

Plant dispersal, neighbourhood size and isolation by distance

BRYAN K. EPPERSON

126 Natural Resources Building, Michigan State University, East Lansing, MI 48824, USA

Abstract

A theoretical relationship between isolation by distance or spatial genetic structure (SGS) and seed and pollen dispersal is tested using extensive spatial-temporal simulations. Although for animals Wright's neighbourhood size $N_e = 4\pi\sigma_t^2$ has been ascertained also, where σ_t^2 is the axial variance of distances between parents and offspring, and it was recently confirmed that $N_e = 4\pi(\sigma_f^2 + \sigma_m^2)/2$ when dispersal of females and males differ, the situation for plants had not been established. This article shows that for a very wide range of conditions, neighbourhood size defined by Crawford's formula $N_e = 4\pi(\sigma_s^2 + \sigma_p^2/2)$ fully determines SGS, even when dispersal variances of seed (σ_s^2) and pollen (σ_p^2) differ strongly. Further, self-fertilization with rate s acts as zero-distance pollen dispersal, and $N_e = 4\pi[\sigma_s^2 + \sigma_p^2(1-s)/2]$ fully determines SGS, for most cases where there are both likely parameter values and substantial SGS. Moreover, for most cases, there is a loglinear relationship, $I(1) = 0.587 - 0.117 \ln(N_e)$, between SGS, as measured by $I(1)$, Moran's coefficient for adjacent individuals, and N_e . However, there are several biologically significant exceptions, namely for very low or large N_e , SGS exceeds the loglinear values. There are also important exceptions to Crawford's formula. First, plants with low seed dispersal, high outcross pollen dispersal and high selfing rate show larger SGS than predicted. Second, in plants with very low (near zero) seed dispersal, selfing decreases SGS, opposite expectations. Finally, in some cases seed dispersal is more critical than pollen dispersal, in a manner inconsistent with Crawford's formula.

Keywords: dispersal, isolation by distance, Moran's I-statistic, neighbourhood size, spatial autocorrelation, spatial genetic structure

Received 23 February 2007; revision accepted 24 May 2007

Wright's (1943) original model of isolation by distance showed how genetic structure within populations is generated by limits to parent-offspring dispersal distances. He quantified isolation by distance with the average inbreeding coefficient within various size 'blocks' or areas of a population. Malécot (1948) reparameterized isolation by distance in a more complete way by using spatial probabilities of identity by descent between all pairs of individuals, metrics that are closely aligned with the spatial autocorrelation I-statistic of Moran (1950). Isolation by distance displays spatial genetic correlations that are relatively large for pairs of adjacent or nearby individuals and generally decrease smoothly as the distance of separation increases (e.g. Sokal & Wartenberg 1983). Both inbreeding coefficients and correlations are smaller in populations that have larger Wright's neigh-

bourhood size, which is a function of the variance of dispersal distance. Although in the literature Wright's neighbourhood size is sometimes loosely referred to as a 'panmictic unit', Wright (1946) himself stated that there is not likely to be panmixia within neighbourhoods and indeed generally spatial genetic structure (SGS) exists within neighbourhoods (e.g. Epperson 1995). Nonetheless, Wright's neighbourhood size, N_e , is a critical parameter that generally controls the degree of isolation by distance in the animal models of Wright and Malécot and SGS (e.g. Rohlf & Schnell 1971; Sokal & Wartenberg 1983). Moreover, N_e has proved to be a useful and popular measure of dispersal distance standardized for population density (e.g. if spatial distance is measured on a scale such that an individual takes up on average 1.0 units of area, then $N_e = 4\pi\sigma^2$, where σ^2 is the axial variance of dispersal distance measured on the same scale).

Since Wright (1943) formulated his isolation-by-distance model, most of the theoretical results on the case of isolation

Correspondence: Bryan K. Epperson, Fax: (517) 432-1143; E-mail: Epperson@msu.edu

by distance of greatest interest, namely that for relatively short distances in populations existing on the plane, came from computer simulation studies, in part because mathematical analyses fail (e.g. diffusion methods — Nagylaki 1978). In particular, an explicit, essentially one-to-one relationship was recently demonstrated between the degree of spatial autocorrelation and Wright's neighbourhood size, under a very wide range of conditions (Epperson 2005). However, nearly all simulation studies, as well as most mathematical results, including Wright (1943) and Malécot (1948), have been based on dispersal of male and female parents, not pollen and seed. In a series of studies, Crawford (1984a, b) used a geometric explanation to argue for a particular transformation of pollen and seed dispersal into male-parent-offspring distance and female-parent-offspring distance. He did not attempt, nor indeed is it possible, to carry this transformation through Wright's and Malécot's mathematical equations to determine the degree of isolation by distance, because both assumed the two parent-offspring distances are independent, which generally will not be true for plants. Although, the assumption of independence may not always be valid in animals, e.g. where fertilized females may move considerable distances before giving birth, dependence is omnipresent in plants. Nonetheless, the transformation has become widely used in the plant experimental literature (e.g. Govindaraju 1988; Heuertz *et al.* 2003; Smouse & Robledo-Arnuncio 2005; Robledo-Arnuncio & Austerlitz 2006). The validity of the transformation, particularly the ranges of pollen and seed dispersal distance variances over which it may determine SGS, still has not been characterized. The ability to empirically determine rigorous and confident relationships between plant dispersal characteristics and SGS requires direct theoretical study.

Although Wright (1946, 1969) studied one case with no female dispersal, the case of unequal sex-specific parent-offspring distances was not characterized for either plants or animals until recent simulation studies. In plants, paternal genetic contributions are distributed and dispersed through both pollen and seed, while female contributions are only through seed; hence, their dispersal distances cannot be the same unless pollen dispersal is zero. In addition, whenever there is any seed dispersal, the distances of dispersal of the parental contributions to any given offspring are correlated in plants, which violates primary assumptions made in both Wright's and Malécot's approaches. Thus, questions remained regarding both the appropriateness of Crawford's formula as well as Wright's original formulation when the sexes have differing dispersal characteristics, and hence the generality of Wright's neighbourhood size as a predictor of spatial genetic structure.

Recently, simulation studies have shown that nearly identical spatial autocorrelations are produced for a wide variety of cases for animals in which there are differing

female (σ_f^2) vs. male (σ_m^2) axial variances of dispersal distance, but similar total neighbourhood sizes (e.g. Epperson 1995, 2005). Total neighbourhood is defined as $N_e = 4\pi\sigma_i^2$, where $\sigma_i^2 = (\sigma_f^2 + \sigma_m^2)/2$ (Crawford 1984b), i.e. parent-offspring distance is taken as the average for the two sexes. Based on numerous simulation studies, an empirical relationship has been observed, $I(1) = 0.544 - 0.102 \ln N_e$ (Epperson 2005), where \ln is the natural logarithm and $I(1)$ is Moran's correlation coefficient for adjacent individuals, which has high statistical power as an indicator of global autocorrelation under isolation-by-distance processes (Epperson 2003). Total neighbourhood size determined $I(1)$ regardless of the relative roles of female and male animal dispersal. Female dispersal variance was modelled to range from zero up to very large numbers, in effect infinity, and male dispersal distance was modelled to range from very small to very large biological values (e.g. Epperson 1995, 2003, 2005). Although this invariance suggests that details of dispersal are not critical in the determination of SGS, it does not necessarily mean that Crawford's relationship of seed and pollen dispersal to female and male dispersal is valid, even over the same range of distances.

It has been repeatedly pointed out in the recent literature that seed dispersal can be more important than pollen dispersal (Hamrick & Nason 1996; Kalisz *et al.* 2001; Hardesty *et al.* 2005). While this is in some ways obvious from Crawford's formula, $\sigma_i^2 = \sigma_s^2 + \sigma_p^2/2$, where the dispersal distance variance of seed (σ_s^2) contributes twice as much as that of pollen (σ_p^2), it remains unclear whether or not this is the only factor, especially for extreme values of either σ_s^2 or σ_p^2 . For example, if σ_p^2 is very large or effectively infinite as what may occur in some wind-pollinated plants, the formula would predict zero spatial genetic structure, yet highly limited seed dispersal should result in some excess of spatial proximity of half-sibs and thus SGS (Kalisz *et al.* 2001). Hence, questions remain on when unequal paternal and maternal distances, as translated from seed and pollen dispersal and including the fact that seed dispersal moves both parental contributions, fit the predictions originally modelled by Wright and Malécot.

The primary goal of this article is to use extensive spatial-temporal computer simulations to test the relationship of total dispersal variance or neighbourhood size with spatial genetic structure, when plant biological processes and factors of dispersal are taken into account. Dispersal of pollen and seed are modelled explicitly, and both are simulated to range widely and cover biologically relevant conditions. In addition, self-fertilization can occur with high frequency, where pollen movement is zero-distance; hence, the effect of selfing on neighbourhood size and SGS is also investigated. Previously, low levels of selfing were modelled and observed to have little effect on spatial genetic structure (Epperson 1990). In the first study, Crawford (1984a) provided a compelling argument against

previous formulations, and in favour of the idea that selfing acts on SGS mainly through reducing the axial variance of pollen dispersal distance, where self-pollinations are simply zero-distance pollen movements. However, no analytical theory has specifically addressed this concept. In the present study, effects of selfing are studied by using low, medium and high selfing rates, together with a wide range of seed dispersal and outcross-pollination distance variances. In total, the results cover a wide range of conditions and provide a direct and broadly applicable theoretical basis for how pollen and seed dispersal determine SGS.

Methods

Simulations

Replicated spatial-temporal simulations of isolation by distance were conducted, having input variables that control seed and pollen dispersal and selfing rate and output measures of SGS. Each simulated population consisted of 10 000 individuals with diploid genotypes, located on a 100×100 square lattice (a conceptual approximation to continuous space), and each was initialized with a random distribution of diploid genotypes at an arbitrary locus, in Hardy–Weinberg proportions. A locus had five alleles, each with initially equal frequency 0.2. For loci with five alleles, the SGS of different alleles are almost stochastically independent (Epperson 2003), hence replication of the process was enhanced. Allele frequencies changed very little during the course of a simulation. Previous simulation studies showed that numbers of alleles and a wide range of frequencies have no effects on values of Moran's I-statistics (Epperson 2003).

Each set of simulations had a unique combination of dispersal and selfing, with 100 simulations per set. Most details of the simulation programme, which uses Monte Carlo methods to simulate stochastic generations of life cycles, were described previously (Epperson 1990). In the present article, the programme was restructured and rewritten to model pollen and seed dispersal, rather than male and female dispersal. Self-pollination was also added. In general, limited dispersal is modelled in the following way. The location of the seed parent of an offspring was chosen at random (using two uniform pseudo-random numbers to choose the two coordinates) from one of the nearest N_s neighbours, where N_s generally varied among sets. Then, the male parent of the seed was chosen at random from the nearest N_p neighbours of the seed parent. N_p also varied among sets. This uniform distance function may be considered unrealistic for many species in which the probability of dispersal decays smoothly with distance. However, it is justified by the fact that it and many other non-normal (including other platykurtotic as well exponential and other leptokurtotic) forms of the dispersal

Table 1 Values of Wright's neighbourhood size, N_e , for the first group of sets having both seed and pollen dispersal but only incidental selfing, i.e. $s = 0$, $s^* = 1/N_p$. Sets were numbered consecutively down rows, in columns 1–4, e.g. set 11 had $N_p = 121$ and $N_s = 49$, with $N_e = 113.10$

| | | N_p | 9 | 25 | 121 | 625 |
|-------|--------------|--------------|-------|-------|--------|--------|
| N_s | σ_s^2 | σ_p^2 | 0.67 | 2.0 | 10.0 | 50.32 |
| 9 | 0.67 | | 12.57 | 20.94 | 71.21 | 324.55 |
| 25 | 2.0 | | 29.32 | 37.70 | 87.96 | 341.30 |
| 49 | 4.0 | | 54.45 | 62.83 | 113.10 | 366.44 |
| 81 | 6.67 | | 87.96 | 96.34 | 146.61 | 399.95 |

curve have very little effect on spatial structure, rather it is the standardized neighbourhood size that matters, over a wide range of conditions (e.g. Rohlff & Schnell 1971; Barton *et al.* 2002; Epperson 2005; Lee & Hastings 2006). Dispersal was 'reflected' off of the boundaries (Epperson 1990).

The first group of 16 sets directly and solely addressed possibly contrasting functional roles of seed vs. pollen distributions in the absence of self-fertilization beyond a small incidental rate of $1/N_p$. The wide range of values of N_p vs. N_s , as well as the corresponding axial variances (Wright 1943; Crawford 1984b) for pollen, σ_p^2 , and seed, σ_s^2 , and Crawford's total neighbourhood size N_e given by $4\pi(\sigma_s^2 + \sigma_p^2/2)$, are shown in Table 1. Together, these 16 sets represent the range over which there is both biologically significant spatial genetic structure and biologically relevant relationships of pollen to seed dispersal. Since seed dispersal has twice the effect, even values of N_s of 100 would produce little SGS. A much larger value (625) of N_p was included, to mimic near-panmixia and to heighten possible contrasts between seed vs. pollen dispersal effects.

As in most other simulation studies, each simulation was run for 200 generations, by which time a population has been at quasi-equilibrium for over 100 generations (e.g. Sokal & Wartenberg 1983; Epperson 2005). In total, 1600 space–time simulations were run and analysed for this group of simulations. In this group as well as the others, the statistics were further averaged over five alleles for each simulation; hence, extensive replication was achieved.

A second group of simulation sets combined the effects of higher rates of selfing and their possible interactions with seed and (outcross) pollen dispersal. This group had a subset of the pollen and seed dispersal regimes in the first group combined with three levels of selfing. Selfing was added to the simulation model by imposing a probability (actualized by a uniform random number), s , that the pollen parent of a seed produced by the seed-parent at location x , was the same individual at x . When a selfing event did not occur, then the pollen parent was chosen at random from the N_p plants centred about x . The total selfing

| | N_s | 9 | 9 | 9 | 9 | 25 | 25 | 25 | 49 | 49 |
|-----|-------|-------|-------|-------|--------|-------|-------|--------|--------|--------|
| s | N_p | 9 | 25 | 121 | 625 | 25 | 121 | 625 | 121 | 625 |
| 0.1 | | 12.15 | 19.69 | 64.93 | 292.95 | 36.44 | 81.68 | 309.70 | 106.81 | 334.83 |
| 0.5 | | 10.45 | 14.67 | 39.79 | 166.47 | 31.42 | 56.55 | 183.21 | 81.68 | 208.35 |
| 0.9 | | 8.80 | 9.63 | 14.66 | 39.99 | 26.39 | 31.42 | 56.74 | 56.55 | 81.86 |

| | N_p | 9 | 25 | 49 | 81 | 121 | 225 | 625 |
|-----|---------|--------|--------|--------|--------|--------|--------|--------|
| s | $1/N_p$ | 0.1111 | 0.0400 | 0.0204 | 0.0123 | 0.0083 | 0.0044 | 0.0016 |
| 0.9 | | 0.9111 | 0.9040 | 0.9020 | 0.9012 | 0.9008 | 0.9004 | 0.9002 |
| 0.5 | | 0.5555 | 0.5200 | 0.5102 | 0.5062 | 0.5042 | 0.5022 | 0.5008 |
| 0.1 | | 0.2000 | 0.1360 | 0.1184 | 0.1111 | 0.1075 | 0.1040 | 0.1014 |

rate, s^* is equal to $s + (1 - s)/N_p$ (note the negligible effects of population boundaries on N_p are ignored here). It should be noted that outcross pollinations (at rate $1 - s^*$) have a correlation between the two axial dispersal distances, and while this technically violates assumptions of Wright (1943) and Crawford (1984b), it is of negligible size. Values of s were taken as 0.1, 0.5, and 0.9. A value of 0.1 represents the low selfing rate typical of nonobligate outcrossing plants, and 0.9 typical of selfers. The value 0.5 is not often observed in plants, but was included to confirm any trends associated with selfing rate.

The criteria for choosing the subset of the combinations of N_p and N_s from the first group to use in the second group was that all biologically relevant ranges be included, at least where substantial SGS is possible for the higher selfing rates. High values of N_p were one focus, as the effects of selfing in reducing pollen dispersal could be amplified. Moreover, such cases were of particular interest since there are virtually no theoretical results on SGS in plants having a mixture of wind-dispersed outcrossing and high selfing rates. In total, 27 sets, or 2700 simulations were conducted for the second group of simulations. Table 2 shows the combinations of N_p and N_s for the sets in the second group. It also gives the neighbourhood size that would obtain if selfing were treated as zero distance pollen movement, $N_0 = 4\pi[\sigma_s^2 + \sigma_p^2(1 - s)/2]$.

A third group of sets of simulations was conducted for populations having higher rates of selfing but no seed dispersal ($N_s = 1$). Initially, several sets were conducted and these gave surprising results, which seemingly might make a critical contrast between plant dispersal vs. animal dispersal. Hence, a large number of sets having zero seed dispersal were conducted to substantiate a potentially new phenomenon. Values of N_p ranged from nine to 625 and s again took the various values 0.1, 0.5, and 0.9. Table 3

Table 2 Values of the predicted neighbourhood size, if selfing were treated as zero distance pollen movement, $N_0 = 4\pi[\sigma_s^2 + \sigma_p^2(1 - s)/2]$, in the second group of sets, which had seed and pollen dispersal and higher rates of selfing. Sets were numbered consecutively down rows, e.g. set 8 had $N_s = 9$, $N_p = 121$ and $s = 0.5$

Table 3 Values of total selfing rate, $s^* = s + (1 - s)/N_p$, used in the third group of simulations, which have zero seed dispersal. Sets were numbered consecutively down rows, in columns 1–7, e.g. set 8 had $N_p = 49$ and $s = 0.5$, and $s^* = 0.5102$

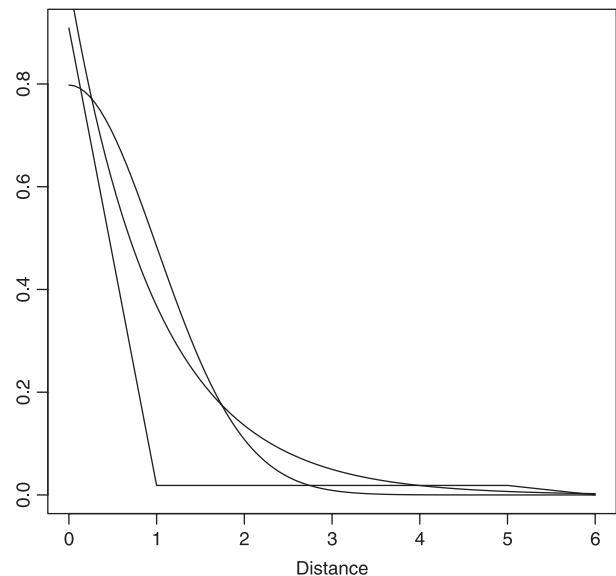


Fig. 1 Contrasting pollen dispersal distance functions, all having the same axial dispersal variance $\sigma_p^2 = 1.0$. Shown are the probability density functions for the normal and the exponential distributions, and the probability mass function for simulations with $N_p = 121$ and selfing rate 0.9.

gives the values of s^* . In total, 21 sets or 2100 simulations were conducted for the third group of simulations.

Figure 1 shows an example (set 13, no seed dispersal, $N_p = 121$, $s = 0.9$) of how extremely non-normal the dispersal distance distribution can be when high rates of selfing are considered as zero-distance pollen dispersal. Wright (1943) assumed that dispersal distance is normally distributed and Malécot's (1948) results make approximations which lead to isolation by distance relying only on the dispersal variance, not on the 'shape' of the distribution.

Although numerous simulation studies have shown that the shape of the dispersal curve is largely unimportant (Rohlf & Schnell 1971; Barton *et al.* 2002; Epperson 2005; Lee & Hastings 2006; Robledo-Arnuncio & Austerlitz 2006), such extreme forms have not been investigated.

Statistical characterization of populations

The spatial distributions of genotypes were characterized by computing the statistics at generation 200 for each simulation run, as has been done in most previous studies. This choice is largely arbitrary but unbiased and efficient. During the period from about 50 to several thousand generations, simulated populations exist in a quasi-stationary phase (e.g. Sokal & Wartenberg 1983; Epperson 1990); hence, any other generation in the range would also have been adequate. It is much more informative to replicate processes over entire simulations rather than over generations.

In preparation for calculation of Moran's I-statistic for the individual genotypes, for each allele, the genotype at each location i was converted into the values $X_i = 0, 0.5, \text{ or } 1.0$ according to the numbers (none, one, two) of that allele that were carried in the genotype. For each allele, Moran's I-statistics were calculated based on these numeric values. For each simulation, pairs or 'joins' were classified according to distance classes, D , in multiples of lattice units. For example, distance class one was 0–1.5 units, i.e. it included nearest neighbours (distance = 1.0), analogous to single space rook's moves in the game of chess and diagonal neighbours (distance = $2^{1/2}$ or $c. 1.41$). Distance class two included all pairs separated by distances in the range of (effectively) 1.5–2.5, and for example included second nearest neighbours analogous to two-step rook moves, as well as other pairs. To reduce the amount of output to manage, only distance classes one through 30 were recorded, and it is noted that by distance class 30, the correlograms have approached asymptotic values. To calculate Moran's I-statistic for each distance class D , let $Z_i = X_i - q$, where q is the mean of the genetic values also equal to the allele frequency in the simulated population. The equation for specific distance class D is written: $I(D) = [n \sum_i \sum_j w_{ij}(D) Z_i Z_j] / W_D \sum_i z_i^2$, where $w_{ij}(D)$ are binary (0,1) variables specifying the inclusion (1) or exclusion (0) of the pair of individuals i and j in class D , and W_D is the sum of the weights or twice the number of joins or pairs placed into class D (Cliff & Ord 1981). Under the randomization null hypothesis, these statistics are completely free of assumptions about the underlying distributions (Cliff & Ord 1981). Hence, for each simulation, I-statistics were calculated for each distance class and for each of five alleles.

For each set of simulations, the means and variances of individual I-statistics were calculated, separately for each distance class, across five alleles and the 100 simulations per set. All standard deviations were small, ranging from

$c. 0.025$ for large I (small N_e) down to $c. 0.005$ for small I (large N_e). Moreover, the standard errors on set averages would be smaller still by a factor of the square root of 500, hence in the range of 0.0002–0.001. Therefore, further details of the variances are not reported, nor are statistical significance levels for differences among sets. Moran's I is asymptotically normal-distributed, and for example differences in values larger than 0.002 would always be statistically significant, even though usually biologically meaningless. All Moran I-statistics were calculated by the simulation programme.

Most results on spatial genetic structure are presented in terms of the Moran I-statistic for the shortest distance class, $I(1)$. There are several reasons for doing so (Epperson 2005), but most importantly, the use of the entire spatial structure (or as it is characterized by the I-correlogram) is complicated by the fact that estimates for different distances (distance classes) are not independent, but have correlations that depend on the spatial distribution itself. Thus, for example the Q-statistic of Oden (1984), which tests entire correlograms for statistical significance, is conservative. Oden (1984) found that tests of significance for the first distance class and the Q-test have about the same statistical power under the process of isolation by distance. Further, a direct relationship between $I(1)$ and N_e has been observed for the case of male and female dispersal (Epperson 2005), and this relationship may be tested against in the present study of seed and pollen dispersal, to determine if and when both types of dispersal biologies produce the same degree of isolation by distance solely through their effects on total dispersal variance (and N_e).

Additional statistics were computed on set mean values (over alleles and replicate simulations), in order to examine trends, including linear regression, using R 2.4.0 (The R Development Core 2006).

Finally, although inbreeding per se is peripheral to the main research questions, values for the fixation index were calculated and reported. Both the allele-specific maximum-likelihood estimator (MLE) of Brown (1970), and the multi-allelic estimator of Li & Horvitz (1953) were calculated; further, the averages of the MLEs over alleles were calculated for each simulation and then averaged over simulations for each set. Only the values of the Li and Horvitz estimator are presented, because the values of the average MLEs were nearly identical, always differing by less than 0.001.

Results

Spatial genetic structure without selfing

Set average values of Moran's statistic for the first distance class, $I(1)$, for the first group of simulation sets, having both pollen and seed dispersal but only incidental selfing ($s = 0$,

Table 4 Values of Moran I-statistic for the first distance class for the first group of models with seed and pollen dispersal and no selfing, i.e. $s = 0$, $s^* = 1/N_p$. Sets were ordered consecutively down rows, as in Table 1

| | | N_p | 9 | 25 | 121 | 625 |
|-------|--------------|--------------|-------|-------|-------|-------|
| N_s | σ_s^2 | σ_p^2 | | | | |
| 9 | 0.67 | | 0.313 | 0.226 | 0.097 | 0.032 |
| 25 | 2.0 | | 0.173 | 0.145 | 0.071 | 0.022 |
| 49 | 4.0 | | 0.105 | 0.094 | 0.057 | 0.017 |
| 81 | 6.67 | | 0.069 | 0.063 | 0.042 | 0.014 |

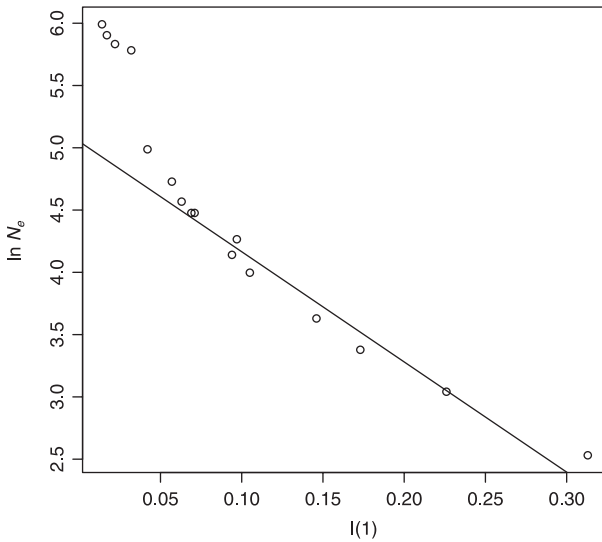


Fig. 2 Values of Moran's I for distance class one ($I(1)$) for the 16 sets in the first group, having various levels of pollen and seed dispersal (Table 1) and only incidental selfing (with rate $1/N_p$), as a function of the natural logarithm of the total neighbourhood size, N_e . Also shown is the regression for sets 1–12.

$s^* = 1/N_p$), are shown in Table 4. Because the relationship of $I(1)$ to N_e appeared to be nearly exponential (Fig. 2), linear regression of $I(1)$ on $\ln N_e$ was conducted. The regression had slope -0.073 and intercept 0.425 , results similar to the regression observed for male and female animal dispersal, $I(1) = 0.544 - 0.102 \ln N_e$, obtained in earlier studies (Epperson 2005). Moreover, just as in the latter case, values of $I(1)$ for very large N_e tended to be underestimated by linear regression. In other words, in both studies the loglinear relationship does not hold for very large N_e . Specifically, the rate of decrease of $I(1)$ with increasing N_e is markedly higher among the four sets having $N_e > 150$. Hence, a second analysis was done after omitting models with $N_e > 150$, and the regression was $I(1) = 0.571 - 0.113 \ln N_e$ (Fig. 2), for all practical purposes identical to the animal dispersal cases (Epperson 2005). The adjusted R-squared

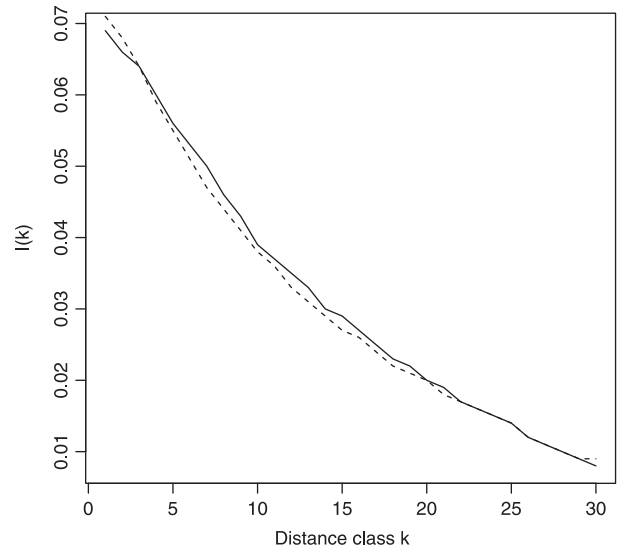


Fig. 3 Average I-correlograms for two sets in the first group of simulations, having both seed and pollen dispersal and with only incidental selfing (with rate $1/N_p$). The solid line is for set 4 ($N_p = 9$, $N_s = 81$) and the dashed line for set 10 ($N_p = 121$, $N_s = 25$). The two sets had very different levels of seed vs. pollen dispersal, but identical neighbourhood size, $N_e = 87.96$ (Table 1).

value was 0.96 compared to 0.90 for the regression using all 16 sets, indicating a very tight fit. Thus, despite the wide range of levels of pollen and seed dispersal represented in the sets, the value of $I(1)$ was fully predicted by the logarithm of Crawford's measure of N_e , as long as the latter was in the range of $c. 10$ – 150 (Fig. 2). For example, set 4 ($\sigma_p^2 = 0.67$, $\sigma_s^2 = 6.67$) and set 10 ($\sigma_p^2 = 10.0$, $\sigma_s^2 = 0.67$), which had identical N_e (87.96 , Table 1) but very different pollen vs. seed dispersal, produced very similar values of $I(1)$, 0.069 and 0.071 , respectively. Finally, it should be noted that the set with minimal pollen and seed dispersal, set 1, produced a value, 0.313 , that is somewhat larger than predicted by the regression, i.e. somewhat of an outlier (Fig. 2). The same was true for the case of smallest N_e for male and female dispersal (Epperson 2005).

Entire I-correlograms were also similar for dispersal models having the similar total neighbourhood size. An example, for sets 4 and 10, which had very different seed vs. pollen dispersal levels, as noted above, but identical values of N_e and very similar correlograms, is shown in Fig. 3.

Spatial genetic structure with selfing

Mean values of $I(1)$ for the second group of simulation sets, having nonzero σ_s^2 and σ_p^2 but with substantial selfing ($s = 0.1, 0.5$ or 0.9), are shown in Table 5. As expected, values of $I(1)$ were larger when selfing rate was increased, and smaller as either σ_s^2 or dispersal variance for outcross pollinations were increased.

| | N_s | 9 | 9 | 9 | 9 | 25 | 25 | 25 | 49 | 49 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| s | N_p | 9 | 25 | 121 | 625 | 25 | 121 | 625 | 121 | 625 |
| 0.1 | | 0.319 | 0.236 | 0.104 | 0.038 | 0.150 | 0.078 | 0.026 | 0.059 | 0.021 |
| 0.5 | | 0.345 | 0.287 | 0.160 | 0.079 | 0.168 | 0.109 | 0.048 | 0.076 | 0.035 |
| 0.9 | | 0.374 | 0.357 | 0.300 | 0.212 | 0.188 | 0.171 | 0.127 | 0.102 | 0.082 |

Table 5 Values of $I(1)$ for simulation sets shown that had partial selfing, and both seed and pollen dispersal, as laid out in Table 2

In order to compare the observed values of $I(1)$ to those predicted if selfing is treated precisely as zero-distance pollen movement, hence predicted neighbourhood $N_0 = 4\pi[\sigma_s^2 + \sigma_p^2(1-s)/2]$, attention must be paid to which part of the range of values of both N_e , taken as if $s = 0$ or equivalently selfing is ignored in the calculation of N_e , and N_0 are found in the distribution shown in Fig. 2. For sets in which both N_e and N_0 are less than $c. 150$, we would expect the change in $I(1)$ due to selfing to be linearly related to $(\ln N_0 - \ln N_e)$. Indeed, using the empirical result from the first group of simulations (where $s = 0$, $s^* = 1/N_p$), the slope of increase would be 0.113. Alternatively, values of $I(1)$ could be predicted directly from the equation $I(1) = 0.571 - 0.113 \ln N_0$, and in fact observed values were quite close to those predicted in this manner (results not shown), for these simulation sets with $N_0 < 150$. However, a more precise prediction can be obtained by using the values of $I(1)$ obtained in sets having the same values of N_s and N_p but with $s = 0$ (Table 4), and setting $I(1)_{\text{predicted}} = I(1)_{s=0} - 0.113 (\ln N_0 - \ln N_{e,s=0})$. For example, for set 8, $N_s = 9$, $N_p = 121$, $s = 0.5$, $N_0 = 39.79$, and the observed value, 0.160, can be compared using set 9 of the first group ($N_s = 9$, $N_p = 121$, $s = 0$), which had $N_{e,s=0} = 71.21$ (Table 1), and the value 0.097 was found for $I(1)_{s=0}$ (Table 4). For this example, $I(1)_{\text{predicted}} = 0.097 - 0.113 (\ln(39.79) - \ln(71.21)) = 0.163$. Sets 1–9, 13–16, and 22–24 had both N_e and N_0 less than 150, and the deviations of the observed values from those predicted by this manner are shown in Table 6. All deviations are small, either in absolute values or as percentages of $I(1)_{\text{predicted}}$, or both. There was a slight trend for observed $I(1)$ to exceed $I(1)_{\text{predicted}}$ for small N_e and the opposite for large N_e . Nonetheless, overall the SGS created when selfing is added fits very closely expectations when selfing is treated as zero-distance pollen movement, for these sets.

Predicted values for the remaining sets, those with very high dispersal ($N_p = 625$) among outcross pollinations, were generated in two distinct ways. In cases (sets 12, 21, and 27) where N_0 is of moderate size (< 150) but the neighbourhood size for the corresponding set with $s = 0$, $N_{e,s=0}$ is large, say > 150 (note all of these sets have $N_p = 625$ and $s = 0.9$), the I-statistics cannot be assumed to be log-linearly related to the contrasts in neighbourhood size. The contrasts span the critical part of the distribution (Fig. 2). However, because N_0 is moderate, the equation

Table 6 Values of observed $I(1) - I(1)_{\text{predicted}}$ for those sets in the second group (seed and pollen dispersal and selfing) that have both N_e and $N_0 < 150$, as described in the text

| | N_s | 9 | 9 | 9 | 25 | 25 | 49 |
|-----|-------|-------|-------|--------|-------|--------|--------|
| s | N_p | 9 | 25 | 121 | 25 | 121 | 121 |
| 0.1 | | 0.002 | 0.003 | -0.003 | 0.000 | -0.001 | -0.004 |
| 0.5 | | 0.011 | 0.021 | -0.003 | 0.001 | -0.012 | -0.018 |
| 0.9 | | 0.021 | 0.043 | -0.025 | 0.002 | -0.016 | -0.033 |

$I(1) = 0.571 - 0.113 \ln N_0$ is a valid predictor. Values of the observed values minus the predicted are 0.056, 0.011 and 0.007, for sets 12, 21, and 27, respectively. Only set 12 shows a non-negligible difference. It is worth noting that set 12 has minimal seed dispersal, maximum N_p and maximum selfing rate. In the remaining cases (sets 10, 11, 19, 20, 25, 26) where N_0 (as well the neighbourhood size for the corresponding set with $s = 0$) is large, greater than 150, the equation cannot be used, indeed for N_e exceeding 159 the equation predicts negative values. An alternative is to use the slope measured for the four sets (sets 12–16) in the first group ($s = 0$) of simulations (Table 4) having large values of N_e , i.e. using a slope of -0.081 in place of -0.113 . Consequently, the deviations were -0.002 , -0.007 , -0.004 , -0.024 , -0.003 , and -0.028 , for sets 10, 11, 19, 20, 25 and 26, respectively. Deviations were large only for sets 20 and 26, and then only as percentages.

Spatial genetic structure with selfing and no seed dispersal

In the third group of sets, where there was no seed dispersal, low to moderate rates of selfing had very little effect on spatial genetic structure (Table 7). Perhaps surprisingly, even when $s = 0.5$, and hence the variance of all pollen dispersal distances was halved, the values of $I(1)$ were virtually unchanged from the case with $s = 0$. In contrast, when the selfing rate was high, $s = 0.9$, the values of $I(1)$ were strongly affected, however, in the direction opposite those based on the argument that s contributes as zero-distance pollen dispersal—values of $I(1)$ are decreased, and opposite the effect observed when there were even small amounts of seed dispersal.

Table 7 Values of $I(1)$ for the third group of simulation sets, having no seed dispersal, and varying rates of selfing, s , and dispersal of outcross pollinations. Values in the last row are included for comparison and were obtained previously (Epperson 2005). Maximum values for a given value of N_p are indicated in bold, minimum values in italics

| | $N_p = 9$ | $N_p = 25$ | $N_p = 49$ | $N_p = 81$ | $N_p = 121$ | $N_p = 225$ | $N_p = 625$ |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| $s = 0.9$ | 0.360 | 0.184 | 0.107 | 0.068 | 0.047 | 0.027 | 0.010 |
| $s = 0.5$ | 0.450 | 0.247 | 0.146 | 0.096 | 0.069 | 0.039 | 0.013 |
| $s = 0.1$ | 0.454 | 0.250 | 0.150 | 0.097 | 0.067 | 0.037 | 0.012 |
| $s = 0.0$ | 0.454 | 0.248 | 0.149 | 0.096 | 0.066 | 0.036 | 0.010 |

Examined in finer detail, the array of values of $I(1)$ exhibits complex behaviour. Whereas $I(1)$ is clearly and sharply lower for $s = 0.9$ (except for a 'tie' with $s = 0$ when $N_p = 625$), the maximum of $I(1)$ for the same N_p but across s values does not occur at a consistent value of s . Although the differences among values across $s = 0, 0.1$ and 0.5 are small, many of them would be statistically significant, given the small standard errors (ranging from 0.0002 to 0.001). As outcross pollen dispersal increases, the value of s giving maximum $I(1)$ increases. Both the behaviour of the maximum and the large reductions caused by large selfing rates suggest that although increased dispersal generally homogenizes the population, when there is no seed dispersal and little effective pollen dispersal, there is not enough dispersal to effect the role of locally spreading local stochastic changes. This apparent conundrum may not be unexpected. For example, for set 1, if $N_0 = 4\pi[\sigma_s^2 + \sigma_p^2(1-s)/2]$ is used, then N_0 would equal $4\pi[0 + 0.67(0.1)/2] = 0.42$. There is so little movement that stochastic events (mainly in the form of selfed progeny of heterozygotes) are scarcely shared among adjacent and neighbouring locations. Hence, even a slight increase in N_0 , e.g. $s = 0.5$ in set 2 (where N_0 is 2.1), can greatly increase local sharing without contributing much to larger scale homogenization, and thus cause an increase in $I(1)$. Dispersal allows nearby locations to share stochastic inputs, yet it also 'smooths' out local stochastic fluctuations: there is a tension between these two effects, but normally and for most of the parametric range, the latter dominates, and the former effect would be unnoticeable. Analogous effects occur in processes of isolation by distance among discrete populations (Epperson 2003). The effects are important only at the extremes of the parametric ranges, i.e. extremely low seed dispersal and high selfing; nonetheless these extremes do occur occasionally in nature.

In the cases where selfing sharply reduced $I(1)$, correlograms also exhibited markedly different shapes (e.g. Fig. 4). Short distance values were uniformly decreased, but more interesting, the distances at which the correlograms became negative, the 'X-intercept' (Sokal & Wartenberg 1983; Epperson 1990) were reduced substantially, again demonstrating the reduced spatial scale over which localized stochastic changes were shared over time.

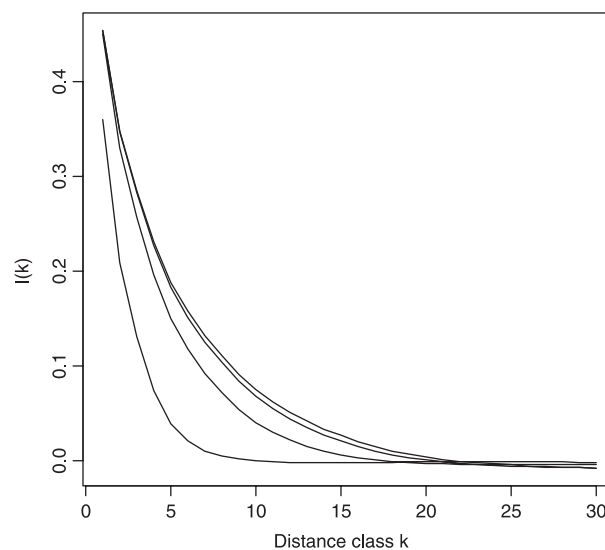


Fig. 4 Average I -correlograms for sets 1–3 ($N_p = 9$) of the third group of simulation sets, which have no seed dispersal but with various levels of selfing. Also shown for comparison is the correlogram for the analogous case (same outcross pollen dispersal function and no seed dispersal) with selfing rate zero from Epperson (2005). Curves from uppermost to bottom are for selfing rates 0.0, 0.1, 0.5, and 0.9, respectively.

Fixation index

Although the focus of this article is on the effects of different types of dispersal on spatial genetic structure per se, the influence of spatial genetic structure on the fixation index is also important. Tables 8 and 9 give the fixation indices for the sets with $s > 0$. As expected, values increase as selfing rates increase or as neighbourhood sizes decrease. Comparisons of observed F to those expected under the mixed mating model (which takes into account only selfing and ignores biparental inbreeding as well as spatial distribution, see e.g. Ennos & Clegg 1982), F_{σ} , are also displayed in Tables 8 and 9. Cases with lower dispersal and greater spatial genetic autocorrelation exhibit greater excesses, attributable to greater biparental inbreeding. Among cases in which there was no seed dispersal (Table 9), those with smaller selfing rates show greater inflation; greater outcrossing implies greater incidence of

| | N_s | 9 | 9 | 9 | 9 | 25 | 25 | 25 | 49 | 49 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| s | N_p | 9 | 25 | 121 | 625 | 25 | 121 | 625 | 121 | 625 |
| (a) | | | | | | | | | | |
| 0.1 | | 0.297 | 0.196 | 0.098 | 0.061 | 0.151 | 0.090 | 0.061 | 0.085 | 0.061 |
| 0.5 | | 0.551 | 0.478 | 0.389 | 0.346 | 0.427 | 0.378 | 0.344 | 0.366 | 0.343 |
| 0.9 | | 0.896 | 0.878 | 0.852 | 0.830 | 0.855 | 0.843 | 0.829 | 0.834 | 0.827 |
| (b) | | | | | | | | | | |
| 0.1 | | 0.186 | 0.123 | 0.041 | 0.008 | 0.078 | 0.033 | 0.008 | 0.028 | 0.008 |
| 0.5 | | 0.166 | 0.127 | 0.052 | 0.012 | 0.076 | 0.041 | 0.010 | 0.029 | 0.009 |
| 0.9 | | 0.059 | 0.053 | 0.032 | 0.011 | 0.030 | 0.023 | 0.010 | 0.014 | 0.008 |

Table 8 (a) Observed values of the fixation index, F , for simulation sets shown that had partial selfing (group 2), and both seed and pollen dispersal, as laid out in Table 2. (b) Excess of observed values over those expected under the mixed mating model

| s | N_p | 9 | 25 | 49 | 81 | 121 | 225 | 625 |
|-----|-------|--------|--------|--------|--------|--------|--------|--------|
| (a) | | | | | | | | |
| 0.9 | | 0.8367 | 0.8248 | 0.8215 | 0.8202 | 0.8195 | 0.8188 | 0.8185 |
| 0.5 | | 0.3846 | 0.3514 | 0.3425 | 0.3389 | 0.3371 | 0.3353 | 0.3340 |
| 0.1 | | 0.1111 | 0.0730 | 0.0629 | 0.0588 | 0.0568 | 0.0549 | 0.0534 |
| (b) | | | | | | | | |
| 0.9 | | 0.0557 | 0.0281 | 0.0168 | 0.0098 | 0.0065 | 0.0044 | 0.0011 |
| 0.5 | | 0.2245 | 0.1121 | 0.0631 | 0.0412 | 0.0286 | 0.0161 | 0.0043 |
| 0.1 | | 0.2851 | 0.1333 | 0.0769 | 0.0488 | 0.0332 | 0.0174 | 0.0049 |

Table 9 (a) Expected values of F for the third group of simulation sets, having no seed dispersal, and varying rates of selfing, s , and dispersal of outcross pollinations. (b) Excess of observed values over those expected under the mixed mating model

inbreds caused by biparental productions of offspring, and this results in overall greater excesses of inbreeding. The same trend is strong in the sets that did have seed dispersal (Table 8). However, there were exceptions where excesses were greater for $s = 0.5$ than for $s = 0.1$. As observed above, higher selfing rates increase spatial genetic structure when there is seed dispersal, hence even though higher selfing reduces the frequency of outcrosses, those outcrosses will be more correlated, given the same pollen dispersal function. Conversely, selfing does not increase SGS when there is no seed dispersal (indeed at high rates it can reduce it), and only the effect of decreased outcrossing is observed. Finally, spatial genetic structure would bias the estimator of s based on F much more when s is small. A combination of very low dispersal and relatively low true selfing rate would make F_e entirely unsuitable for estimating s .

Discussion

In the first group of simulations (for which selfing was minimal) spatial genetic structure is almost fully determined by the total neighbourhood size defined by Crawford (1984b), $N_e = 4\pi(\sigma_s^2 + \sigma_p^2/2)$. Sets having the same value of N_e but different variances of pollen (σ_p^2) and seed (σ_s^2) dispersal distances produced essentially identical values of Moran's I-statistic for nearest neighbours ($I(1)$), even when the differences in variances were

very large. SGS was very similar to that produced in models mimicking 'animal' dispersal, where diploid females and males disperse, at least for most of the parametric range. Hence, the results show that Crawford's formula is valid for outcrossers in most cases of biological relevance. Indeed, values of $I(1)$ were loglinearly related to N_e among the study simulations, as they were for animal dispersal, as long as N_e was not either large (> 150) or very small (< 10). Indeed, linear regression of $I(1)$ on the natural logarithm of N_e yielded $I(1) = 0.571 - 0.113 \ln N_e$, very similar to that obtained for animal dispersal, $I(1) = 0.544 - 0.102 \ln N_e$ (Epperson 2005). Small differences either in intercept or slope could be due to differences in numbers of simulations near either end of the range $10 < N_e < 150$. When N_e exceeds 150 SGS is very weak, and Wright (1943) predicted SGS to be very weak when $N_e > 200$. Hence, the results show that for most cases ($N_e < 150$) where there is substantial SGS, Crawford's formula is appropriate for outcrossing plants.

Values of $I(1)$ for very large N_e tended to be underestimated by the regression equation, both in the present study and in Epperson (2005). While formal comparisons cannot be made (since it was not possible to set the values of N_e equal in the two studies), none of the models studied produced SGS that was obviously inconsistent with those produced by animal models having similar values of N_e . At the opposite end of the spectrum, set 1, having a neighbourhood size that is nearly the smallest conceivable for

an outcrossing plant, produced correlations that were slightly underestimated by regression, and this was also true for animal models having very low N_e (Epperson 2005). Hence, even at the extremes, Crawford's formula is not called into question for outcrossers.

SGS, and in particular values of $I(1)$, can be used to provide indirect estimates of N_e in natural populations (Epperson 2005). The results of this study can be used to refine such methods, by combining them with other studies. The fact that the present results were loglinear but only when $N_e < 150$ reinforces the findings of Epperson (2005), and it implies that the estimator should only be used for those cases. In order to obtain a refined estimator, regression was conducted on mean $I(1)$ values from 20 sets, 11 from the present study and nine from Epperson (2005), all cases with $N_e < 150$, and it generated the equation $I(1) = 0.587 - 0.117 \ln(N_e)$ (R-squared = 0.96), nearly identical to that for the present study. Thus the refined estimator of N_e is $\exp[(0.587 - I(1))/0.117]$.

The results of the second group of sets, in which there was both substantial selfing and seed and pollen dispersal, showed that as long as there is at least a small amount of seed dispersal, the effect of selfing on $I(1)$ is fully captured by its reduction in the variance of dispersal distances of pollen, in almost all normal conditions. Selfing occurred with rate s^* equal to $s + (1 - s)/N_p$, where s may be taken as selfing rate controlled by various mechanisms and $1/N_p$ is an increment caused by incidental self-pollination, a parameterization that bears a convenient relationship to Crawford's formula. The predicted neighbourhood size (and its resultant SGS) if selfing acts on SGS only as zero-distance pollen dispersal distance is $N_0 = 4\pi[\sigma_s^2 + \sigma_p^2(1 - s)/2]$. All values of deviations of observed $I(1)$ from those predicted by N_0 were small. In all cases where N_0 and the usual N_e if selfing were ignored ($N_{e,s=0}$, i.e. not including self-pollinations in the dispersal variance) were not too large (< 150), selfing increased SGS by amounts indistinguishable from $0.113(\ln N_{e,s=0} - \ln N_0)$, precisely as if σ_p^2 were multiplied by $1 - s$. There was a slight trend for observed $I(1)$ to exceed $I(1)_{\text{predicted}}$ for small N_e and the opposite for relatively large N_e . Nonetheless, overall the SGS created when selfing is added usually fits very closely expectations when selfing is treated solely as reducing pollen dispersal variance, as long as there is at least a small amount of seed dispersal.

For the cases where either N_0 or $N_{e,s=0} > 150$, the results were also highly consistent with selfing acting solely by reducing the pollen dispersal variance, with a few exceptions. One set (set 12) showed values that were considerably larger than expected. This case had the unique combination of lowest (but nonzero) seed dispersal, highest selfing rate ($s = 0.9$) and highest dispersal variance for outcross pollen σ_p^2 . Hence, in this one case, there is a significant interaction between selfing and dispersal, such that it is not

appropriate to use N_0 as a modification of Crawford's formula. Some other differences were observed when N_0 was very large and $I(1)$ very small, which could be important in terms of percentage changes in $I(1)$ but not in absolute values. However, in these cases values of $I(1)$ are so small that it would usually be difficult or unadvisable to use them to estimate dispersal from SGS data in real populations. Nonetheless, such exceptions show that selfing does not always act on SGS solely through Crawford's formula. Assumptions made by Crawford (1984b) and Wright (1943) can be substantially violated and this can cause their results to become invalid predictors of the SGS caused by isolation by distance.

In the third group of simulation sets, where there was substantial selfing but no seed dispersal, SGS exhibited complex behaviour as a function of selfing rates and outcross pollen dispersal variance. The general trend that selfing decreased SGS was perhaps surprising, and opposite expectations based on modifying Crawford's formula by treating self-pollination as zero-distance pollen dispersal. This result points out the dual effects of dispersal. The predominant effect is one of homogenizing the population at both small and large spatial scales, i.e. smoothing out localized differences in genotypic frequencies. However, dispersal also allows adjacent or nearby plants to share local stochastic changes (namely, the role of chance in gamete-types segregating from heterozygotes), and hence genetic correlations. Such duality is also observed in models of isolation by distance among discrete populations. Migration between adjacent populations makes them share stochastic changes (inputs) caused by genetic drift within populations and causes them to be spatially correlated, yet migration over time also homogenizes populations at larger spatial scales and tends to reduce autocorrelation. The importance of localized sharing of stochastic inputs is explicit in the Space Time Autoregressive (STAR) representation of migration-drift processes, where STAR can be inverted into a Space Time Moving Average process, which acts as a (sharing) filter on white noise (Epperson 2003). Thus, gene flow binds populations together yet can also reduce correlations among spatially proximal populations. Normally, the increased 'smoothing' effects of increasing dispersal greatly outweigh those from increased sharing. In the present simulations, in those cases having no seed dispersal, when selfing rates are large ($s = 0.9$), irrespective of σ_p^2 for outcrosses, dispersal is so limited there is very little spatially proximal sharing of localized stochastic inputs. Hence, relative decreases in selfing rate (e.g. to $s = 0.5$) greatly increase the relative amount of sharing, more than offsetting the increased spatial smoothing effects of dispersal, and thus increasing the SGS correlations.

Although among sets having the same outcross pollination dispersal distance variance but different selfing rates ($s = 0, 0.1, 0.5, \text{ and } 0.9$) the minimum of $I(1)$ was always

lowest for the case with highest selfing rate ($s = 0.9$), the maximum of $I(1)$ did not occur at a consistent value of s . Further, although the differences are small across values of $s = 0, 0.1$ and 0.5 , many would be statistically significant and hence are real effects. As outcross pollen dispersal increases, the value of s that gives the highest value of $I(1)$ tends to increase. This complex behaviour can also be explained by tradeoffs in the dual effects of dispersal mentioned above. The retardation of SGS caused by high selfing rates when seed dispersal was minimal was also apparent (e.g. Fig. 4) in reductions in the distance at which I -correlograms become negative, the so-called X -intercept (Sokal & Wartenberg 1983). The spatial scale of autocorrelation can be reduced by as much as 50% when the selfing rate is 0.9.

The contrast of the behaviour of SGS with selfing for zero seed dispersal (Table 7) vs. that when seed dispersal is nonzero but nearly minimal (sets 1–3 of the second group, Table 5) is extreme, and this contrast shows that very small changes in amounts of seed dispersal can have critical effects. This contrast is not evident in the theory of Wright (1943) and Malécot (1948), nor in the formula of Crawford (1984b). Further, for populations of plants where the selfing rate is high and seed dispersal distances are fairly short, e.g. plants having 'gravity'-dispersed seed (Hamrick & Godt 1990), this result indicates that predicted SGS may depend critically on the population density, demographics and particular components of ecology. For example, in a canopy-gap filling, tropical tree species having large, gravity-dispersed seed or fruits and low population density, most regeneration could be effectively replacement events of the maternal parent. If the species also is highly selfing, then it could exhibit SGS like that modelled as zero seed dispersal. In contrast, in denser populations, SGS should behave more like those simulation sets where there was seed dispersal.

In total, the results showed that for the most part, seed dispersal is not 'more important' than pollen dispersal, with respect to isolation by distance, beyond Crawford's argument that it contributes twice as much to the neighbourhood size. Nonetheless, the contrast described above does provide an instance where it is indeed more critical than pollen dispersal. In addition, the results of the first group of simulations show that plants that have highly dispersed pollen but moderate or lower seed dispersal will have greater amounts of SGS than might be expected from Crawford's formula and the loglinear relationship between SGS and neighbourhood size. Hence, the results help to clarify the relative roles of seed vs. pollen dispersal under various biological conditions.

The results also further substantiate how little the shape of the dispersal distribution affects SGS. Even when the selfing rate is high and causes the distribution to be extremely different from the normal one assumed by Wright (1943),

negligible effects were generally observed. The non-normality included in the present study (e.g. Fig. 1) was often much more extreme than in previous studies, which generally also found negligible effects (Rohlf & Schnell 1971; Barton *et al.* 2002; Epperson 2005; Lee & Hastings 2006). The cases where SGS differed strongly from Crawford occurred only with zero seed dispersal, and there any skewness effects are confounded with the other strong (and counter to Crawford) effects of high selfing rate.

Given the lack of effects of non-normality, it is noteworthy that Wright's model and Crawford's formula are still not entirely precise for some parameter combinations with seed dispersal greater than zero. One possible explanation is that Wright assumed that male and female dispersal are independent, whereas in plants the male parental gamete and the female parental gamete are dispersed together in the seed. Hence, the dispersal distances of the two gametes are correlated. Moreover, the correlations are highest when there is little pollen movement, such as when selfing rate is high. In total, the present results show that for most cases such correlations do not substantially change the SGS predicted, but in some cases they may.

It is worth noting that the fixation index at the spatial quasi-stationary state was virtually indistinguishable from the equilibrium value F_e predicted by the rate of selfing in the (nonspatial) mixed mating model (Weir & Cockerham 1973) in all cases where SGS was weak, as would be expected. Further as expected, cases with stronger SGS had increasingly large excesses of observed F over F_e . Nonetheless, the results quantify effects of biparental inbreeding in explicit space-time processes. In some cases, if selfing rate were inferred from F_e there would be little bias, and the present results could be used to account for bias. Use of F_e as an estimator of s is generally considered to be an inaccurate method, yet the present results also indicate that (assuming the equilibrium assumption is valid) it can be accurate. However, when the true selfing rate is low and SGS is strong, then the estimator is entirely inappropriate.

Summary

The results of this study show how SGS is determined by Crawford's (1984b) definition of neighbourhood size for plants, under a wide range of conditions. For outcrossers, SGS is almost fully determined, regardless of the relative sizes of seed (σ_s^2) vs. pollen (σ_p^2) dispersal variances. In fact, for most cases where there is substantial SGS, the Moran statistic for neighbouring plants, $I(1)$, bears a loglinear relationship with N_e . For very small or very large N_e , the relationship is no longer loglinear, but the level of SGS still fits that expected from Crawford's formula. Importantly, when pollen dispersal is great and seed dispersal small, SGS is stronger than expected based on the loglinear equation. The results also showed that self-fertilization affects SGS

solely as zero-distance pollen movement, through reducing the variance of pollen dispersal distance σ_p^2 in Crawford's formula and thereby increasing $I(1)$, for most cases of biological importance. However, this relationship is violated when there is a combination of relatively low seed dispersal, high selfing rate and large dispersal distance variance for outcross pollinations. The results further showed that when seed dispersal is highly limited, selfing can have complex, even opposite effects on SGS, and clearly there Wright's and Crawford's formulae break down. This behaviour illustrates the duality of gene dispersal: it allows localized sharing of local stochastic changes; and it causes spatial smoothing or homogenizing of the population. The former has not been widely acknowledged in the literature. Finally, while seed dispersal generally acts as double the pollen dispersal, there are cases in which seed dispersal plays a more critical role, in a manner inconsistent with Crawford's formula.

Acknowledgements

The author thanks Terry Crawford for several important insights, and Kim Scribner, Yu-Wen Pan and an anonymous reviewer for helpful comments on an earlier version of the manuscript.

References

- Barton NH, Depaulis F, Etheridge AM (2002) Neutral evolution in spatially continuous populations. *Theoretical Population Biology*, **61**, 31–48.
- Brown AHD (1970) The estimation of Wright's fixation index from genotypic frequencies. *Genetica*, **41**, 399–406.
- Cliff AD, Ord JK (1981) *Spatial Processes*. Pion, London.
- Crawford TJ (1984a) What is a population? In: *Evolutionary Ecology* (ed. Shorrocks B), pp. 135–173. Blackwell Scientific Publications, Oxford, UK.
- Crawford TJ (1984b) The estimation of neighborhood parameters for plant populations. *Heredity*, **52**, 273–283.
- Ennos RA, Clegg MT (1982) Effect of population substructuring on estimates of outcrossing rate in plant populations. *Heredity*, **48**, 283–292.
- Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics*, **124**, 757–771.
- Epperson BK (1995) Fine-scale spatial structure: correlations for individual genotypes differ from those for local gene frequencies. *Evolution*, **49**, 1022–1026.
- Epperson BK (2003) *Geographical Genetics*. Princeton University Press, Princeton, New Jersey.
- Epperson BK (2005) Estimating dispersal from short distance spatial autocorrelation. *Heredity*, **95**, 7–15.
- Govindaraju DR (1988) Life histories, neighbourhood sizes, and variance structure in some North American conifers. *Biological Journal of the Linnean Society*, **35**, 69–78.
- Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates, Sunderland, Massachusetts.
- Hamrick JL, Nason JD (1996) Consequences of dispersal in plants. In: *Population Dynamics in Ecological Space and Time* (eds Rhodes OEJ, Chesser RK, Smith MH), pp. 203–236. The University of Chicago Press, Chicago.
- Hardesty BD, Dick CW, Kremer A, Hubbell S, Bermingham E (2005) Spatial genetic structure of *Simarouba amara* Aubl. (Simaroubaceae), a dioecious, animal-dispersed Neotropical tree, on Barro Colorado Island, Panama. *Heredity*, **95**, 290–297.
- Heuertz M, Vekemans X, Hausman JF, Palada M, Hardy OJ (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, **12**, 2483–2495.
- Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history, and selection. *Evolution*, **55**, 1560–1568.
- Lee CT, Hastings A (2006) Non-equilibrium genetic structure is insensitive to the shape of the dispersal distribution. *Evolutionary Ecology Research*, **8**, 279–293.
- Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics*, **5**, 107–117.
- Malécot G (1948) *Les Mathématiques de l'Hérédité*. Masson, Paris.
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.
- Nagylaki T (1978) A diffusion model for geographically structured populations. *Journal of Mathematical Biology*, **9**, 101–114.
- Oden NL (1984) Assessing the significance of a spatial correlogram. *Geographical Analysis*, **16**, 1–16.
- R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Robledo-Arnuncio JJ, Austerlitz F (2006) Pollen dispersal in spatially aggregated populations. *American Naturalist*, **168**, 500–511.
- Rohlf FJ, Schnell GD (1971) An investigation of the isolation-by-distance model. *American Naturalist*, **105**, 295–324.
- Smouse PE, Robledo-Arnuncio JJ (2005) Measuring the genetic structure of the pollen pool as the probability of paternal identity. *Heredity*, **94**, 640–649.
- Sokal RR, Wartenberg DE (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics*, **105**, 219–237.
- Weir BS, Cockerham CC (1973) Mixed self and random mating at two loci. *Genetical Research*, **21**, 247–262.
- 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 (1969) *Evolution and the Genetics of Populations*, Vol. 4. *Variability Within and Among Natural Populations*. University of Chicago Press, Chicago.

Bryan Epperson, Professor at Michigan State University, studies theoretical and statistical aspects of geographical genetics, and uses molecular markers to study the spatial population genetics of trees and other species.
