Fine-Scale Spatial Genetic Structure with Nonuniform Distribution of Individuals

Agnès Doligez, Claire Baril[1] and Hélène I. Joly

CIRAD-Forêt, B. P. 5035, 34032 Montpellier Cedex 01, France

Manuscript received March 3, 1997
Accepted for publication October 10, 1997

ABSTRACT

This paper presents the first theoretical study of spatial genetic structure within nonuniformly distributed continuous plant populations. A novel individual-based model of isolation by distance was constructed to simulate genetic evolution within such populations. We found larger values of spatial genetic autocorrelations in highly clumped populations than in uniformly distributed populations. Most of this difference was caused by differences in mean dispersal distances, but aggregation probably also produced a slight increase in spatial genetic structure. Using an appropriate level of approximation of the continuous distribution of individuals in space, we assessed the potential effects of density, seed and pollen dispersal, generation overlapping, and overdominance selection at an independent locus, on fine-scale genetic structure, by varying them separately in a few particular cases with extreme clumping. When selfing was allowed, all these input variables influenced both aggregation and spatial genetic structure. Most variations in spatial genetic structure were closely linked to variations in clumping and/or local density. When selfing was not allowed, spatial genetic structure was lower in most cases.

BEGINNING with Wright (1943), many authors have studied the combined effects of genetic drift and restricted gene dispersal on the evolution and structure of genetic diversity within large continuous populations. They have derived several analytical results on spatial genetic structure under the assumptions of the isolation by distance model, among which is a general relationship between kinship and distance (Malécot 1948). But these results cannot be directly applied to studies of real populations (Epperson 1993), mainly because genealogical data are seldom available, and most of the time only spatial distributions of genotypes are known. Moreover, this analytical approach implicitly ignores stochastic effects during dispersal events, although these effects are critical in the development of patchy structures of genotypes that result from isolation by distance (Epperson 1995a).

More recently, a few authors have used individual-based stochastic simulation models to describe spatial patterns of genotypic variation within continuous uniformly distributed populations, under various conditions of gene dispersal, mutation, migration, allele frequency, turnover of individuals, and selection (Rohlf and Schnell 1971; Turner *et al.* 1982; Sokal and Wartenberg 1983; Bos and Van der Haring 1988; Sokal *et al.* 1989; Epperson 1990; Ohsawa *et al.* 1993; Berg and Hamrick 1995; Epperson 1995a,b). Several autocorrelation statistics were used to describe this structure, join-count statistics being the most informative. These simulation studies showed that large homozygote patches developed when gene dispersal was spatially restricted. Spatial autocorrelations were reduced by the occurrence of mutation-additive selection directly at the locus under study or by migration from an infinite source population (Epperson 1990). Slatkin and Barton (1989), Kawata (1995), and Ronfort and Couvet (1995) have simulated continuous populations with nonuniform distributions of individuals. But they used only global F_{ST} statistics to describe spatial genetic structure, because detailed description of this structure was not their primary concern.

There is a complete lack of theoretical studies describing spatial genetic structure within plant populations with nonuniform and thus more realistic spatial distributions of individuals. Most distributions are either random or clumped in nature (see Hubbell 1979 and Armesto *et al.* 1986 for examples in temperate and tropical forests). Therefore, before introducing various density- (or distance-) dependence mechanisms in simulated populations [as did Kawata (1995) and Ronfort and Couvet (1995) to avoid clumping], on which few quantitative data are available for natural plant populations (Antonovics and Levin 1980), it appears necessary to explore the potential relationships between spatial distribution of individuals and spatial genetic structure.

Corresponding author: Hélène I. Joly, CIRAD-Forêt, U. R. Diversité et Amélioration Génétique, Baillarguet, B. P. 5035, 34032 Montpellier Cedex 01, France. E-mail: joly.h@cirad.fr

[1] *Present address:* GEVES, La Minière, 78285 Guyancourt Cedex, France.

The study of these relationships presents an opportunity to understand better the determinants of spatial genetic structure in real populations. In addition, it may have several fields of application, for example, inference of the factors responsible for spatial patterns observed in natural populations, or prediction of the effects of various management practices on the evolution of diversity level and structure to provide guidelines for *in situ* conservation programs.

Therefore we propose a novel individual-based stochastic model of isolation by distance allowing for different kinds of spatial distributions, from uniform to highly clumped. Our objective in this paper is to determine which of several input variables (density, dispersal, generation overlapping, and selection) may affect spatial genetic differentiation within the population under extreme clumping conditions. We have used Monte Carlo simulations to examine the consequences of separate variations of these variables in a few particular cases. Spatial distribution of individuals was characterized by means of a statistic based on nearest neighbor distances, and spatial genetic structure by means of autocorrelation statistics. Whenever some effect on spatial genetic structure was found, we have proposed hypotheses to explain the mechanisms involved and have described relationships with the spatial distribution of individuals.

METHODS

The model: Each simulated population consisted of 10,000 mature, hermaphroditic, diploid individuals. Individuals were located at the intersection points of a square grid of size $L \times L$, with L measured in grid units. There was at most one individual at each intersection point. Thus the grid unit was the minimum distance allowed between two individuals. The grid size could be larger than the population size. There was no density-dependent regulation, apart from the minimum spacing of individuals due to their location on a grid.

Each individual of the population was characterized by its location on the grid (x,y coordinates), its genotype at a neutral marker locus, and its genotype at each of a few loci undergoing symmetrical overdominance selection with multiplicative effects between selected loci. Individual fitness was defined as $(1\text{-}S)^k$ where S was the selection coefficient (identical for all selected loci) and k was the number of homozygous selected loci. All loci were bi-allelic and segregated independently from each other. There was no mutation and no migration.

Time was discrete and reproduction occurred synchronously throughout the population. In each time cycle, a number of new individuals was created. In case of nonoverlapping generations, 10,000 new individuals were created in each cycle, and all individuals alive in the preceding cycle died. In case of overlapping generations, the population was divided into C age classes of identical size, and $10{,}000/C$ new individuals were created in each cycle. Only those $10{,}000/C$ individuals alive in the preceding cycle that belonged to the oldest age class died, and all others stayed alive and moved up one age class. Thus in all cases, the population size was kept constant through cycles, it took C cycles for a complete turnover of individuals, and the lifespan of each individual was C cycles. In what follows, we will refer to generation overlapping as C, that is, the maximum number of successive reproduction cycles that might occur between the birth of two mates, including the cycles during which these mates were created.

To create each new individual in a given cycle, the following procedure was used (all distances in grid units): (1) One mother individual was drawn at random among all individuals alive in the preceding cycle. (2) One point of the grid was then drawn at random among all grid points located within a circle of radius D centered on the mother. (3) If no individual (be it a new individual or an individual of the preceding cycle in case of generation overlapping) was already present at that point, a father individual was then drawn at random among all individuals of the preceding cycle located within a circle of radius P centered on the mother. Otherwise, the procedure was started again from step (1). (4) At each selected locus, Mendelian segregation of genes was assumed to obtain the genotype of the new individual. Lastly, a number was drawn at random from a uniform [0; 1] distribution and the new individual was actually created if this number was smaller than its fitness value (see definition above). Otherwise no new individual was created. The whole procedure [steps (1)–(4)] was carried out as many times as necessary to create the appropriate number of new individuals in each cycle.

For a given simulation, all individuals were either self-compatible (could self) or self-incompatible (could not self). When all individuals were self-compatible, mothers were one of the potential fathers of their own progeny, so that outcrossing rate was determined by local density. When all individuals were self-incompatible, isolated individuals could not reproduce and outcrossing rate was 1. It should be noted that no gametophytic or sporophytic self-incompatibility genetic system was used here, but only the possibility or impossibility for individuals to self.

The grid was assumed toroidal (*i.e.*, the top and bottom of the grid were joined together, and so were the two sides), to avoid edge effects. Such connections inevitably lead to some spatial overlapping, which is expected to prevent the development of spatial genetic structure for small grid sizes. We checked that a grid size of 100×100 was sufficient for this overlapping effect to become negligible, by comparing spatial genetic autocorrelations obtained for values of L increasing from 25 to 125 by 25, for $D = P = C = 1$ and $S = 0$ (data not shown). Moreover, as grid sizes were large, this assumption was expected not to affect, too much, comparisons with previous simulation studies such as Epperson's (1990, 1995a,b), which rely on different assumptions concerning edge effects.

Simulation program: The simulation program was written in C, and run on Hewlett-Packard work stations under a Unix operating system. We used the random number generator recommended by Sedgewick (1988), which was based on an additive congruential method, because of its efficiency and good statistical properties. One separate seed was used for each replicate simulation. Each seed was validated by a chi-square test of uniform distribution using a sequence of 30 million numbers.

Simulation runs: To create initial populations for each simulation, the coordinates and genes of individuals were drawn at random, with initial allele frequency 0.5 at each locus. All simulations were run over $200 \times C$ cycles. Previous simulation studies of the isolation by distance model with nonoverlapping generations have shown that 200 generations were enough for spatial genetic structure to stabilize in most cases (Sokal and Wartenberg 1983; Epperson 1990). Note that cases with and without overlap of generations were compared for the same number of complete turnovers rather than for the same number of generations (with the generation length defined as the average time between the births of parents and

TABLE 1

Values of input variables used in the simulations

Set	Subset	Input variables							
		L	*D*	*P*	*C*	*S*	*GS*	*GA*	*W*
Grid area constant									
1	*a*[a]	**100**	**1**	**1**	1	0.0	**60**	3,600	**1**
	b	**200**	**2**	**2**	1	0.0	**30**	3,600	**2**
	c	**300**	**3**	**3**	1	0.0	**20**	3,600	**3**
	d	**400**	**4**	**4**	1	0.0	**15**	3,600	**4**
	e	**500**	**5**	**5**	1	0.0	**12**	3,600	**5**
	f[a]	**1,000**	**10**	**10**	1	0.0	**6**	3,600	**10**
	g	**1,500**	**15**	**15**	1	0.0	**4**	3,600	**15**
	h	**2,000**	**20**	**20**	1	0.0	**3**	3,600	**20**
Grid spacing constant									
2	*a*	**100**	10	10	1	0.0	5	**25**	1
	b	**500**	10	10	1	0.0	5	**625**	1
	ref	**1,000**	10	10	1	0.0	5	**2,500**	1
	d	**1,500**	10	10	1	0.0	5	**5,625**	1
	e[a]	**2,000**	10	10	1	0.0	5	**10,000**	1
3	ref	1,000	**10**	10	1	0.0	5	2,500	1
	b	1,000	**20**	10	1	0.0	5	2,500	1
	c	1,000	**50**	10	1	0.0	5	2,500	1
	d	1,000	**100**	10	1	0.0	5	2,500	1
4	ref	1,000	10	**10**	1	0.0	5	2,500	1
	b	1,000	10	**20**	1	0.0	5	2,500	1
	c	1,000	10	**50**	1	0.0	5	2,500	1
	d	1,000	10	**100**	1	0.0	5	2,500	1
5	ref	1,000	10	10	**1**	0.0	5	2,500	1
	b	1,000	10	10	**5**	0.0	5	2,500	1
	c[a]	1,000	10	10	**10**	0.0	5	2,500	1
6	ref	1,000	10	10	1	**0.0**	5	2,500	1
	b	1,000	10	10	1	**0.2**	5	2,500	1
	c	1,000	10	10	1	**0.4**	5	2,500	1
	d[a]	1,000	10	10	1	**0.5**	5	2,500	1
	e	1,000	10	10	1	**0.6**	5	2,500	1
	f	1,000	10	10	1	**0.8**	5	2,500	1
	g	1,000	10	10	1	**1.0**	5	2,500	1

Headings: *L*, length of the grid (in grid units); *D*, maximum speed dispersal distance (in grid units); *P*, maximum pollination distance (in grid units); *C*, number of age classes; *S*, selection coefficient; *GS*, grid spacing (in any arbitrary unit; *GS* is used only for computation of d_w values); *GA*, grid area [$GA = (L \times GS)^2$]; *W*, width of first 30 distance classes for spatial autocorrelations (in grid units).

[a] Subsets for which the evolution of spatial statistics in time was recorded for the first 10 replicates. The subset of simulations labeled ref is common to sets *2–6*.

offspring), so that the total number of individuals created during each simulation run was constant. Fifty replicate simulations were run for each combination of values of the following input variables (Table 1): length of grid (*L*), maximum seed dispersal distance (*D*), maximum pollen dispersal distance (*P*), number of age classes (*C*), and selection coefficient (*S*).

The objective of simulations in set *1* was to determine the effects, on spatial genetic structure, of approaching the continuous distribution of individuals by decreasing grid spacing with a fixed global density. Note that subset *1a* (uniform distribution and minimal dispersal) is very close to the models of Turner *et al.* (1982), Sokal and Wartenberg (1983), and Epperson (1990). The results of set *1* were then used to choose, for the reference situation in sets *2–6*, the grid spacing that yielded the best compromise between the level of approximation of the continuous distribution and the computation time required. The objective of simulations in sets *2–6* was to study the effects of the variables listed above on spatial genetic structure within clumped populations.

In sets *1* and *2*, the total number of grid points ($L \times L$) varied. But in set *1* grid spacing was varied in order to keep the total grid area constant, whereas in set *2* the total grid area was varied in order to keep grid spacing constant. In all other sets, both grid spacing and total grid area were kept constant, and one other input variable varied (set *3*: *D*; set *4*: *P*; set *5*: *C*; set *6*: *S*). In set *6*, there was only one selected locus. All simulations were carried out once assuming self-compatibility (SC) and once assuming self-incompatibility (SI). As a complement to set *6*, a few additional simulations with SC were run, to assess the effect of 10 selected loci with much lower selection coefficients (0.01, 0.02, 0.05, 0.10, 0.15, and 0.20).

Statistics measured: The following variables were recorded every $5 \times C$ cycles: outcrossing rate after selection (effective outcrossing rate, T_e); allele frequency (P_A); genotypic frequencies (P_{AA}, P_{aa}, and P_{Aa}). To characterize spatial genetic structure, we used two complementary spatial autocorrelation statistics (Upton and Fingleton 1985): Moran's *I*-statistics for individual allele frequencies [the method of Heywood (1991)], which give a summary measure of genetic similarity between individuals as a function of distance; and join-count statistics, which more accurately describe spatial distribution of genotypes, since they display all the detailed information summarized in Moran's *I* (Epperson 1995b). To characterize the spatial distribution of individuals, we used Diggle's d_w statistic, which is based upon the distributions of nearest neighbor distances (Upton and Fingleton 1985) and measures the deviation from a random distribution. We chose this statistic among all the ones that have been developed to detect departures from randomness so far because it is useful for small as well as for large populations, unlike statistics based on sampling or quadrat counting. Therefore, it makes it easier to compare theoretical results with data obtained for real populations even of relatively small sizes. Statistics involving all inter-plant distances and not merely nearest neighbor distances (such as Ripley's statistic) and statistics using tessellations are more informative but considerably more demanding on computer time. How all these statistics quantify various aspects of nonrandomness is an open issue.

These spatial statistics were computed after the last cycle for all simulations. To get an idea of their patterns of evolution in time, we also computed these statistics every $10 \times C$ cycles for only a few simulation subsets with contrasted values of input variables, because of the very large amount of computation time required (Table 1). Subsets with the highest dispersal distances were not included because spatial genetic structure was very weak after the last cycle. All genetic statistics (fixation indices, *I*- and join-count statistics) were computed both at the neutral locus and at the selected loci.

Moran's *I*-statistic for individual allele frequencies is defined as:

$$I = \frac{n}{\sum_{i,j} w_{ij}} \frac{\sum_{i,j} w_{ij}(x_i - x)(x_j - x)}{\sum_i (x_i - \bar{x})^2}$$

where *n* is the number of individuals in the population (*e.g.*, 10,000), w_{ij} is a weight whose value decreases with the geographic distance between individuals *i* and *j*, x_i and x_j are the individual allele frequencies of individuals *i* and *j*, respectively (1 for genotype *AA*, 0.5 for genotype *Aa*, and 0 for genotype

aa), and $\bar{x}$ is the mean allele frequency in the population. This statistic was computed for different distance classes by setting w_{ij} to one when the distance between individuals *i* and *j* belonged to the distance class under consideration, and to zero otherwise. For simulation set *1* (Table 1), 30 distance classes were used. The first class was]0; d + 0.5], and subsequent classes were]k; k + d], where k varied from d + 0.5 to 29d + 0.5 by d, and d was the distance class width (see Table 1). For all other simulation sets, wider distance classes were used for large distances in order to save computing time, with a total of 37 distance classes. The first class was]0; 1.5], the 29 following classes were]k; k + 1], where k varied from 1.5 to 29.5 by 1, and the last seven classes were]k, k + 10], where k varied from 29.5 to 99.5 by 10. We performed no test of the null hypothesis of random distribution of genotypes with respect to individual locations, because there is no evidence that a normal approximation can be used for the sampling distribution of *I*-statistics for individual allele frequencies, and a Monte Carlo approach would have been far too time-consuming.

Join-count statistics were computed as follows. Using the same distance classes as for *I*-correlograms, all (10,000 × 9,999)/2 pairs of individual genotypes were classified according to both distance and type of join (*AA-AA*, *aa-aa*, *Aa-Aa*, *AA-Aa*, *Aa-aa*, or *AA-aa*), which allowed computation of all $n_{ij}(k)$ [where $n_{ij}(k)$ is the number of joins between genotypes *i* and *j* in distance class *k*]. The expected values $\mu_{ij}(k)$ and standard deviations $SD_{ij}(k)$ of each $n_{ij}(k)$ under the null hypothesis, H_0, that genotypes are distributed at random with respect to individual location, are given in Upton and Fingleton (1985) for $i = j$ and $i \neq j$. Under H_0, the statistics $\mathrm{SND}_{ij}(k) = [n_{ij}(k) - \mu_{ij}(k)]/SD_{ij}(k)$ are distributed asymptotically as standard normal deviates (Cliff and Ord 1981).

Diggle's d_w statistic is defined as follows. Let the random variable W be the nearest neighbor distance. Under the assumption of random spatial distribution of individuals (not only on grid intersections, but continuously over the whole area), the expected cumulative distribution function of W is $G(w) = 1 - (\exp - \rho\pi w^2)$, where ρ is the mean intensity of individuals per unit area. Then $d_w = \sup |G(w) - \hat{G}(w)|$, where $\hat{G}(w)$ is the observed distribution function for the sample individuals on grid intersections. d_w takes values between 0 and 1. If $G(w) > \hat{G}(w)$, there is excess uniformity; if $\hat{G}(w) > G(w)$, there is excess clumping. For each simulation, we tested the null hypothesis of random distribution of individuals on the grid using a Monte Carlo approach since no approximation of the sampling distribution of d_w is known. For each combination of grid spacing and total grid area, we generated 1000 separate random configurations of 10,000 points on the intersections of the grid, and computed d_w for each configuration. These 1000 simulated values were used as an estimate of the sampling distribution of d_w under the assumption of random distribution of individuals.

RESULTS

Effect of grid spacing under constant global density and dispersal distances: For all simulations of set *1*, patterns of mean *I*-correlograms (Figure 1) and SND-correlograms were typical of a patchy distribution of genotypes, with large homozygote patches separated by smaller and less homogenous heterozygote patches, as already described in previous simulation studies. Moran's *I*-values were highly positive in short and intermediate distance classes, which means that two individuals separated by such a given distance had more similar genotypes than expected under the assumption of random distribution of genotypes in space.

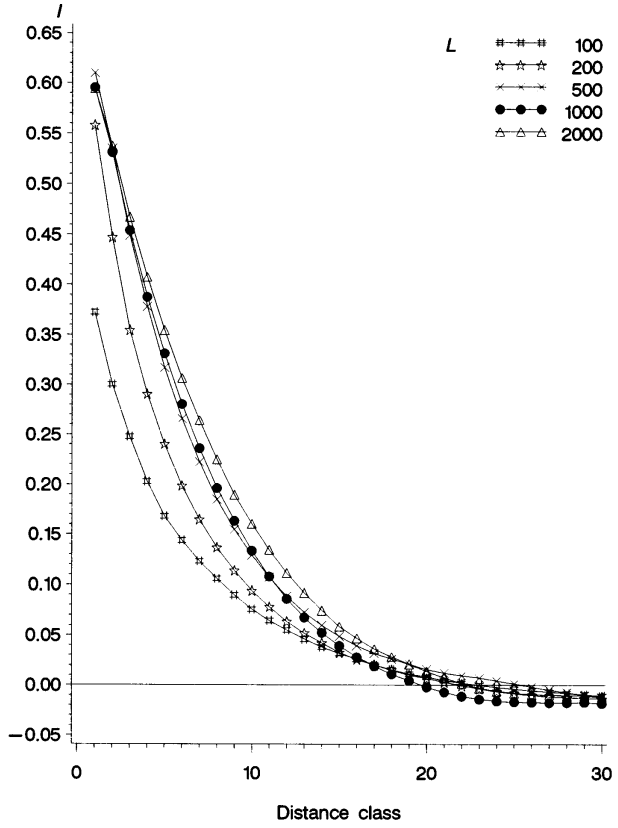

Figure 1.—Average *I*-correlograms after the last cycle for subsets *1a* (L = 100), *1b* (L = 200), *1e* (L = 500), *1f* (L = 1000), and *1h* (L = 2000), with SC.

With self-compatibility (SC), Moran's *I*-values in small and intermediate distance classes increased when L was increased. They stabilized for $L \geq 1000$ (Figure 2). Values of *X*-intercepts varied little with L. This means that two individuals separated by a given small or intermediate distance had more similar genotypes when the minimum distance allowed between individuals (grid spacing) was smaller.

At short to intermediate distances, when L was increased, SNDs for joins between like homozygotes decreased, SNDs for *Aa-Aa* joins increased, there was no important change in SNDs for joins between one heterozygote and one homozygote, and SNDs for *AA-aa* joins decreased. Such variations reflected larger distances between unlike homozygote patches, higher concentration of heterozygotes, and reduced homogeneity of homozygote patches. These conflicting effects resulted in increased autocorrelations for individual allele frequencies.

The aforementioned changes were closely related to changes in the amount of aggregation, as shown by both maps of individuals (Figure 3) and values of Diggle's d_w statistic (Figure 4), but also to changes in effec-

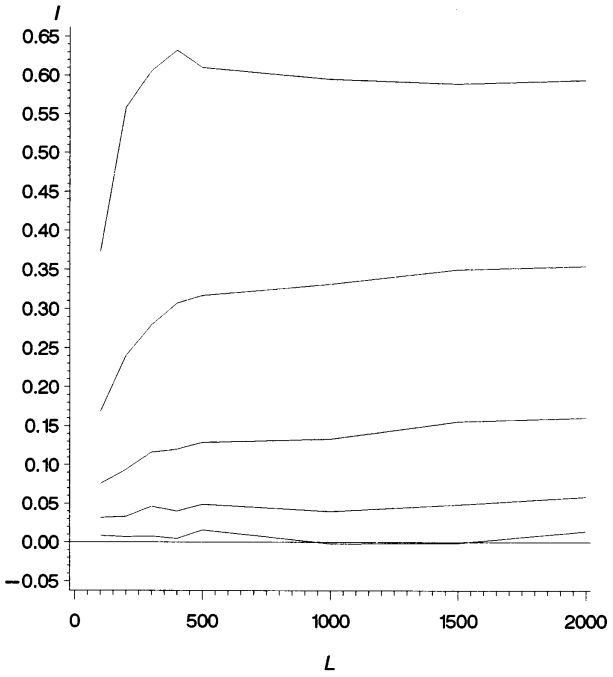

Figure 2.—Average *I*-statistics in distance classes 1, 5, 10, 15, and 20 (in this order from top to bottom) after the last cycle for all subsets of simulation set *1*, with SC.

Figure 3.—Spatial distribution of individual genotypes in a subpart (of size 500 grid units × 700 grid units) of the total grid after the last cycle in one replicate of subset *1f*, with SC; −, o, |, for genotype *AA*, *Aa*, and *aa*, respectively.

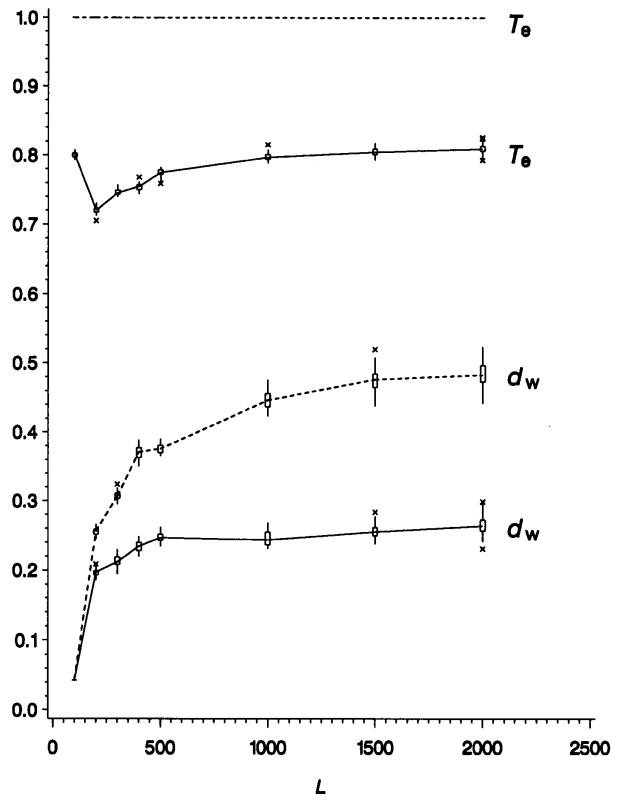

Figure 4.—Plot of Diggle's d_w statistic and effective outcrossing rate T_e for all subsets of simulation set *1*, with SC (——) or with SI (- - - - - - -), as functions of *L*. The bottom and top edges of each box are located at the sample 25th and 75th percentiles; the vertical lines (whiskers) are drawn from the box to the most extreme point within 1.5 interquartile range (which is the distance between the 25th and the 75th sample percentiles), and any value more extreme than this is marked with an x.

tive dispersal distances (Table 2). The higher values of Moran's *I*-statistics for cases with nonuniform distribution (subsets *1b* to *1h*), compared to the case with uniform distribution (subset *1a*), are consistent with the lower dispersal distances and variances in dispersal distances. However, it is not excluded that the higher aggregation also had some effect on its own, in the following way: unlike homozygote patches became more and more distant, which could partly explain both the decrease in SNDs for like homozygote joins and the decrease in SNDs for *AA-aa* joins. Heterozygotes grouped at the fewer areas of contact between homozygote patches, instead of almost continuously bordering homozygote patches as was the case when individuals were uniformly distributed (*i.e.*, for $L = 100$). This led to a reduced number of heterozygote patches in the population, which could contribute to the increase in SNDs for *Aa-Aa* joins. Lastly, in these contact areas, there was more mixing of genotypes within distance classes, which could also enhance the decrease in SNDs for like homozygotes. However, the simulations of set *1* do not

TABLE 2

Mean and variance of effective dispersal distances for simulations of set *1*

	Mean effective dispersal distance					
	Seed		Pollen		σ^2_{axial}	
Subset	With SC	With SI	With SC	With SI	With SC	With SI
a	0.800	0.801	0.800	1.000	0.600	0.650
b	0.680	0.680	0.528	0.734	0.374	0.415
c	0.677	0.678	0.519	0.697	0.361	0.395
d	0.658	0.659	0.501	0.668	0.338	0.369
e	0.679	0.678	0.527	0.683	0.360	0.390
f	0.670	0.670	0.530	0.668	0.352	0.378
g	0.667	0.667	0.532	0.665	0.350	0.375
h	0.666	0.666	0.535	0.664	0.350	0.374

σ^2_{axial} is the variance of the parent-offspring axial dispersal distance computed according to Crawford (1984).

allow one to distinguish between the direct effect of clumping, if any, and the effect of mean dispersal distances.

With self-incompatibility (SI), autocorrelations at short and intermediate distances were always larger for nonuniform ($L \geq 200$) than for uniform ($L = 100$) spatial distributions of individuals, except in the first distance class where autocorrelations for $L = 100$ were similar to those for large L values (Figure 5). Autocorrelation values first increased with increasing L, and then stabilized for $L \approx 1000$–1500. This general pattern could be due, as above, to the decrease in mean dispersal distances, as well as to the large increase in clumping of individuals. Heterozygote patches were fewer, and thus values of SNDs for *Aa-Aa* joins were larger. This effect on Moran's *I*-values was opposed to the effects of reduced homogeneity of homozygote patches and lower distance between unlike homozygote patches. The particular pattern observed in the first distance classes (decrease with large L values), seemed to be primarily due to very high local densities leading to more mixing of heterozygotes with homozygotes at short distances, as shown by a decrease in SNDs for *Aa-Aa* joins. *X*-intercepts of SND-correlograms for *AA-AA* joins decreased with increasing L values, reflecting a reduction in the size of homozygote patches.

Introduction of SI had no effect on *I*-correlograms (Figures 1 and 5) and SND-correlograms, for $L = 100$. However for $L = 100$, patches were slightly larger with

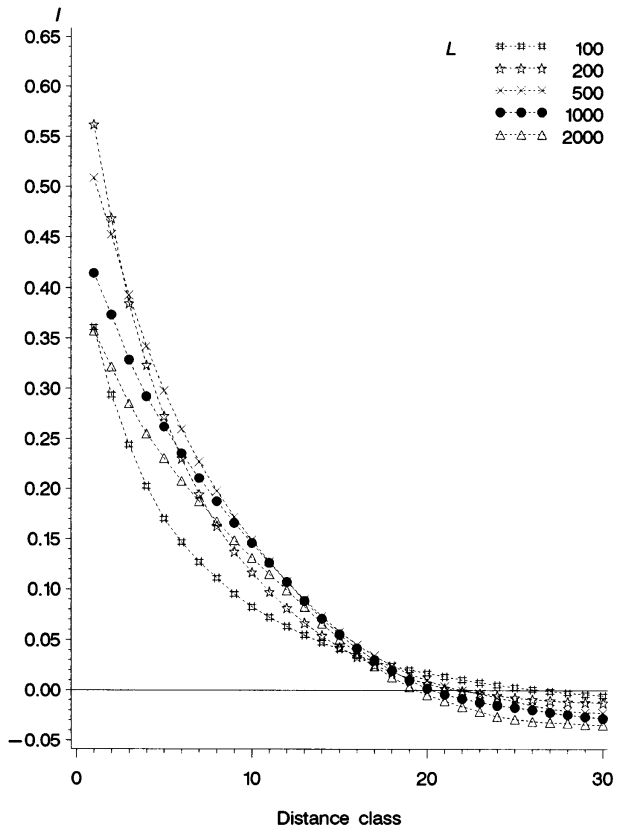

Figure 5.—Average *I*-correlograms after the last cycle for subsets *1a* ($L = 100$), *1b* ($L = 200$), *1e* ($L = 500$), *1f* ($L = 1000$), and *1h* ($L = 2000$), with SI.

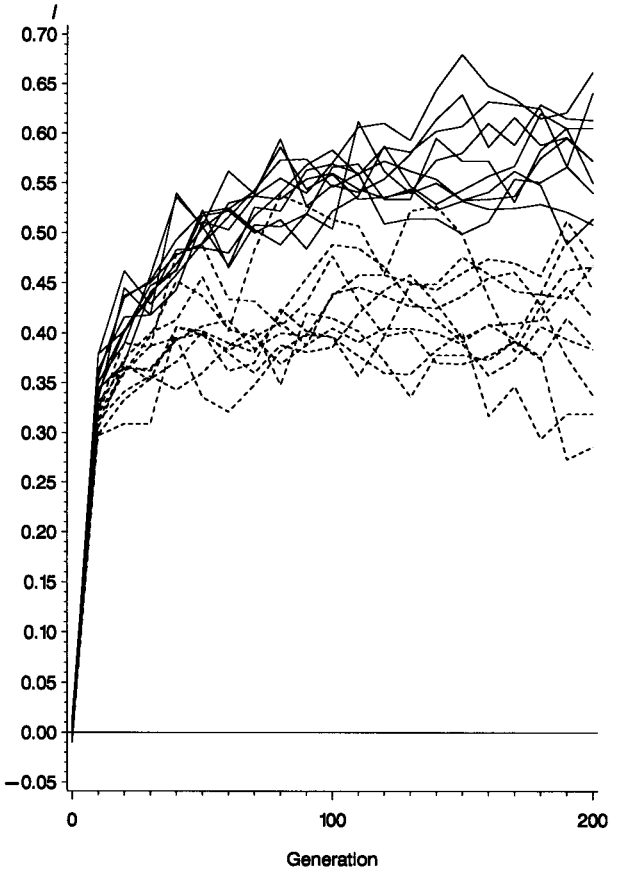

Figure 6.—Plot of *I*-statistics in the first distance class for subset *1f* ($L = 1000$; $D = P = 10$; $C = 1$; $S = 0$), as a function of time, for 10 individual replicates with SC (——) and 10 individual replicates with SI (- - - - - - -).

TABLE 3

Spatial distribution of individuals, outcrossing rate, and spatial genetic structure, within simulated populations

		Observed d_w		Simulated d_w		T_e	I (First distance class)	
Set	Subset	With SC	With SI	Min	Max	With SC	With SC	With SI
1	*a*	0.04 ± 0.000	0.04 ± 0.000	—	—	0.80 ± 0.004	0.37 ± 0.025	0.36 ± 0.032
	b	0.20 ± 0.005	0.26 ± 0.005	0.20	0.24	0.72 ± 0.005	0.56 ± 0.024	0.56 ± 0.034
	c	0.21 ± 0.008	0.31 ± 0.006	0.12	0.15	0.75 ± 0.004	0.61 ± 0.032	0.56 ± 0.040
	d	0.23 ± 0.007	0.37 ± 0.009	0.09	0.13	0.75 ± 0.006	0.63 ± 0.034	0.54 ± 0.045
	e	0.25 ± 0.006	0.38 ± 0.007	0.08	0.12	0.77 ± 0.006	0.61 ± 0.038	0.51 ± 0.051
	f	0.25 ± 0.010	0.45 ± 0.013	0.03	0.07	0.80 ± 0.006	0.60 ± 0.044	0.41 ± 0.061
	g	0.26 ± 0.010	0.48 ± 0.018	0.02	0.04	0.80 ± 0.006	0.59 ± 0.052	0.37 ± 0.081
	h	0.26 ± 0.015	0.48 ± 0.017	0.01	0.04	0.81 ± 0.007	0.59 ± 0.041	0.36 ± 0.082
2	*a*	0.04 ± 0.000	0.04 ± 0.000	—	—	1.00 ± 0.001	0.01 ± 0.005	0.01 ± 0.006
	b	0.14 ± 0.007	0.14 ± 0.008	0.08	0.12	0.92 ± 0.003	0.25 ± 0.032	0.24 ± 0.042
	ref	**0.25 ± 0.010**	**0.45 ± 0.013**	**0.03**	**0.06**	**0.80 ± 0.006**	**0.62 ± 0.056**	**0.43 ± 0.062**
	d	0.40 ± 0.011	0.61 ± 0.009	0.02	0.04	0.74 ± 0.008	0.80 ± 0.043	0.54 ± 0.068
	e	0.51 ± 0.007	0.71 ± 0.008	0.01	0.04	0.72 ± 0.008	0.87 ± 0.035	0.62 ± 0.091
3	*b*	0.10 ± 0.009	0.44 ± 0.019	0.03	0.06	0.73 ± 0.005	0.31 ± 0.051	0.14 ± 0.048
	c	0.05 ± 0.006	0.37 ± 0.029	0.03	0.06	0.70 ± 0.005	0.06 ± 0.047	0.02 ± 0.029
	d	0.04 ± 0.005	0.21 ± 0.010	0.03	0.06	0.70 ± 0.004	0.00 ± 0.052	0.00 ± 0.034
4	*b*	0.25 ± 0.010	0.26 ± 0.009	0.03	0.06	0.94 ± 0.003	0.47 ± 0.044	0.44 ± 0.067
	c	0.24 ± 0.010	0.25 ± 0.009	0.03	0.06	0.99 ± 0.001	0.17 ± 0.039	0.17 ± 0.045
	d	0.25 ± 0.010	0.24 ± 0.011	0.03	0.06	1.00 ± 0.001	0.06 ± 0.041	0.06 ± 0.034
5	*b*	0.32 ± 0.009	0.41 ± 0.010	0.03	0.06	0.84 ± 0.005	0.80 ± 0.040	0.73 ± 0.065
	c	0.33 ± 0.008	0.42 ± 0.012	0.03	0.06	0.85 ± 0.005	0.83 ± 0.053	0.75 ± 0.069
6	*b*	0.38 ± 0.018	0.47 ± 0.014	0.03	0.06	0.86 ± 0.008	0.48 ± 0.061	0.39 ± 0.062
	c	0.42 ± 0.014	0.47 ± 0.012	0.03	0.06	0.89 ± 0.006	0.47 ± 0.087	0.41 ± 0.074
	d	0.43 ± 0.018	0.46 ± 0.012	0.03	0.06	0.89 ± 0.007	0.43 ± 0.067	0.42 ± 0.094
	e	0.43 ± 0.016	0.46 ± 0.014	0.03	0.06	0.89 ± 0.006	0.43 ± 0.075	0.41 ± 0.071
	f	0.38 ± 0.017	0.45 ± 0.016	0.03	0.06	0.87 ± 0.007	0.50 ± 0.075	0.46 ± 0.079
	g	0.24 ± 0.011	0.45 ± 0.013	0.03	0.06	0.80 ± 0.006	0.62 ± 0.064	0.44 ± 0.063

Values of observed d_w (Diggle's statistic, measuring clumping), T_e (effective outcrossing rate), and I (Moran's spatial autocorrelation statistic for individual allele frequencies, measuring spatial genetic structure) in the first distance class are means over replicates ± SD, for the neutral locus after the last cycle. Values of simulated d_w are the minimum and maximum values of the simulated distribution of d_w (1000 replicate random configurations).

SI than with SC, as shown by values of *X*-intercepts. For larger L values, Moran's *I*-statistics at short and intermediate distances were lower with SI than with SC. This difference was greater for $L \geq 1000$ than for $L < 1000$. Differences in outcrossing rates (Figure 4) could not explain these differences in spatial autocorrelations. Conversely, differences in the level of clumping (as measured by d_w, Figure 4) and mean pollen dispersal distances (Table 2) were linked to these differences in spatial autocorrelations. With SI, clumps were denser, larger, and more isolated from each other than with SC, because isolated individuals could not reproduce. This increase in local density yielded more mixing of homozygotes with heterozygotes at short and intermediate distances, which could partly explain the decrease in autocorrelations obtained with the introduction of SI, in combination with the larger effective pollen dispersal distances.

Evolution in time from random initial state: In all cases for which we recorded the evolution of spatial structure in time (Table 1), Moran's *I*-statistics (Figure 6) and d_w initially increased rapidly, and then considerably more slowly. Although in some cases they were still increasing after $100 \times C$ cycles (*e.g.*, with selection), their mean value always varied little between cycles $100 \times C$ and $200 \times C$. Therefore the time required for spatial genetic structure to stabilize was similar in nonuniformly and uniformly distributed populations. Fixation indices and effective outcrossing rates showed a similar evolution in time (*i.e.*, rapid increase or decrease in earlier cycles, and little change in later cycles). For genetic variables, variability between replicates tended to increase with time (Figure 6). In some cases, it was relatively high after the last cycle (Table 3).

Effect of global density (simulation set *2*): Moran's *I*-statistics always increased with decreasing global density (Figure 7). This resulted from an increase in SNDs for like homozygote joins, an increase in SNDs for *Aa-Aa* joins, and a decrease in SNDs for *AA-aa* joins. This increase in spatial genetic structure was attributable primarily to the decrease in mean relative dispersal distances (measured in units of average interplant abso-

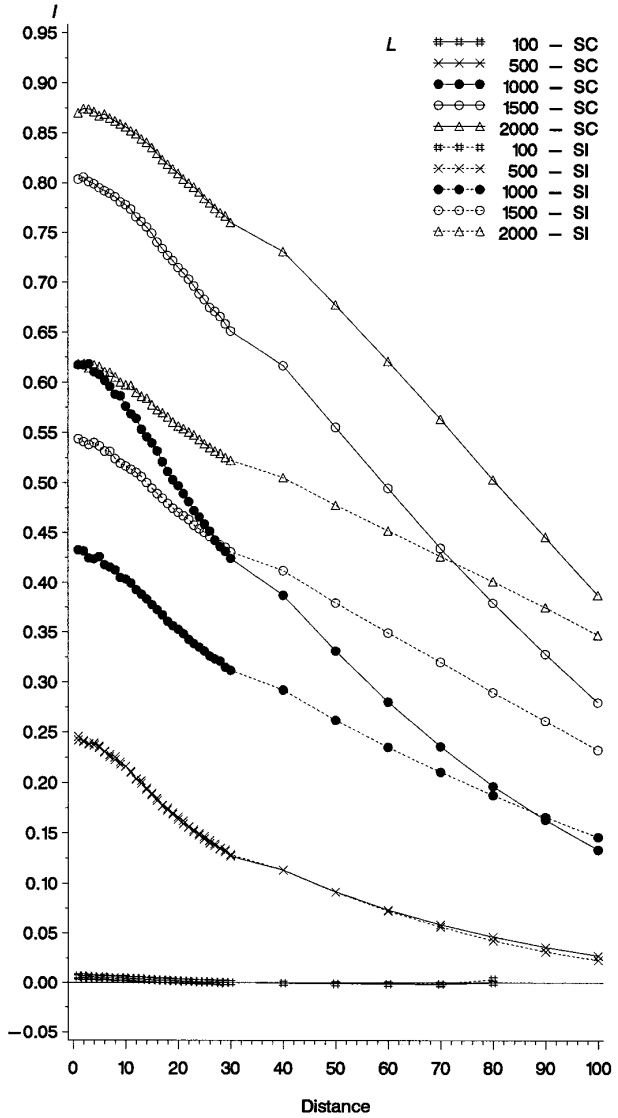

Figure 7.—Average *I*-correlograms after the last cycle for the reference subset ($L = 1000$) and subsets *2a* ($L = 100$), *2b* ($L = 500$), *2d* ($L = 1500$), and *2e* ($L = 2000$), with SC or SI.

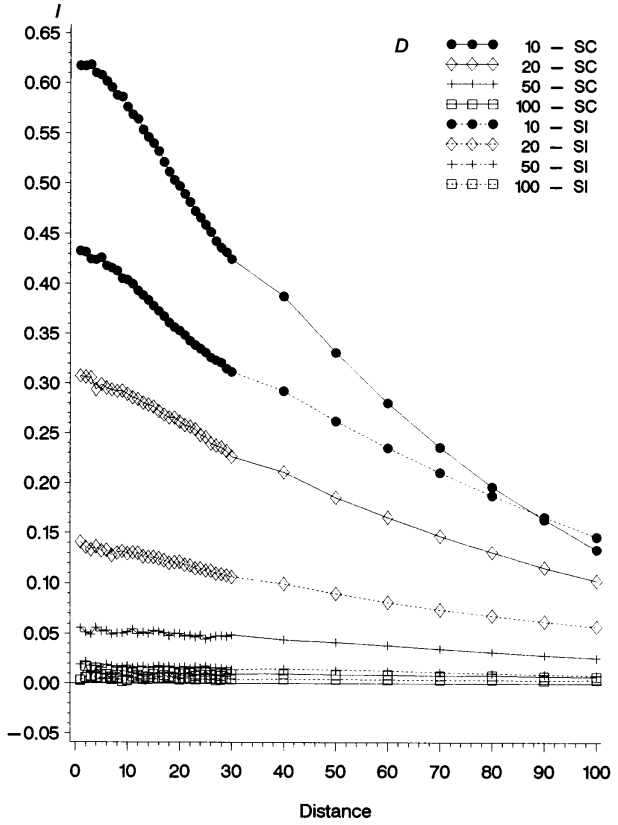

Figure 8.—Average *I*-correlograms after the last cycle for the reference subset ($D = 10$) and subsets *3b* ($D = 20$), *3c* ($D = 50$), and *3d* ($D = 100$), with SC or SI.

lute distances). To test whether the growing isolation of clumps, which resulted in increased d_w values, had any additional effect, we computed relative dispersal distances for the first nine replicates of subsets *1a* ($L = 100$, $D = P = 1$) and *1f* ($L = 1000$, $D = P = 10$) with SC. Simulations of these two subsets produced very different spatial distributions of individuals (uniform and clumped, respectively), whereas we expected them to yield very similar relative dispersal distances. However, mean seed and pollen relative dispersal distances were slightly higher for subset *1a* (0.021 and 0.021, respectively) than for subset *1f* (0.018 and 0.014, respectively), and it was not possible to know whether this difference was sufficient to explain the large discrepancy observed in Moran's *I*-values (Table 3). So we could not make any conclusion about the existence of any additional effect of clumping in this case.

Note that on Figures 7 to 9, each correlogram consists of two different parts, corresponding to the two widths that were used for distance classes (one grid unit for small distances and 10 grid units for larger distances). These two parts cannot be compared within each correlogram because Moran's *I*-value is dependent on the width of distance class. But each part can be compared among different correlograms.

The increase in Moran's *I*-statistics with decreasing global density was less pronounced with SI than with SC for large *L* values (Figure 7), which resulted from lower SNDs for *Aa-Aa* and *AA-Aa* joins, and higher SNDs for *AA-aa* joins. This could be due to differences in relative dispersal distances, clumping, and/or outcrossing. With SC, most clumps were small and consisted of one homozygote patch, and there was substantial selfing. With SI, owing to larger, denser, and more isolated clumps, together with the absence of selfing, heterozygosity and d_w were higher, and considerable local mixing of genotypes was maintained even at very low global densities.

Effects of seed and pollen dispersal (simulation sets *3* and *4*): Moran's *I*-values always decreased with increasing *D* (maximum seed dispersal distance) (Figure 8). They were always lower with SI than with SC, due to a smaller excess of *Aa-Aa* joins and a smaller deficit of

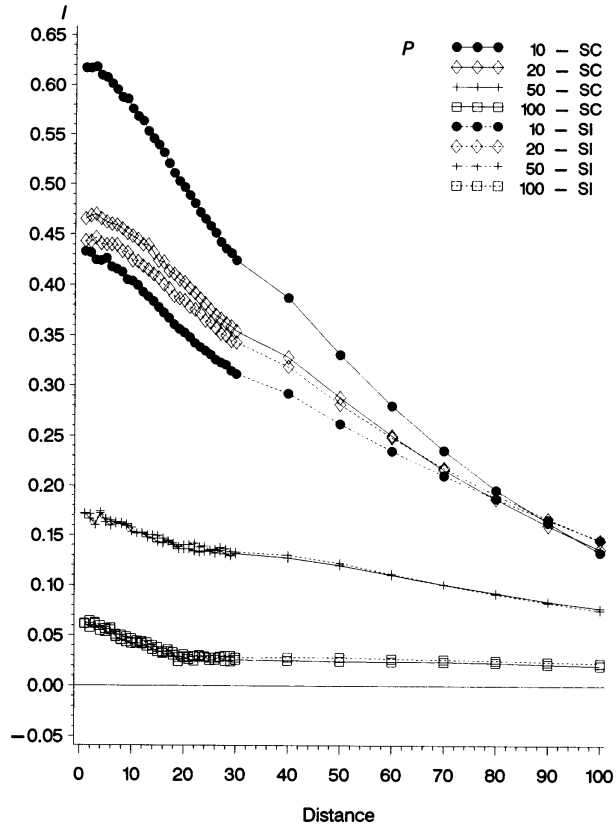

Figure 9.—Average *I*-correlograms after the last cycle for the reference subset ($P = 10$) and subsets *4b* ($P = 20$), *4c* ($P = 50$), and *4d* ($P = 100$), with SC or SI.

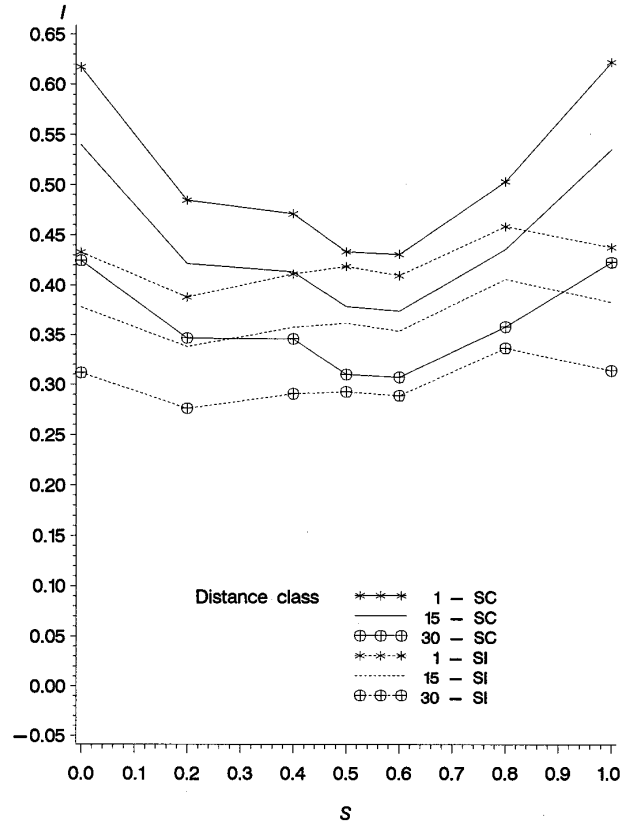

Figure 10.—Average *I*-statistics in three distance classes after the last cycle for all subsets of simulation set *5*, with SC or SI.

AA-aa joins. Both higher local densities and the absence of selfing seemed to be responsible for these differences. With SC, individuals were significantly clumped for $D \leq 20$, and their distribution became random for $D \geq 50$ (Table 3). With SI, d_w values also decreased with increasing *D*, but spatial distribution of individuals remained highly clumped even for large *D* values; all individuals of the population were then grouped within a single large clump spread over approximately half the whole grid area, because isolated individuals did not reproduce.

Moran's *I*-values decreased with increasing *P* (maximum pollination distance) for $P \geq 20$ (Figure 9). In these cases, *I*-values, spatial distribution of individuals, and outcrossing rates were similar with SI and with SC, which was due to the fact that pollination distances were large enough for spatially isolated individuals not to be reproductively isolated. For $P = 10$ (reference subset), *I*-values were lower with SI than with SC and d_w values were higher, because spatially isolated individuals were reproductively isolated. In this case, clumping developed with SI and there was substantial selfing with SC ($T_e \approx 0.8$).

Comparing the effects of *D* and *P* (Figures 8 and 9) shows that a given increase in *D* had a stronger decreasing effect on Moran's *I*-statistics than the same increase in *P*. This difference presumably resulted from a combination of two causes: (1) the difference in ploidy level between seeds and pollen (and thus the difference in the number of genes dispersed during each dispersal event), and (2) the difference in relative dispersal distances (when *D* was increased, effective pollen dispersal distances were expected to be larger and interplant distances lower, whereas when *P* was increased, both effective seed dispersal distances and interplant distances were expected not to change). The opposed effects of differences in local density on Moran's *I* were weaker because when *D* was increased it decreased, whereas when *P* was increased, it essentially did not change. The effect of outcrossing was also opposed to this ploidy and relative dispersal effects, leading to more difference between the effects of *D* and *P* with SI than with SC, since outcrossing rate decreased with increasing *D*, whereas it increased with increasing *P*.

Effect of generation overlapping (simulation set *5*): Moran's *I*-statistics always increased when *C* (generation overlapping) was increased (Table 3). This resulted from an increase in SNDs for like homozygote joins and *Aa-Aa* joins, and a decrease in SNDs for *AA-aa* joins. However, the largest difference was found between the reference subset (no overlap) and *5b* or *5c* (some overlap), and the difference between subsets *5b* ($C = 5$) and *5c* ($C = 10$) was much smaller. This effect

Figure 11.—Spatial distribution of individual genotypes in a subpart (of size 500 grid units × 700 grid units) of the total grid after the last cycle in one replicate of subset *6d* (L = 1000; $D = P = 10$; $C = 1$; $S = 0.5$), with SC; −, o, |, for genotype *AA*, *Aa*, and *aa*, respectively.

of overlapping was presumably due to the increased kinship between mates, which is expected to result from the combination of restricted gene flow and mating of individuals with their ancestors. Specific computations are required to test for this hypothesis. However, mean values of the probability that, at the neutral locus, two genes randomly drawn without replacement among all the genes of the population were identical by descent, were higher with overlap than without (0.013, 0.032, and 0.037 for the reference subset and subsets *5b*, *5c*, respectively), which strongly supports this interpretation. *I*-correlograms were always lower with SI than with SC, which was probably due to the absence of selfing, higher local density, and lower relative dispersal distances.

Effect of selection (simulation set *6*): In this section, all genetic values reported are for the unlinked neutral locus, unless otherwise stated. With SC, Moran's *I*-statistics decreased for S (selection coefficient) from 0 to 0.6, and then increased again for S up to 1; *I*-values were similar for $S = 1$ and $S = 0$ (Figure 10). These changes were related to changes in d_w values (Table 3) and in local density (as illustrated by the differences between Figure 3 and Figure 11), which were both highest for intermediate S values. These variations in clumping resulted in more spatial mixing of genotypes and larger outcrossing rates for intermediate than for extreme S values, which could explain the increase in heterozygosity and the lower spatial genetic autocorrelations. In additional simulations with uniform distribution of individuals (with SC, $L = 100$, $D = P = C = 1$, $S = 0.5$, 10 replicates), the average Moran's I value in the first distance class was 0.36 (SD 0.028). This result confirms that the effect of selection on spatial genetic structure was due to variations in clumping. In turn, these changes in clumping could result from the fact that the mean effective number of offspring produced by homozygotes at the selected locus was lower than the mean number of offspring produced by heterozygotes (data not shown). But this hypothesis has not been tested and there might be alternative explanations. Since gene flow was restricted enough to allow for the development of homozygote patches, mating between like genotypes was enhanced. Therefore, at the selected locus, heterozygotes on average produced a larger proportion of heterozygote offspring (*i.e.*, offspring more likely to survive in the presence of overdominance selection) than did homozygotes. This relative reproductive advantage of heterozygotes over homozygotes was further increased by a significant part of selfing among matings. When the proportion of heterozygotes in the population was intermediate (*i.e.*, for $S \sim 0.5$), heterozygotes at the selected locus would have the highest relative reproductive advantage, which would lead to clumping of surviving offspring around the locations of mothers heterozygous at the selected locus. However, the global variance in the number of surviving offspring was only slightly, and not significantly, larger for intermediate than for extreme S values (results not shown). When $S = 0$ (no selection) or $S = 1$ (100% heterozygotes at the selected locus), heterozygotes at the selected locus had no particular advantage, which could explain why there was less clumping of individuals.

With SI, differences in *I*-correlograms between subsets of set *6* were much lower than with SC (Figure 10). This is consistent with the above interpretation, since with SI, heterozygotes at the selected locus have less reproductive advantage, because of both lower spatial autocorrelations and the absence of selfing. This was confirmed by the absence of systematic difference between the mean number of offspring produced by homozygotes and heterozygotes at the selected locus (data not shown).

The case of study with only one selected locus and the widest possible range for S values (from 0 to 1) was useful to help understand the mechanisms through which spatial genetic structure at neutral loci may be modified by the presence of selection at other, inde-

pendent loci. However, the complementary simulations carried out with 10 selected loci and selection coefficients from 0.01 to 0.20, represented situations far more likely in real populations. For selection coefficients ≥0.02, d_w statistics had mean values similar to or greater than the highest values obtained with only one selected locus (*i.e.*, 0.43 for $S = 0.6$), and *I*-correlograms were lower.

Relationships between clumping, local density, and spatial genetic structure: For simulation sets *2–6*, the relationship between spatial genetic structure and spatial distribution of individuals (as measured by Diggle's d_w) was different among sets (Figure 12, A and B). In some cases (for variations in *S* with SC, or introduction of SI), increased clumping was related to lower spatial autocorrelations. This pattern presumably resulted from the increase in both outcrossing rate and local density, which led to an increase in heterozygosity and spatial mixing of genotypes. Conversely, in other cases (for variations in *C* with SC, variations in *D*, or variations in global density), increased clumping was related to higher spatial autocorrelations. This common pattern, however, had three different explanations. First, the changes in kinship directly due to variations in generation overlapping (*C*) were strong enough to counter and overcome the aforementioned changes in clumping. Second, reduced maximum seed dispersal distances (*D*) resulted in less spatial mixing of genotypes and therefore more local genetic differentiation, in spite of higher local densities and clumping. Lastly, when global density was decreased, d_w increased, but this reflected variations in clumping different from those described above (smaller and more numerous clumps developed with lower local density, and thus a lower outcrossing rate, instead of fewer and larger clumps with higher local density). Meanwhile, relative dispersal distances decreased, leading to less mixing of genotypes and more local fixation of alleles.

DISCUSSION

Nonuniform spatial distribution of individuals: The present study is the first one to give a quantitative description of fine-scale spatial organization of genotypes in theoretical plant populations with nonuniform spatial distribution of individuals. Studying cases of extreme aggregation has allowed us to bring to light some potential effects of clumping on spatial genetic structure. Our main findings are the following: (1) the presence of strong clumping leads to a slight increase in spatial autocorrelations compared with uniformly distributed populations, independently of the effect of dispersal distances; (2) global density, seed and pollen dispersal, overlap of generations, overdominance selection at unlinked loci, and self-incompatibility may all influence aggregation; and (3) the effect of each of

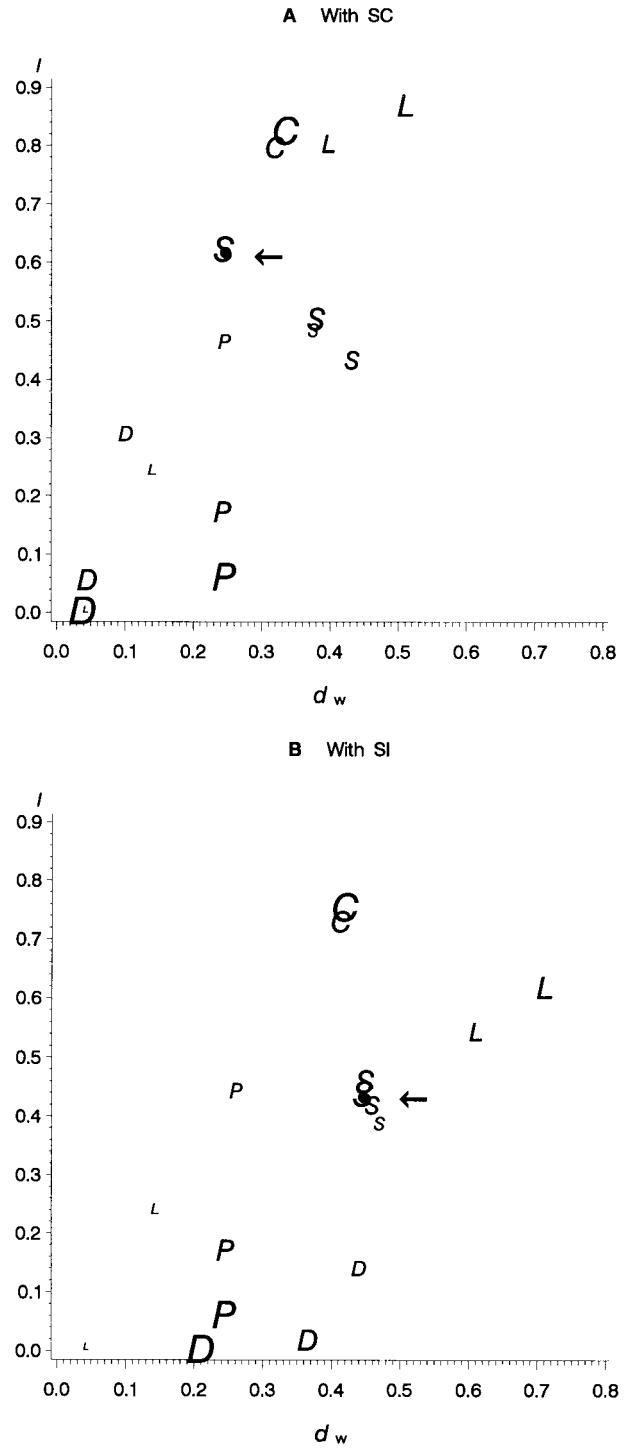

Figure 12.—Plot of average *I*-statistics in the first distance class as a function of average d_w statistics after the last cycle for all simulation subsets, except those of set *1* and subsets *6c* ($S = 0.4$) and *6e* ($S = 0.6$), which were omitted for clarity of the figure, with SC (A) or with SI (B). ●, reference subset (marked by an arrow); for all other subsets, the size of the letter used in the plot is proportional to the value of the input variable, which differs from the reference (see Table 1).

these variables on spatial genetic structure, except maybe global density, may be affected by clumping.

Before discussing these results in more detail, it is worth examining the appropriateness and expected consequences of three assumptions that might be considered as major limits to our model: the absence of microenvironmental heterogeneity, the absence of density-dependent survival, and the uniform distribution of dispersal distances.

Our simulations with restricted dispersal produced very clumped spatial distributions of individuals, sometimes with large clumps separated by large empty areas. Such extreme configurations can be found in some species, in particular in tropical forest tree species (*e.g.*, d_w values > 0.20 were observed for several tropical forest tree species in French Guiana; CIRAD-Forêt, unpublished data). But then, spatial aggregation may be due primarily to microenvironmental heterogeneity of survival rates rather than to restricted dispersal. In such cases, metapopulation models would certainly be more appropriate. Our model is very different in that clumps can change sizes, shapes, and locations over time. Nevertheless, in many cases, there is as yet no evidence that habitat characteristics have any influence on spatial distribution of individuals. In a neotropical forest, Hubbell and Foster (1986) showed that 119 out of 239 woody plant species were indifferent to topography (for presence/absence), and Welden *et al.* (1991) showed that 123 out of 148 species were indifferent to canopy height (for sapling survival).

Evidence of density-dependent mortality in natural plant populations is scarce. Condit *et al.* (1992) showed that 65 out of 80 species in a neotropical forest exhibited no significant reduction in sapling recruitment probability when close to rather than far away from conspecific adults. One of Antonovics and Levin's (1980) conclusions, after reviewing available evidence of density-dependent regulation in natural plant populations, was that it is "sparse, often circumstantial, and primarily from populations of dominant or abundant species." Furthermore, authors working on this topic have more often been concerned with "yield" than with survival up to the age of reproductive maturity. In our simulations, individuals were located at the intersection points of a grid, which might be considered as a minimum density-dependent regulation accounting for the physical space occupied by each adult plant. The effect of introducing additional density-dependent mortality would probably be to reduce both aggregation and spatial genetic structure through an increase in effective dispersal distances, and thus to modify accordingly the effects of the input variables studied here on spatial genetic structure. It would yield more realistic distributions of individuals, lying somewhere in a continuum of distributions, of which uniform and highly clumped are the extremes.

We assumed uniform distributions of pollen and seed dispersal distances, both to save computer time and for consistency with previous simulation studies. For highly clumped distributions such as those obtained in our simulations, one important consequence of this assumption, together with the assumption that pollinator movement is independent of local plant density, is that unoccupied habitat between patches is an absolute barrier to gene flow when gaps are larger than maximum dispersal distances. This is not supported by the empirical evidence available to date. The consequence of introducing a low proportion of long-distance dispersal, in addition to high amounts of short-distance dispersal would certainly be comparable to the effect of random long-distance immigration as modelled by Epperson (1990). Immigration only slightly reduced spatial autocorrelations, compared to dispersal or selection. Furthermore, as argued by Epperson and Li (1997), leptokurtosis does not have much effect on Wright's neighborhood size, and thus on F_{ST}, unless the kurtosis parameter is very large. However, relationships between F_{ST} and spatial autocorrelations are complex (Barbujani 1987; Epperson and Li 1996), and little is known of the effect of platykurtosis. More direct evidence of the effect of the dispersal regime on Moran's *I*-values is needed, and could be obtained by extending the comparisons of Ohsawa *et al.* (1993) to larger dispersal values. Nevertheless, our simulations presumably exaggerate spatial genetic structure compared to that of real populations, since more leptokurtotic dispersal could be expected to produce more clumping and higher local densities.

Therefore, even though our model may lack realism because it is based on very simple assumptions about dispersal and the spatial distribution of individuals, the study of its behavior in a few cases of extreme clumping has provided results very useful for a better understanding of the distributions of genotypes observed in real plant populations.

First, we found that allowing for strong clumping leads to higher values of spatial genetic autocorrelations than in uniformly distributed populations. Simulations in set *1* do not allow us to distinguish between the direct effect of clumping, if any, and the effect of mean dispersal distances, since the only way to obtain clumped distributions is to change dispersal distances. However, comparison with the results of Epperson (1995b) strongly suggests that clumping has some effect of its own, in addition to the effect of dispersal. In both his and our simulations, the variance of the parent-offspring axial dispersal distance (σ^2_{axial}) is not proportional to Wright's neighborhood area, because axial dispersal distances are not normally distributed. But in our simulations, for subsets *1b* to *1h* with SC, σ^2_{axial} values were between 0.34 and 0.37, and *I*-values in the first distance class were all higher than 0.5, whereas in Epperson's (1995b) simulation set *1*, I_c was lower than 0.5 for $\sigma^2_{axial} = N_e/4\pi = 0.33$. Moreover, since local

density is higher in our simulations, the neighborhood size is certainly higher too. In our simulation set *1* with SI, Moran's *I*-values were not a monotonous function of mean dispersal distances, which is additional evidence that these distances are not the only factor influencing spatial genetic structure in nonuniformly distributed populations. In addition to the higher local densities, the effect of clumping seems to be mainly due to the fact that under nonuniform distributions, homozygote patches can be separated either by empty grid areas or by heterozygote patches, whereas with a uniform distribution they can be separated only by heterozygote patches.

For given values of maximal dispersal distances and global density, both mean dispersal distances and Moran's *I*-values stabilized when a continuous distribution of individuals was approached (for L around 1000 in simulations of set *1*). We chose this particular case as the reference case to study the separate effects of a few demographic (global density, generation overlapping), reproductive (gene dispersal, self-incompatibility), and genetic (selection) factors, on the spatial distribution of individuals and genotypes. Note that we mainly used small seed dispersal distances relative to global density. Thus we obtained highly clumped distributions, the study of which is useful to understand better the potential determinants of aggregation and spatial genetic structure.

Second, our simulations have shown that aggregation can be influenced by self-compatibility, pollination distances, generation overlapping, and selection, in addition to the more straightforward effects of seed dispersal distances and global density. These previously undescribed mechanisms are potentially active in all plant species. Thus further studies are needed to assess their effective importance in real populations.

Third, we found that these changes in spatial distributions of individuals in turn have major consequences on the way dispersal, generation overlapping, selection, and self-compatibility influence spatial genetic structure, as will now be discussed. In all cases, part of the effect of clumping is likely to be due to changes in relative dispersal distances, since clumping is expected to reduce mean interplant distances. However, our simulations did not allow us to dissociate these effects of relative dispersal from hypothetical specific effects of clumping, since we did not systematically record relative dispersal distances.

Relationships between clumping, local density, and spatial genetic structure: We have found no systematic relationship between variations in clumping as measured by d_w, and spatial genetic structure, in the few particular cases studied here. There are two reasons for this. First, two input variables (the maximum seed dispersal distance and the amount of generation overlapping) had direct effects on spatial genetic structure that were opposed to and stronger than their indirect effect through variations in clumping. Second, d_w statistics measure the extent to which distributions deviate from randomness, but they are not sufficient to entirely characterize clumping. The same d_w value can be obtained in situations showing very different local densities. In the present study, local density (of which outcrossing rate before selection is a direct measure here) was positively related with d_w within all simulation sets except the one involving variations in global density. Therefore, clumping is better described with both the d_w statistic and outcrossing before selection than with the d_w statistic alone. These results imply that the level of clumping as measured by the d_w statistic only cannot directly be used to predict spatial genetic structure.

Comparative effects of both gene dispersal modes: Previous simulation studies showed that increasing pollination distances with little or no seed dispersal, or simultaneously increasing seed and pollen dispersal distances (with the same neighborhood sizes), both led to reduced fine-scale genetic structure (*e.g.*, Sokal *et al.* 1989; Epperson 1990, 1995a,b). Our study is the first one to report the effect of dispersing seed farther than pollen. It shows that seed dispersal affects spatial genetic structure to a larger extent than pollen dispersal, probably because of differences in relative dispersal and in ploidy between seeds and pollen. It is not clear whether the difference between the effects of D and P would be larger or smaller with a uniform than with a nonuniform distribution of individuals, since the ploidy effect is opposed to the effect of variations in clumping, but at the same time, there would be no differences in relative dispersal distances under a uniform distribution. Therefore, it appears necessary to allow for separate variations of these two modes of gene dispersal in theoretical studies of spatial genetic structure with isolation by distance, whatever the spatial distribution of individuals.

Effects of generation overlapping and selection: Our results suggest that with nonuniform distributions of individuals, the effect of generation overlapping on spatial genetic structure was reduced by variations in clumping. Therefore, we would expect this effect to be even stronger with uniform distributions of individuals. But the effect of generation overlapping has scarcely been studied in such cases. The simulations of Bos and Van Der Haring (1988) revealed no effect of generation overlapping on spatial genetic structure in most cases, probably because pollen flow was not restricted enough in their simulations. But this difference between their results and ours might also be due to differences in the age structure of the population. They used random replacement of individuals, which presumably led to a decreasing age structure and thus to less overlap of generations. As for the results of Berg and Hamrick (1995), they cannot be compared with ours because these authors included both reproductive and nonreproductive individuals in their simulated populations.

Spatial genetic structure at a neutral locus when selection occurs at independent loci has never been described before. Both Sokal *et al.* (1989) and Epperson (1990) studied the effects of selection. But they only described spatial genetic structure at the selected locus itself, and they did not report the indirect effects of this selection at other loci. Our simulations showed that the occurrence of symmetric overdominance selection at a few loci might increase heterozygosity and reduce spatial genetic structure at all loci, not only at directly selected or closely linked ones. This effect seems to be entirely due to changes in outcrossing rate and local density. It does not involve any direct relationship between heterozygosity at selected loci (and thus fitness) and heterozygosity over the whole genome within each individual, as would be expected if selection against inbred individuals were involved (Charlesworth 1991). Chi-square tests of independency between heterozygosity at the selected locus and heterozygosity at the neutral locus were significant ($\alpha = 5\%$) for 49, 47, 37, 32, 21, and 4 replicates out of 50, for the reference subset and subsets *6b* to *6f* with SC, respectively. Therefore it is not an effect at the individual level, but rather at the population level. As a consequence, this effect is expected to be found only in nonuniformly distributed populations with spatial genetic structure and mixed mating, as supported by the few complementary simulations we carried out with selection in a uniformly distributed population. Moreover, only very low selection coefficients (*e.g.*, 0.02) at a few (*e.g.*, 10) selected loci were sufficient to obtain large changes in the spatial features of the population. Even though very few data are available to date, overdominance selection events might not be scarce in natural populations, especially for long-lived perennials [*e.g.*, Bush and Smouse (1991) have found evidence of overdominance effects over the whole life cycle at one allozyme locus in *Pinus taeda*]. However, our results bring to light the potential effects of only one among all possible selection mechanisms, on spatial genetic structure. They will have to be completed by examining alternative hypotheses such as the mutation-dominance model with several values of the dominance coefficient and/or additive combination of effects between selected loci for fitness.

Consequences of self-incompatibility and outcrossing determinism: Our results stress the importance of knowing whether a species is self-compatible or not in order to understand its fine-scale spatial genetic structure. The introduction of self-incompatibility (when defined as the inability to self) can induce very large decreases in spatial genetic structure within populations with nonuniform distributions, whereas it has no effect in uniformly distributed populations. This difference was not directly due to the absence of selfing itself, but to the higher clumping and local density that indirectly resulted from the absence of selfing. Even though the high levels of clumping reached in some of our simulations are not very realistic, as discussed above, these simulations help understand the extent to which the ability of isolated individuals to reproduce through selfing may influence spatial genetic structure in nonuniformly distributed populations. Introducing a gametophytic self-incompatibility genetic system (SI locus with several SI alleles) instead of merely prohibiting selfing is expected to further reinforce the effect of self-incompatibility on spatial genetic structure through a reduction in biparental inbreeding. In the simulations of Bos and Van Der Haring (1988), the introduction of such a system hardly affected local differentiation. However, it might have a stronger effect with a nonuniform distribution of individuals, through variations in clumping.

The effect of overdominance selection on the spatial distribution of individuals and genotypes which was found here is dependent upon both some selfing among matings and a large influence of local density on outcrossing rate. Therefore, it will be particularly important to assess the extent to which the amount of outcrossing is determined by local density in self-compatible species. Relationships between outcrossing rate and density of mature individuals have been found in natural populations for a few species [*e.g.*, Murawski and Hamrick (1991) in tropical trees], but such relationships might be difficult to detect if inbred individuals are selected against (Van Treuren *et al.* 1994).

Consequences for interpretation of *I*-correlograms observed in real populations: Epperson (1995b) has argued that isolation by distance might be sufficient to explain most of the low values of spatial autocorrelations for individual allele frequencies that have been found in many natural plant populations studied so far [see reviews by Heywood (1991) and Epperson (1993)]. Our results suggest that *I*-values are likely to be slightly higher in real populations than those predicted by Epperson (1995b), because of nonuniform spatial distribution of individuals (and generation overlapping in perennials). Nevertheless, even for the cases of extreme clumping considered here, the difference is not large enough to affect his general conclusion that in many cases isolation-by-distance is sufficient to explain the low *I*-values found in real populations, without the need for selectionist arguments.

In conclusion, our results show the potential benefits of enlarging the framework for simulation studies of fine-scale spatial genetic structure, compared to what has been done so far, by taking into account a nonuniform rather than a uniform spatial distribution of individuals. The model presented here is based on very simple assumptions and thus it has many limits. Nevertheless, the study of its behavior in a few particular cases of extreme aggregation has brought to light some mechanisms, never described before, through which variations in clumping might influence spatial genetic structure within plant populations (*e.g.*, the effects of self-incompatibility or overdominance selection

on spatial genetic structure). In particular, our results stress the need to assess the real importance of self-incompatibility, generation overlapping, and overdominance selection in natural populations, as well as the effective influence of local density on outcrossing rate, and the extent of differences between seed and pollen dispersal distances. Further study of the present model is required (1) to determine the conditions under which each potential mechanism described here will effectively play a significant role in shaping fine-scale spatial genetic structure, and (2) to assess the exact contribution of relative dispersal distances to the effect of clumping on genetic structure.

The authors thank B. K. Epperson, J. S. Heywood, and an anonymous reviewer, for constructive comments on the manuscript, as well as D. Doligez for helpful advice on the implementation of simulations, and L. Houde for his help with the use of spatial statistics. This investigation was supported by a grant from the French Ministère de la Recherche to A.D. and funds from CIRAD-Forêt.

LITERATURE CITED

Antonovics, J., and D. A. Levin, 1980 The ecological and genetic consequences of density-dependent regulation in plants. Annu. Rev. Ecol. Syst. **11:** 411–452.

Armesto, J. J., J. D. Mitchell and C. Villagran, 1986 A comparison of spatial patterns of trees in some tropical and temperate forests. Biotropica **18:** 1–11.

Berg, E. E., and J. L. Hamrick, 1995 Fine-scale genetic structure of a turkey oak forest. Evolution **49:** 110–120.

Barbujani, G., 1987 Autocorrelation of gene frequencies under isolation by distance. Genetics **117:** 777–782.

Bos, M., and E. Van Der Haring, 1988 Gene flow in *Plantago.* II. Gene flow pattern and population structure. A simulation study. Heredity **61:** 1–11.

Bush, R. M., and P. E. Smouse, 1991 The impact of electrophoretic genotype on life history traits in *Pinus taeda.* Evolution **45:** 481–498.

Charlesworth, D., 1991 The apparent selection on neutral marker loci in partially inbreeding populations. Genet. Res. **57:** 159–175.

Cliff, A. D., and J. K. Ord, 1981 *Spatial processes.* Pion, London.

Condit, R., S. P. Hubbell and R. B. Foster, 1992 Recruitment near conspecific adults and the maintenance of tree and shrub diversity in a neotropical forest. Am. Nat. **140:** 261–286.

Crawford, T. J., 1984 The estimation of neighbourhood parameters for plant populations. Heredity **52:** 273–283.

Epperson, B. K., 1990 Spatial autocorrelation of genotypes under directional selection. Genetics **124:** 757–771.

Epperson, B. K., 1993 Recent advances in correlation studies of spatial patterns of genetic variation. Evol. Biol. **27:** 95–155.

Epperson, B. K., 1995a Spatial distributions of genotypes under isolation by distance. Genetics **140:** 1431–1440.

Epperson, B. K., 1995b Fine-scale spatial structure: correlations for individual genotypes differ from those for local gene frequencies. Evolution **49:** 1022–1026.

Epperson, B. K., and T. Li, 1996 Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. Proc. Natl. Acad. Sci. USA **93:** 10528–10532.

Epperson, B. K., and T. Li, 1997 Gene dispersal and spatial genetic structure. Evolution (in press).

Heywood, J. S., 1991 Spatial analysis of genetic variation in plant populations. Annu. Rev. Ecol. Syst. **22:** 335–355.

Hubbell, S. P., 1979 Tree dispersion, abundance, and diversity in a tropical dry forest. Science **203:** 1299–1309.

Hubbell, S. P., and R. B. Foster, 1986 Commonness and rarity in neotropical forest: implications for tropical tree conservation, pp. 205–231 in *Conservation Biology: The Science of Scarcity and Diversity,* edited by M. E. Soulé. Sinauer Associates, Sunderland, MA.

Kawata, M., 1995 Effective population size in a continuously distributed population. Evolution **49:** 1046–1054.

Malécot G., 1948 *Les mathématiques de l'hérédité.* Masson, Paris.

Murawski, D. A., and J. L. Hamrick, 1991 The effect of the density of flowering individuals on the mating systems of nine tropical tree species. Heredity **67:** 167–174.

Ohsawa, R., N. Furuya and Y. Ukai, 1993 Effect of spatially restricted pollen flow on spatial genetic structure of an animal-pollinated allogamous plant population. Heredity **71:** 64–73.

Rohlf, F. J., and G. D. Schnell, 1971 An investigation of the isolation-by-distance model. Am. Nat. **105:** 295–324.

Ronfort, J., and D. Couvet, 1995 A stochastic model of selection on selfing rates in structured populations. Genet. Res. **65:** 209–222.

Sedgewick, R., 1988 *Algorithms.* Addison-Wesley, Reading, MA.

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

Sokal, R. R., and D. E. Wartenberg 1983 A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics **105:** 219–237.

Sokal, R. R., G. M. Jacquez and M. C. Wooten, 1989 Spatial autocorrelation analysis of migration and selection. Genetics **121:** 845–855.

Turner, M. E., J. C. Stephens and W. W. Anderson, 1982 Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. Proc. Natl. Acad. Sci. USA **79:** 203–207.

Upton, G. J. G., and B. Fingleton, 1985 *Spatial Data Analysis by Example, Volume I: Point Pattern and Quantitative Data.* John Wiley & Sons, Chichester, U.K.

Van Treuren, R., R. Bulsma, N. J. Ouborg and M. M. Kwak, 1994 Relationships between plant density, outcrossing rates and seed set in natural and experimental populations of Scabiosa columbaria. J. Evol. Biol. **7:** 287–302.

Welden, C., S. W. Hewett, S. P. Hubbell and R. B. Foster 1991 Sapling survival, growth, and recruitment: relationship to canopy height in a neotropical forest. Ecology **72:** 35-50.

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

Communicating editor: B. S. Weir