as $F(r)$ is approximately linear with $r$ or $\ln(r)$. The setting of the first distance interval can affect $\hat{F}_{(1)}$, but as $\hat{F}_{(1)}$ is usually much closer to 0 than to 1 $(1 - \hat{F}_{(1)})$ remains robust.

### *Testing correlograms*

A commonly used procedure consists in determining empirically the frequency distribution of the statistics computed under the null hypothesis of no SGS, by randomly permuting genotypes among locations (Heywood 1991). This is generally applied to average values per distance intervals [e.g. $\hat{F}(r)$ or Moran's *I*-values], which produces as many tests as distance intervals, and a Bonferroni correction is sometimes applied. Applying this correction, the overall test becomes very conservative because $\hat{F}(r)$ values are not independent, but neglecting it, it is liberal if a single significant deviation at a distance interval is taken as demonstrating the occurrence of SGS. Based on these tests it is often concluded that SGS occurs in the range where significant deviations are detected and does not occur elsewhere. However, the randomization procedure cannot be used to determine the 'scales' of SGS. It allows one to test only one null hypothesis: the overall absence of SGS. Therefore, it should be applied to test a single statistic describing SGS, such as the regression slope of $\hat{F}_{ij}$ on $r_{ij}$, $\hat{b}_F$ (this is strictly equivalent to a Mantel test). The power will be maximal when a linear relationship occurs, and any monotonic transformation (i.e. preserving the ranking) of $r_{ij}$ [such as $\ln(r_{ij})$] increasing this linearity could be applied. The advantages of testing $\hat{b}_F$ rather than a set of $\hat{F}(r)$ values is that (i) all the information is used in a single test, (ii) the results are independent of arbitrarily set distance intervals, (iii) the testing power is higher, at least as long as $\hat{F}(r)$ approaches linearity with (transformed) $r$ (Rousset 2000; O. J. Hardy, unpublished). In some case, strange forms of SGS may not result in significant slopes because $\hat{F}(r)$ goes up and down with distance, but with an amplitude exceeding the sampling errors. A Mantel test is inefficient in such cases; inspecting the tests applied to different distance intervals can reveal such a pattern (see Smouse & Peakall 1999 for a global test appropriate for such situations).

### *Choice of the pairwise estimator of genetic relatedness*

It must be noted that, when *testing* SGS, a low sampling variance is required but estimator bias is unimportant. On the contrary, bias matters as much as sampling variance when *quantifying* SGS or estimating $D\sigma^2$. Table 1 shows the relative performance of different relatedness estimators, and an estimator of the $a_r$ parameter (Rousset 2000), to test for SGS using real data sets of 10 populations of different species genotyped with allozyme or microsatellite markers. SGS was tested by assessing the significance of the regression slope of pairwise statistics on ln(distance) using 10 000 randomizations of the individual spatial positions using the software SPAGEDI (Hardy & Vekemans 2002). In most cases, Ritland's estimator (equation 5 in Ritland 1996) proved the most powerful, especially with highly polymorphic markers, and the statistic presented in Loiselle *et al*. (1995), which shares the same statistical properties as Hardy & Vekemans's (1999) estimator derived from Moran's *I* statistic, was the best in a few cases where marker polymorphism was lower (allozymes). The $a_r$ estimator of Rousset (2000) performed poorly, detecting significant SGS at a 5% level in only three populations, whereas Ritland's (1996) and Loiselle *et al*. (1995) estimators detected significant SGS in eight populations. Although Ritland's (1996) estimator seems the best to test SGS, it tends to give downward biased estimates when rare alleles occur (Ritland 1996; O. J. Hardy, unpublished results), so that it may not be the most adequate to quantify SGS and estimate $D\sigma^2$. For these purposes, simulations showed that the statistic presented in Loiselle *et al*. (1995) performs well, even for predominantly selfing species with high $F_I$ (Hardy 2003; O. J. Hardy, unpublished).

### *Sampling strategy*

To characterize SGS best, the range of spatial scales encompassed by the sample should be maximized. This is especially relevant for $D\sigma^2$ inference because it requires a particular window of spatial scales, which is difficult to assess without prior information on $\sigma$. Sampling exhaustively within a confined area ensures a detailed account of

**Table 1** Power of different pairwise statistics to test SGS when applied to various data sets

| Species* | Marker† | Sample size | No. of loci‡ | Total no. of alleles (range per locus)‡ | Mean $H_E$‡¶ | Estimators** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $F_L$ | $F_R$ | $R_{Q\&G}$ | $R_{L\&R}$ | $R_W$ | $R_L$ | $a_r$ |
| *Dicorynia guianensis* | A | 135 | 4 | 12 (2–4) | 0.45 | 0.0048 | 0.0344 | 0.3888 | 0.4454 | 0.1945 | 0.2049 | 0.1710 |
| *Chrysophyllum sanguinolentum* | A | 54 | 4 | 13 (2–4) | 0.52 | 0.0614 | 0.0752 | 0.3583 | 0.0798 | 0.5831 | 0.5809 | 0.7601 |
| *Primula vulgaris* | A | 50 | 7 | 16 (2–3) | 0.24 | 0.0339 | 0.0008 | 0.0047 | 0.0040 | 0.0070 | 0.0113 | 0.0047 |
| *Chamaecrista fasciculata* | A | 88 | 6 | 18 (2–4) | 0.38 | 0.0360 | 0.0629 | 0.1469 | 0.1858 | 0.1009 | 0.1588 | 0.0769 |
| *Arabidopsis lyrata* | A | 59 | 8 | 22 (2–5) | 0.29 | 0.0090 | 0.0131 | 0.0079 | 0.0173 | 0.0303 | 0.041 | 0.0216 |
| *Medicago truncatula* | M | 132 | 5 | 23 (2–6) | 0.35 | 0.0002 | 0.0000 | 0.1361 | 0.1912 | 0.1235 | 0.1103 | 0.1681 |
| *Eryngium alpinum* | M | 100 | 7 | 40 (2–9) | 0.47 | 0.0079 | 0.0000 | 0.0312 | 0.0005 | 0.0056 | 0.0085 | 0.0227 |
| *Centaurea corymbosa* | M | 50 | 9 | 41 (2–11) | 0.5 | 0.0092 | 0.0007 | 0.1692 | 0.0013 | 0.2458 | 0.3030 | 0.4726 |
| *Sextonia rubra* | M | 163 | 4 | 48 (9–15) | 0.83 | 0.0012 | 0.0000 | 0.0512 | 0.0022 | 0.1109 | 0.1222 | 0.0684 |
| *Quercus robur* | M | 100 | 6 | 104 (10–27) | 0.871 | 0.2724 | 0.0286 | 0.0507 | 0.0604 | 0.0409 | 0.0347 | 0.0793 |

The regression slopes of the pairwise values between individuals on the logarithm of the spatial distance were compared with the distribution for 10 000 randomizations of the genotypes (*cf.* Mantel test). The *P*-values that there is no SGS are reported. Doubled and single underlines indicate the best and second best estimators, respectively.

*See Table 2 for the references to each species.

†Markers: A, allozymes; M, microsatellites.

‡Only polymorphic loci are considered.

§The original data sets were sub-sampled to reduce the testing power and avoid that $P < 10^{-4}$ for all estimators.

¶$H_E$, gene diversity.

**Estimators: $F_L$ (Loiselle *et al.* 1995); $F_R$ (equation 5 in Ritland 1996); $R_{Q\&G}$ (Queller & Goodnight 1989); $R_{L\&R}$ (Lynch & Ritland 1999); $R_W$ (Wang 2002); $R_L$ (Li *et al.* 1993); $a_r$ (Rousset 2000). The correspondence with the nomenclature used by Van de Casteele *et al.* (2001) is the following: $F_R = \hat{r}_{C0}/2$; $R_{Q\&G} = \hat{r}_{Qrsl,avi}$; $R_{L\&R} = \hat{r}_R$; $R_L = \hat{r}_{Swavl}$ (see Van de Casteel *et al.* 2001 for computational details of $R_{Q\&G}$ and $R_L$).

 

very fine-scale SGS but misses larger scales. On the other hand, sampling one individual at each node of a regular grid over a larger scale will miss fine-scale SGS. Hence, we do not recommend these strategies in general (but the first one may be useful if the data are also used for parentage analysis). Sampling along transects can be a useful alternative, because small to large scales are well represented (given enough pairs of individuals) and mapping individuals is easy. With a single transect in two dimensions, the largest distances will correspond to pairs between the extremities of the transect, which may result in fairly stochastic $F(r)$ estimates at long distances. Hence, two or more transects disposed along, e.g. a cross, a triangle or a square, should be better. However, it is important to verify that the mesh of the transect is fine enough that the consequences of very restricted dispersal events will be detected. Stratified sampling with nearby individuals sampled at different locations, and then organized into subgroups, can be another alternative, though mapping can be more difficult. This strategy can ensure a more homogeneous coverage of a study site than transects, but some intermediate distance intervals may contain a few pairs of individuals. [See Epperson & Li (1996) and Leblois *et al*. (2003) for other discussions on sampling strategies.]

Better assessment of SGS can be obtained by increasing the sample size and/or scoring more markers. How should we distribute the efforts? A good criterion is the sampling variance of the *Sp* statistic, which depends essentially on the sampling variance of $\hat{b}_F$ ($V_b$). It can be shown that, in the absence of SGS, $V_b$ is proportional to the sampling variance of $\hat{F}_{ij}$ ($V_F$), as well as to $1/n^2$, where $n$ is the sample size (O. J. Hardy, unpublished). With the estimator of Ritland (1996), $V_F$ is roughly inversely proportional to the total number of alleles minus the number of loci (lets call this difference $A$). Thus, when $n$ or $A$ is doubled, $V_b$ is divided by four or two, respectively, suggesting that, at equal cost, one should put more effort into sampling individuals than into scoring markers. However, simulations show that under isolation by distance (i.e. significant SGS), $V_b$ remains proportional to $1/A$ but is less than proportional to $1/n^2$. Thus, depending on the strength of isolation by distance, doubling $n$ could reduce $V_b$ by less than a factor two (O. J. Hardy, unpublished). Therefore, there is no general rule; except that sampling more individuals should be much more efficient than scoring more markers in populations where SGS is weak (say $N_b > 100$).

## A reanalysis of spatial autocorrelation studies in plants

### *Introduction*

As stated before, many empirical studies in plants have been devoted to characterizing SGS at the scale of single populations, often using the approach of spatial autocorrelation (reviewed in Heywood 1991; Ennos 2001). Two main points emerge from these studies: (i) a statistically significant pattern of SGS is observed in the vast majority of studies; and (ii) the pattern of SGS is often detected only at the shortest spatial scales investigated. These two observations have usually been interpreted as a consequence of an isolation-by-distance process with restricted seed dispersal within plant populations (e.g. Streiff *et al*. 1998; Gehring & Delph 1999). However, the patterns of SGS could not be directly compared among species because of (i) a lack of a synthetic statistic independent of the spatial scale and insensitive to the particular sampling scheme used; and (ii) the variation among studies in the details of the spatial autocorrelation method applied (e.g. different relatedness statistics computed, estimates computed for each allele or each locus separately vs. multiallelic multilocus estimates). The theoretical framework described in the first part of this paper provides such comparative tools. Firstly, it is possible to transform values of different relatedness statistics into a common statistic, for instance to transform values of the often used Moran's $I$ statistic applied to individual allele frequencies to a kinship coefficient corresponding to the statistic of Loiselle *et al*. (1995). Secondly, the *Sp* statistic, $-\hat{b}_F/(1-\hat{F}_{(1)})$, allows comparisons among species and can be computed based on published data.

We compiled published and unpublished data from 47 plant species to compare their patterns of SGS and tested whether these patterns were related to life-history traits of the species, such as plant mating system, life form, or pollen and seed dispersal mechanisms, as well as to population density. The approach is analogous, although much more limited in the number of species compared, to the survey of allozyme variation within and among plant populations by Hamrick & Godt (1990).

### *Materials and methods*

We considered studies of SGS in plants using codominant nuclear markers that provided either of the following results: (i) values of $\hat{b}_F$ and $\hat{F}_{(1)}$, respectively, the regression slope of $\hat{F}_{ij}$ on $\ln(r)$ and the kinship coefficient between adjacent individuals (for species 1, 6, 8, 19, 28, 37, 46; see references in Table 2); (ii) tabulated values of an average genetic relatedness statistic between pairs of individuals separated by given distance intervals, i.e. the data that are usually plotted on spatial autocorrelograms (for species 14, 29, 32, 40); or (iii) graphical representation of spatial autocorrelograms for individual alleles, individual loci, or multiallelic multilocus statistics (for species 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 15, 17, 20, 21, 22, 24, 27, 31, 33, 34, 36, 38, 43, 45); and (iv) unpublished studies for which full datasets were kindly provided by the authors and which were

**Table 2** Biological characteristics and statistics of SGS in plant species for which SGS data has been reanalysed. The species are listed by decreasing order of the *Sp* statistic

| | Species | Reference | Marker* | npop† | Breeding system‡ | Life form§ | Pollen dispersal¶ | Seed dispersal** | $F_1$†† | $F_{(1)}$‡‡ | *Sp* statistic |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Phaseolus lunatus* | Zoro Bi *et al.* (1997) | A | 1 | S | H | AP | GS | 0.717 | 0.206 | 0.26316 |
| 2 | *Cryptotaenia canadensis* | Williams (1994) | A | 2 | S | H | AP | GS | 0.609 | 0.255 | 0.17479 |
| 3 | *Plantago major* | Van Dijk (1987) | A | 1 | S | H | WP | GS | 0.863 | 0.148 | 0.12310 |
| 4 | *Androcymbium gramineum* | Caujapé-Castells & Pedrola-Monfort (1997) | A | 1 | M | H | AP | GS | 0.249 | 0.181 | 0.11111 |
| 5 | *Osmorhiza claytonii* | Williams (1994) | A | 1 | S | H | AP | AS | 0.953 | 0.390 | 0.09975 |
| 6 | *Medicago trunculata* | Bonnin *et al.* (2001) | M | 2 | S | H | AP | AS | 0.925 | 0.200 | 0.05446 |
| 7 | *Helicteres brevispira* | Franceschinelli & Kesseli (1999) | A | 1 | M | ST | AP | GS | 0.070 | 0.147 | 0.05291 |
| 8 | *Vouacapoua americana* | Dutech *et al.* (2002) | M | 2 | O | T | AP | GS | 0.068 | 0.075 | 0.03934 |
| 9 | *Sanicula odorata* | Williams (1994) | A | 2 | M | H | AP | AS | 0.344 | 0.105 | 0.03649 |
| 10 | *Arabidopsis lyrata* | Schierup unpublished | A | 4 | SI | H | AP | GS | 0.057 | 0.035 | 0.03268 |
| 11 | *Ancistrocladus korupensis* | Foster & Sork (1997) | A | 3 | M | ST | AP | WS | 0.455 | 0.150 | 0.03266 |
| 12 | *Hibiscus moscheutos* | Kudoh & Whigham (1997) | A | 2 | O | ST | AP | GS | 0.123 | 0.030 | 0.02565 |
| 13 | *Trillium grandiflorum* | Kalisz *et al.* (2001) | A | 1 | O | H | AP | AS | –0.020 | 0.122 | 0.02494 |
| 14 | *Eurya emarginata* | Chung & Epperson (2000) | A | 1 | SI | T | AP | GS | 0.000 | 0.055 | 0.02457 |
| 15 | *Calluna vulgaris* | Mahy *et al.* (1999) | A | 2 | O | ST | AP | GS | 0.050 | 0.048 | 0.02183 |
| 16 | *Primula elatior* | Van Rossum unpublished | A | 1 | SI | H | AP | GS | 0.233 | 0.162 | 0.02041 |
| 17 | *Silene acaulis* | Gehring & Delph (1999) | A | 5 | O | H | AP | GS | 0.093 | 0.097 | 0.01947 |
| 18 | *Sorbus torminalis* | Oddou-Muratorio unpublished | M | 1 | SI | T | AP | AS | –0.017 | 0.037 | 0.01689 |
| 19 | *Taraxacum sect. Ruderalia* | Meirmans *et al.* (2003) | A | 1 | O | H | AP | GS | –0.094 | 0.090 | 0.01639 |
| 20 | *Ipomopsis aggregata* | Campbell & Dooley (1992) | A | 1 | SI | H | AP | GS | 0.110 | 0.053 | 0.01626 |
| 21 | *Beta vulgaris* | Laporte *et al.* (2001) | MR | 2 | SI | H | WP | GS | 0.053 | 0.048 | 0.01526 |
| 22 | *Rhus javanica* | Epperson (2000) | A | 1 | SI | T | AP | AGS | 0.000 | 0.020 | 0.01445 |
| 23 | *Primula veris* | Van Rossum *et al.* unpublished | A | 4 | SI | H | AP | GS | 0.324 | 0.026 | 0.01400 |
| 24 | *Psychotria nervosa* | Dewey & Heywood (1988) | A | 1 | SI | ST | AP | AS | 0.034 | 0.019 | 0.01227 |
| 25 | *Primula vulgaris* | Van Rossum & Triest (2003) | A | 3 | SI | H | AP | AGS | 0.082 | 0.029 | 0.01115 |
| 26 | *Eryngium alpinum* | Gaudeul unpublished | M | 2 | O | H | AP | GS | 0.075 | 0.046 | 0.01100 |
| 27 | *Carapa procera* | Doligez & Joly (1997) | A | 1 | M | T | AP | AS | –0.210 | 0.033 | 0.01087 |
| 28 | *Centaurea jacea* | Hardy & Vekemans (2001) | A | 1 | SI | H | AP | GS | 0.104 | 0.181 | 0.01087 |
| 29 | *Pinus strobus* | Epperson & Chung (2001) | A | 1 | O | T | WP | WS | 0.078 | 0.032 | 0.01076 |
| 30 | *Centaurea corymbosa* | Hardy & Gonzalez-Martinez unpublished | M | 1 | SI | H | AP | GS | 0.028 | 0.050 | 0.01066 |
| 31 | *Psychotria officinalis* | Loiselle *et al.* (1995) | A | 1 | SI | ST | AP | AS | 0.055 | 0.047 | 0.01026 |
| 32 | *Acer saccharum* | Perry & Knowles (1991) | A | 3 | O | T | WP | WS | 0.000 | 0.018 | 0.01016 |
| 33 | *Caladenia tentactulata* | Peakall & Beattie (1996) | A | 1 | M | H | AP | GS | 0.130 | 0.034 | 0.00906 |
| 34 | *Quercus petraea* | Streiff *et al.* (1998) | M | 1 | O | T | WP | AGS | 0.059 | 0.031 | 0.00826 |
| 35 | *Chrysophyllum sanguinolentum* | H. Caron unpublished | A | 1 | O | T | AP | AS | –0.005 | 0.041 | 0.00826 |
| 36 | *Lesquerella fendleri* | Cabin (1996) | A | 1 | SI | H | AP | GS | 0.016 | 0.042 | 0.00755 |
| 37 | *Chamaecrista fasciculata* | Fenster *et al.* (2003) | A | 12 | M | H | AP | GS | 0.098 | 0.035 | 0.00746 |
| 38 | *Delphinium nuttallianum* | Williams & Waser (1999) | A | 4 | O | H | AP | GS | 0.099 | 0.018 | 0.00624 |
| 39 | *Sextonia rubra* | Hardy unpublished | M | 1 | O | T | AP | AS | 0.087 | 0.017 | 0.00610 |
| 40 | *Neolitsea sericea* | Chung *et al.* (2000) | A | 3 | SI | T | AP | AS | –0.073 | 0.002 | 0.00567 |

 

**Table 2** *Continued*

| | Species | Reference | Marker* | npop† | Breeding system‡ | Life form§ | Pollen dispersal¶ | Seed dispersal** | $F_I$†† | $F_{(1)}$‡‡ | *Sp* statistic |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | *Arabidopsis halleri* | Vekemans unpublished | M | 1 | SI | H | AP | GS | –0.031 | 0.161 | 0.00471 |
| 42 | *Dicorynia guianensis* | H. Caron unpublished | A | 1 | O | T | AP | AGS | –0.199 | 0.044 | 0.00460 |
| 43 | *Larix laricina* | Knowles *et al.* (1992) | A | 1 | O | T | WP | WS | 0.024 | 0.020 | 0.00450 |
| 44 | *Eperua grandiflora* | H. Caron unpublished | A | 1 | O | T | AP | GS | –0.297 | 0.021 | 0.00406 |
| 45 | *Quercus robur* | Streiff *et al.* (1998) | M | 1 | O | T | WP | AGS | 0.077 | 0.011 | 0.00298 |
| 46 | *Fraxinus excelsior* | Heuertz *et al.* (2003) | M | 1 | O | T | WP | WS | 0.029 | 0.029 | 0.00196 |
| 47 | *Virola michelii* | H. Caron unpublished | A | 1 | SI | T | AP | AS | 0.214 | 0.002 | 0.00031 |

*Type of genetic marker: A, allozymes; M, microsatellites; MR, microsatellites and nuclear random fragment length polymorphisms.
†Number of population samples analysed.
‡Plant breeding system: S, predominantly selfing; M, mixed-mating system; O, predominantly outcrossing; SI, self-incompatible.
§Plant life form: H, herbaceous species; ST, small trees or shrubs; T, trees.
¶Mode of pollen dispersal: AP, animal-dispersed pollen; WP, wind-dispersed pollen.
**Mode of seed dispersal: GS, gravity-dispersed seeds; AS, animal-dispersed seeds; WS, wind-dispersed seeds; AGS, mixed animal/gravity dispersed seeds.
††Wright's coefficient of inbreeding.
‡‡Kinship coefficient between adjacent individuals.

used to compute statistics $\hat{b}_F$ and $\hat{F}_{(1)}$, using the software SPAGEDI (Hardy & Vekemans 2002; http://www.ulb.ac.be/sciences/lagev) (for species 10, 16, 18, 23, 25, 26, 30, 35, 39, 41, 42, 44, 47).

For cases (ii) and (iii), when multiallelic multilocus estimates of pairwise relatedness coefficients [i.e. $\hat{F}(r)$] were not available (for species 3, 5, 12, 24, 32, 33, 43), they were calculated as averages by weighting the relatedness coefficient of each allele $k$ by its polymorphism index $\bar{p}_k(1 - \bar{p}_k)$ according to Loiselle *et al.* (1995), where $\bar{p}_k$ is its population allelic frequency. For case (iii), numerical values of the average pairwise relatedness coefficients for each spatial distance interval were deduced by digitalization of the published spatial autocorrelograms and manual recording of the plotted data using graphical software. Data presented as Moran's *I* statistics were transformed to values of kinship coefficients [$\hat{F}(r)$] using the following formula (Hardy & Vekemans 1999): $\hat{F}(r) = I(r)\ (1 + \hat{F}_I)/2$, where $\hat{F}_I$ is the inbreeding coefficient. For cases (ii) and (iii), values of $\hat{F}_{(1)}$ were taken from the average kinship coefficient over the smallest distance interval, and $\hat{b}_F$ was computed as the slope of the linear regression of the $\hat{F}(r)$ coefficients on ln($r$). In all cases, the *Sp* statistic was then computed as $-\hat{b}_F/(1 - \hat{F}_{(1)})$. In several species, SGS data were available for several populations. We then report the averages of $F_I$, $F_{(1)}$ and the *Sp* statistic over populations.

For cases (i) and (iv), $\hat{b}_F$ is thus computed from a regression using pairwise relatedness coefficients between each pair of individuals, $\hat{F}_{ij}$, whereas for cases (ii) and (iii) the regression is performed on averages of these coefficients over given distance intervals, $\hat{F}(r)$. To check whether the latter procedure biases the estimation of the *Sp* statistic, we applied it to 18 species for which we had the complete data set and compared the SGS statistics computed with both procedures. We obtained a coefficient of determination $R^2$ of 0.978 and the following equation: $Sp_{\text{full\_regression}} = -0.0035 + 1.045 \times Sp_{\text{average\_intervals}}$; which shows that the two procedures give essentially the same estimates.

Based on information from the published studies, we determined categories of breeding systems, life form and pollen and seed dispersal as follows (Table 2): (i) plant breeding systems classified as (a) predominantly selfing (S, average selfing rate higher than 90% and $F_I > 0.5$), (b) mixed-mating system (M, average selfing rate between 10% and 90%), (c) predominantly outcrossing (O, average selfing rate lower than 10% but not described as self-incompatible), (d) described as possessing a self-incompatibility system (SI); (ii) life form classified as (a) herbs (H), (b) small trees or shrubs (ST), or (c) trees (T); (iii) mode of pollen dispersal classified as (a) animal-dispersed (AP), (b) wind-dispersed (WP); (iv) mode of seed dispersal classified as (a) gravity-dispersed (GS), (b) animal-dispersed (AS) (c) mixed gravity-animal-dispersed (AGS) (d) wind-dispersed (WS). Differences among categories in average values of the *Sp*

**Table 3** Effect of biological characteristics of plant species on statistics of SGS and on the inbreeding coefficient

| | | *Sp* statistic | | $F_{(1)}$ | | $F_I$ | | $F_I - F_{(1)}$ | |
|---|---|---|---|---|---|---|---|---|---|
| Effect | *n* | Mean | SD* | Mean | SD | Mean | SD | Mean | SD |
| *Breeding system* | | | | | | | | | |
| Selfing (S) | 5 | 0.1431 | 0.0799 | 0.240 | 0.092 | 0.813 | 0.146 | 0.574 | 0.154 |
| Mixed mating (M) | 7 | 0.0372 | 0.0367 | 0.098 | 0.064 | 0.162 | 0.215 | 0.064 | 0.184 |
| Outcrossing (O) | 18 | 0.0126 | 0.0101 | 0.044 | 0.032 | 0.014 | 0.110 | −0.030 | 0.116 |
| Self-incompatible (SI) | 17 | 0.0134 | 0.0077 | 0.057 | 0.056 | 0.070 | 0.103 | 0.013 | 0.112 |
| | ANOVA† | $P < 0.001$ | | $P < 0.01$ | | $P < 0.001$ | | $P < 0.001$ | |
| *Life form* | | | | | | | | | |
| Herbaceous (H) | 24 | 0.0459 | 0.0643 | | | | | | |
| Small trees (ST) | 6 | 0.0259 | 0.0156 | | | | | | |
| Trees (T) | 17 | 0.0102 | 0.0096 | | | | | | |
| | ANOVA | $P < 0.01$ | | | | | | | |
| *Pollen dispersal* | | | | | | | | | |
| Animal-dispersed (AP) | 17 | 0.0171 | 0.0142 | | | | | | |
| Wind-dispersed (WP) | 6 | 0.0064 | 0.0038 | | | | | | |
| | ANOVA | n.s.‡ | | | | | | | |
| *Seed dispersal* | | | | | | | | | |
| Gravity-dispersed (G) | 6 | 0.0281 | 0.0166 | | | | | | |
| Wind-dispersed (W) | 5 | 0.0120 | 0.0121 | | | | | | |
| Animal-dispersed (A) | 8 | 0.0088 | 0.0050 | | | | | | |
| | ANOVA | n.s. | | | | | | | |

*SD, standard deviation.
†level of significance of a one-way ANOVA.
‡n.s., not significant.

statistic were determined using one-way analysis of variance ANOVAs. Data were log-transformed to satisfy the normality hypothesis. Because almost all herbaceous species investigated were pollinated by insects and had gravity-dispersed seeds, statistical analyses of the effect of propagule dispersal were restricted to tree and small tree species.

The effect of population density on the pattern of SGS was tested on four species for which spatial autocorrelograms were available in two populations differing for plant density (2, 6, 9, 12) and one species with data from five populations (17). To test the overall difference between low- and high-density populations, we performed a *t*-test for paired comparisons. For species 17, we discarded the population with intermediate plant density and arbitrarily separated the other populations into two pairs of low vs. high density populations. To test whether gene dispersal is affected by plant density, we computed inferred estimates of the parameter $\sigma^2$, half the average squared axial parent–offspring distance, based on the pattern of SGS according to the following formula: $< \sigma^2 > = 1/(4\pi Sp\delta)$, where $\delta$ is the estimate of population density (expressed as the number of individuals per square metre). Note that by doing so we implicitly assume that SGS is representative of an isolation-by-distance pattern at equilibrium, which is not demonstrated. Moreover, $\delta$ is the census rather than the effective density, so that the estimates are probably biased. Nevertheless, making the reasonable assumption that $D/\delta$ does not vary substantially among populations of the same species, the relative $< \sigma^2 >$ estimates remain comparable.

### *Results and discussion*

$\hat{F}_I$, $\hat{F}_{(1)}$ and the *Sp* statistic are shown for each species in Table 2. $\hat{F}_I$ ranged from −0.297–0.953, with an average of $0.141 \pm 0.271$ (SD) whereas $\hat{F}_{(1)}$ ranged from 0.002 to 0.390 with an average of $0.077 \pm 0.079$ (SD). The *Sp* statistic ranged from 0.00031 (in *Virola michelii*, a species for which the pattern of SGS was not significantly different from a random distribution of genotypes) to 0.263 (in the predominantly selfing species *Phaseolus lunatus*), with an average of $0.0304 \pm 0.0491$.

One-way analyses of variance performed using biological characteristics as main effects showed that plant breeding system and life form highly significantly influence patterns of SGS (Table 3). On average, *Sp* was 10 times higher in predominantly selfing than in predominantly outcrossing or self-incompatible species, with intermediate values for species with mixed mating systems. Similarly, *Sp* was about four times higher in herbaceous as compared to tree species,

**Table 4** Effect of plant density on the *Sp* statistic and on inferred standard deviation of gene dispersal in five species

| Species | Low density population | Density* | *Sp* | <σ>† | High density population | Density* | *Sp* | <σ>† |
|---|---|---|---|---|---|---|---|---|
| 2 *Cryptotaenia canadensis* | LD | 0.10 | 0.3107 | 1.60 | HD | 6.00 | 0.0388 | 0.58 |
| 6 *Medicago trunculata* | aude 3 | 0.71 | 0.0627 | 1.33 | aude 1 | 1.32 | 0.0462 | 1.14 |
| 9 *Sanicula odorata* | LD | 0.10 | 0.0548 | 3.81 | HD | 6.00 | 0.0181 | 0.86 |
| 12 *Hibiscus moscheutos* | site 8 | 0.95 | 0.0430 | 1.40 | site 2 | 2.20 | 0.0083 | 2.09 |
| 17 *Silene acaulis* | Krummholz | 0.09 | 0.0395 | 4.73 | slope 2 | 2.52 | 0.0144 | 1.48 |
| 17 *Silene acaulis* | Flat 1 | 0.14 | 0.0230 | 4.98 | Summit | 2.93 | 0.0032 | 2.92 |
| *t*-test for paired comparisons | | | $P < 0.01$ | $P = 0.07$ | | | | |

*density in plants expressed as number of individuals per square meter.
†inferred estimate of the standard deviation of the overall distribution of gene dispersal.
†*P*-values of a two-tailed *t*-test for paired comparisons between low *vs* high density populations.

with intermediate values for small trees and shrubs. In contrast, pollen and seed dispersal modes did not significantly influence patterns of SGS, perhaps in relation to the lower number of species included in these tests, although the trends indicated by *Sp* were in agreement with the expectation of (i) a stronger SGS in species with animal- vs. wind-dispersed pollen, and (ii) a stronger SGS in species with gravity- vs. animal-dispersed seeds.

The higher SGS observed in selfing species is a logical consequence of two phenomena. Firstly, the high level of inbreeding ($F_I$) substantially increases the rate of genetic drift, reducing the effective population density by a factor $1/(1 + F_I)$ (Pollak 1987). Secondly, in outcrossing plant species, pollen dispersal contributes to the overall gene dispersal (σ) whereas in highly selfing species, only seed dispersal contributes to σ. The first effect would increase the *Sp* statistic of highly selfing species by a factor of two compared to outcrossing species. The fact that a factor close to 10 was observed suggests that pollen dispersal generally provides the major contribution to σ within populations of outcrossing species. However, it is also possible that natural selection plays an important role in increasing SGS in selfing species as a response to environmental clines, because of the wide hitchhiking effects expected under high selfing.

Interestingly, the *Sp* statistic was correlated with life form, with roughly two- to four-fold higher values for herbaceous plants than shrubs and trees. This effect was still present when selfing species, which are rare among trees, were removed (data not shown). This is somewhat in contradiction to the results of the literature survey by Ennos (2001) who noted that the majority of woody perennial species did show significant patterns of SGS whereas it was the case for only five out of 11 outcrossing herbs. We suggest that comparisons among species based upon levels of significance of SGS taken from the literature are not reliable however, because of the potentially very different power of tests involving different methodologies, marker types, and sampling schemes. Because we did not gather original datasets for most of the studies included in our survey, we could not apply formal tests of the isolation-by-distance pattern.

Populations differing with respect to plant density were compared for their pattern of SGS (Table 4). In each of the six pairwise comparisons, the *Sp* statistic was consistently and significantly ($P < 0.01$) higher in low density, as compared to high density populations. Inferred estimates of the overall standard deviation of gene dispersal were found to be higher in low- vs. high-density populations, for five of the six pairwise comparisons. However, the difference was only marginally significant ($P = 0.07$ for a two-tailed test). These results show that density is a major determinant of the SGS as it affects the strength of local genetic drift. Indeed, the *Sp* statistic is expected to be inversely proportional to the density under isolation by distance (Heywood 1991). However, the results also suggest that gene dispersal distances were higher in low density populations, partially compensating for the direct effect of density. This is consistent with direct measures of pollen dispersal showing that pollinator flight distances increase when population density decreases (e.g. Fenster 1991; Schmitt 1983).

## Perspectives

### *Confronting direct and indirect estimates of gene dispersal*

A variety of approaches have been developed to assess gene dispersal in natural populations, from the most 'direct' ones, where dispersal events are followed (*in situ* monitoring of propagule movements) or reconstructed (paternity/parentage analyses), providing 'real-time' estimates, to the most 'indirect' ones based on equilibrium SGS, which provide 'historical' estimates. The 'Two-Gener' approach to assess pollen dispersal (Austerlitz & Smouse 2001, 2002; Smouse *et al*. 2001) is somewhat intermediate,

although it is 'indirect' because it is based on the pattern of genetic structure among pollen pools, it nevertheless provides 'real-time' estimates. These methods not only differ in the actual vs. historical nature of their estimates, but also in their precision and reliability with respect to *a priori* assumptions (e.g. the *Sp*-statistic-based approach assumes that SGS is representative of an isolation-by-distance pattern at equilibrium). Although 'direct' approaches such as parentage analyses are potentially more reliable and precise, they require exhaustive sampling of adults as well as highly polymorphic and reliable markers, so that they cannot be applied at a reasonable cost in all situations. Hence, it is important to compare 'direct' and 'indirect' estimates: (i) to validate indirect approaches, which are generally much easier to apply in plant populations, and (ii) to assess if the extent of gene dispersal remains stable across generations. Few such comparisons using the approach presented here are nowadays available, but we can cite the study on the annual legume *Chamaecrista fasciculata* where direct (from measures of seed and pollen movement) and indirect (from the pattern of SGS) *Nb* estimates matched very closely (Fenster *et al*. 2003).

### *Assessment of biparental inbreeding*

Biparental inbreeding is the contribution of mating events among relatives to the overall level of inbreeding within a population, as opposed to the contribution of selfing events. It can be quantified by the average relatedness between mates (excluding selfing). Waller & Knight (1989) expressed it as the 'genotypic correlation' (i.e. the 'relationship' coefficient) between truly outcrossed mates. It is more illuminating to express it as the average kinship coefficient between truly outcrossed mates, $Fx$, because $Fx$ is also the expected inbreeding coefficient of the offspring of these mates (provided that the coefficients are relative to the same reference level of relatedness). Hence, $Fx$ estimates can be compared to values of the statistic describing overall inbreeding, i.e. Wright's inbreeding coefficient, $F_I$. If $F_I = Fx$, mating among relatives is sufficient to explain the observed level of inbreeding. If $F_I > Fx$, selfing must occur. If $F_I < Fx$, inbreeding depression might be occurring.

When biparental inbreeding is the result of limited pollen dispersal within a population showing SGS, $Fx$ can be estimated by integrating the product $P(r) \times \hat{F}(r)$ over distance $r$, where $P(r)$ is the frequency distribution of pollen dispersal distances (excluding selfing). $P(r)$ can be obtained by direct monitoring of pollen dispersal events, for example through a paternity analysis. It can be shown that the relationship between the inbreeding coefficient and the selfing rate, $s$, is $s = 2(F_I - Fx)/(1 + F_I - 2Fx)$ (Fenster *et al*. 2003), reducing to Wright's (1951) formula, $s = 2F_I/(1 + F_I)$, in the absence of biparental inbreeding. Hence, an assessment of $Fx$ can help to distinguish the relative roles of selfing and mating among relatives to the overall level of inbreeding, an alternative to procedures based on progeny analyses (Ritland 2002). Even in the absence of information on $P(r)$, one may just assume that pollen is restricted to very short distances, so that a maximum estimate for $Fx$ would be $F_{(1)}$, the kinship coefficient between adjacent individuals. If $F_I$ is higher than this estimate, selfing is suggested. If it is lower, mating among relatives might exclusively explain the pattern of inbreeding. In strictly outcrossing species, one expects $F_I = Fx$, and a significant deviation might indicate that selective processes (e.g. inbreeding depression) occur ($F_I < Fx$) or that the estimated $F_I$ is biased upwards, for example because of the occurrence of null alleles. In practice the comparison is not easy because precise estimates of $F_I$ and $Fx$ are often difficult to obtain (i.e. to a precision of $10^{-2}$).

From our survey of empirical studies in plants, we observe in Table 3 that $\hat{F}_I >> \hat{F}_{(1)}$ in species with selfing or a mixed mating system, whereas the two estimates are similar in outcrossing species, as expected. One would also expect that $\hat{F}_I \leq \hat{F}_{(1)}$ in species with a self-incompatibility system, because they avoid selfing, but we observed that $\hat{F}_I$ was on average slightly larger than $\hat{F}_{(1)}$. This probably results from the low precision of $F_I$ estimates and potential bias because of null alleles.

### *Inferring seed vs. pollen dispersal with nuclear markers*

The regression approach to estimating dispersal parameters based on the $F(r)$ function exploits the information within a restricted distance range, neglecting what happens when $r < \sigma$. It was shown in Heuertz *et al*. (2003) that the form of $F(r)$ at short distances could be related to the relative contributions to the overall level of gene dispersal, $\sigma$, of seed dispersal, $\sigma_s$, vs. pollen dispersal, $\sigma_p$. More specifically, assuming Gaussian dispersal kernels, when $\sigma_s << \sigma_p$, $F(r)$ at $r < \sigma$ goes above the regression line (concave shape), whereas when $\sigma_s \geq \sigma_p$, $F(r)$ at $r < \sigma$ goes below the regression line (convex shape). Reliable inference of the $\sigma_p/\sigma_s$ ratio, however, requires precise $F(r)$ estimates, which necessitate a strong SGS and/or large sample size and/or many markers. For instance, precision was low when trying to infer $\sigma_p/\sigma_s$ in common ash using about 150 individuals and six very polymorphic microsatellite loci because the SGS was weak ($Sp = 0.002$, Heuertz *et al*. 2003). Another problem is that the shapes of the pollen and seed dispersal kernels also affect the shape of $F(r)$ at short distances. More generally, if $F(r)$ goes above the regression line at short distances, the gene dispersal kernel must be very leptokurtic (fat tail), which necessarily occurs if $\sigma_s << \sigma_p$, but is also the case if the seed and/or pollen dispersal kernels are very leptokurtic, independently of $\sigma_p/\sigma_s$.

Using the $F(r)$ graphs collected for the 47 species presented above, we tried to infer whether there was an overall trend regarding the shape of $F(r)$ estimates at short distance. Using the average $F(r)$ estimates per distance interval, we fitted the data to a polynomial function of third power: $f(r) = a + b\ln(r) + c[\ln(r)]^2 + d[\ln(r)]^3$, looking for the $a$, $b$, $c$ and $d$ parameters that minimize the difference between $f(r)$ and the $F(r)$ estimates. The choice of the third-power polynomial was justified by simulation results showing that it fits $F(r)$ well under isolation by distance for a wide range of conditions (Heuertz *et al*. 2003). The curvature of $f(r)$ is given by its second derivative, so that the initial curvature was estimated as $k = 2c + 6d\ln(r_1)$, where $r_1$ represents the middle of the first distance interval. Of the 47 data sets analysed, 32 (68%) gave $k > 0$, indicating concavity, and 15 gave $k < 0$, indicating convexity. Hence, there is an overall trend suggesting that gene dispersal is highly leptokurtic, possibly because $\sigma_s << \sigma_p$ in many plant species.

## Conclusion

This paper presents a first attempt to compare the within-population SGS among plant species in a quantitative way, based on robust theoretical models. We found clear effects of breeding system and life form on SGS, which mirror the well-established effects of these traits on plant population genetic structure (Hamrick & Godt 1990; Charlesworth & Pannell 2001). Population density was also found to affect the level of SGS. Pollen and seed dispersal, however, were not found to influence patterns of SGS significantly, but the number of species was very limited for these comparisons ($N = 23$). The $Sp$ statistic used for these comparisons is expected to be equal to the inverse of the neighbourhood size, $1/Nb$, under isolation by distance in two-dimensional space. We did not interpret our results in terms of $Nb$ estimates because we could not check for each reviewed species whether the conditions required for $Nb$ estimation (spatial scale, population geometry, stationary SGS) were fulfilled. Nevertheless, the fact that the results were consistent with *a priori* expectations based on knowledge of population density as well as dispersal processes suggests that the SGS observed in most species resulted from isolation by distance.

## Acknowledgements

We are grateful to all authors who have kindly provided their data sets for assessing SGS, in particular I. Bonnin, H. Caron, C. Dutech, M. Gaudeul, A. Kremer, V. Laporte, M. Schierup, R. Streiff, F. Van Rossum and I. Zoro Bi. We thank D. Charlesworth, F. Van Rossum, the subject editor and the two reviewers for their suggestions to improve the manuscript. This work was supported by the Belgian National Fund for Scientific Research (FNRS) where O. J. Hardy is a postdoctoral researcher.

## References

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relationship between $\Phi_{FT}$, pollen dispersal and inter-females distances. *Genetics*, **157**, 851–857.

Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics*, **161**, 355–363.

Barton NH, Depaulis F, Etheridge AM (2002) Neutral evolution in spatially continuous populations. *Theoretical Population Biology*, **61**, 31–48.

Bonnin I, Ronfort J, Wozniak F, Olivieri I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). *Molecular Ecology*, **10**, 1371–1384.

Cabin RJ (1996) Genetic comparisons of seed bank and seedling populations of a perennial desert mustard, *Lesquerella fendleri*. *Evolution*, **50**, 1830–1841.

Campbell DR, Dooley JL (1992) The spatial scale of genetic differentiation in a hummingbird-pollinated plant: comparison with models of isolation by distance. *American Naturalist*, **139**, 735–748.

Caujapé-Castells J, Pedrola-Monfort J (1997) Space-time patterns of genetic structure within a stand of *Androcymbium gramineum* (Cav.) McBride (Colchicaceae). *Heredity*, **79**, 341–349.

Charlesworth D, Pannell JR (2001) Mating systems and population genetic structure in the light of coalescent theory. In: *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 73–93. Blackwell Science, Cambridge.

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

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

Dewey SE, Heywood JS (1988) Spatial genetic structure in a population of *Psychotria nervosa*. I. Distribution of genotypes. *Evolution*, **42**, 834–838.

Doligez A, Joly HI (1997) Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. *Heredity*, **79**, 72–82.

Dutech C, Seiter J, Petronelli 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.

Ennos RA (2001) Inferences about spatial processes in plant populations from the analysis of molecular markers. In: *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 45–71. Blackwell Science, Cambridge.

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

Epperson BK (2000) Spatial genetic structure and non-equilibrium demographics within plant populations. *Plant Species Biology*, **15**, 269–279.

Epperson BK, Li T (1996) Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proceedings of the National Academy of Sciences of the USA*, **93**, 10528–10532.

Epperson BK, Chung MG (2001) Spatial genetic structure of allozyme polymorphisms within populations of *Pinus strobus* (Pinaceae). *American Journal of Botany*, **88**, 1006–1010.

Fenster CB (1991) Gene flow in *Chamaecrista fasciculata* (Leguminosae). I. Gene dispersal. *Evolution*, **45**, 398–409.

Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution*, **57**, 995–1007.

Foster PF, Sork VL (1997) Population and genetic structure of the West African rain forest liana *Ancistrocladus korupensis* (Ancistrocladaceae). *American Journal of Botany*, **84**, 1078–1091.

Franceschinelli EV, Kesseli R (1999) Population structure and gene flow of the Brazilian shrub *Helicteres brevispira*. *Heredity*, **82**, 355–363.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research Cambridge*, **6**, 95–107.

Gehring JL, Delph LF (1999) Fine-scale genetic structure and clinal variation in *Silene acaulis* despite high gene flow. *Heredity*, **82**, 628–637.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown HD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates Inc., Sunderland, MA.

Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology*, **12**, 1577–1588.

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 (2001) Patterns of allozymic variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. *Evolution*, **55**, 943–954.

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

Heuertz M, Vekemans X, Hausman J-F, Palada M, Hardy OJ (2003) Estimating seed versus pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, **12**, 2483–2495.

Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, **22**, 335–355.

Husband BC, Barrett SCH (1992) Effective population size and genetic drift in tristylous *Eichhornia paniculata* (Pontediraceae). *Evolution*, **46**, 1875–1890.

Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history and selection. *Evolutionary Biology*, **55**, 1560–1568.

Kimura M, Crow JF (1963) The measurement of effective population number. *Evolution*, **17**, 279–288.

Knowles P, Perry DJ, Foster HA (1992) Spatial genetic structure in two tamarack [*Larix laricina* (Du Roi) K. Koch] populations with differing establishment histories. *Evolution*, **46**, 572–576.

Kudoh H, Whigham DF (1997) Microgeographic genetic structure and gene flow in *Hibiscus moscheutos* (Malvaceae) populations. *American Journal of Botany*, **84**, 1285–1293.

Laporte V, Viard F, Bena G, Valero M, Cuguen J (2001) The spatial structure of sexual and cytonuclear polymorphism in the gynodioecious *Beta vulgaris* ssp. *maritima*: I/ at a local scale. *Genetics*, **157**, 1699–1710.

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.

Li CC, Weeks DE, Chakravarti A (1993) Similarity of DNA finger-prints due to chance and relatedness. *Human Heredity*, **43**, 45–52.

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.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.

Mahy G, Vekemans X, Jacquemart A-L (1999) Patterns of allozymic variation within *Calluna vulgaris* populations at seed bank and adult stages. *Heredity*, **82**, 432–440.

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

Meirmans PG, Vlot EC, Den Nijs JCM, Menken SBJ (2003) Spatial ecological and genetic structure of a mixed population of sexual diploid and apomictic triploid dandelions. *Journal of Evolutionary Biology*, **16**, 343–352.

Peakall R, Beattie AJ (1996) Ecological and genetic consequences of pollination by sexual deception in the orchid *Caladenia tentaculata*. *Evolution*, **50**, 2207–2220.

Perry DJ, Knowles P (1991) Spatial genetic structure within three sugar maple (*Acer saccharum* Marsh.) stands. *Heredity*, **66**, 137–142.

Pollak E (1987) On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, **117**, 353–360.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research Cambridge*, **67**, 175–185.

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. *Heredity*, **88**, 221–228.

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 (2001) Inference from spatial population genetics. In: *Handbook of Statistical Genetics* (eds Balding D, Bishop M, Cannings C), pp. 239–269. John Wiley & Sons Ltd, Chichester.

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

Schmitt J (1983) Density-dependent pollinator foraging, flowering phenology, and temporal pollen dispersal patterns in *Linanthus bicolor*. *Evolution*, **37**, 1247–1257.

Slatkin M, Arter HE (1991) Spatial autocorrelation methods in population genetics. *American Naturalist*, **138**, 499–517.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution*, **55**, 260–271.

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

Van de Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, **10**, 1539–1549.

Van Dijk H (1987) A method for the estimation of gene flow parameters from a population structure caused by restricted gene flow and genetic drift. *Theoretical and Applied Genetics*, **73**, 724–736.

Van Rossum F, Triest L (2003) Spatial genetic structure and reproductive success in fragmented and continuous populations of Primula vulgaris. *Folia Geobotanica*, **38**, 239–254.

Waller DM, Knight SE (1989) Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. II. Outcrossing rates and genotypic correlations. *Evolution*, **43**, 860–869.

Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–1215.

Williams CF (1994) Genetic consequences of seed dispersal in three sympatric forest herbs. II. Microspatial genetic structure within populations. *Evolution*, **48**, 1959–1972.

Williams CF, Waser NM (1999) Spatial genetic structure of *Delphinium nuttallianum* populations: inferences about gene flow. *Heredity*, **83**, 541–550.

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

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Zoro Bi I, Maquet A, Baudoin J-P (1997) Spatial patterns of allozyme variants within three wild populations of *Phaseolus lunatus* L. from the central valley of Costa Rica. *Belgian Journal of Botany*, **129**, 149–155.

---

X. Vekemans is Professor in plant population genetics and recently set up a research team studying population biology and molecular evolution of self-incompatibility systems in plants. He is also involved in investigations on plant population genetic structure. O.J. Hardy is a postdoctoral researcher developing approaches to infer gene dispersal parameters and adaptive responses at quantitative traits from the analysis of spatial genetic structure. He is also developing computer programs for data analyses.

---