Evolution, 57(5), 2003, pp. 995–1007

QUANTIFYING GENE FLOW FROM SPATIAL GENETIC STRUCTURE DATA IN A METAPOPULATION OF *CHAMAECRISTA FASCICULATA* (LEGUMINOSAE)

CHARLES B. FENSTER,[1] XAVIER VEKEMANS,[2,3,4] AND OLIVIER J. HARDY[2,5]
[1]*Department of Biology, H. J. Patterson Hall, University of Maryland, College Park, Maryland, 20742*
E-mail: cf25@umail.umd.edu
[2]*Université Libre de Bruxelles, Laboratoire de Génétique et Ecologie végétales, 1850 Chaussée de Wavre, B-1160 Brussels, Belgium*
[4]*E-mail: xvekema@ulb.ac.be*
[5]*E-mail: ohardy@ulb.ac.be*

Abstract.—An extensive allozyme survey was conducted within a natural ''meta'' population of the native North American annual legume, *Chamaecrista fasciculata* (Leguminosae) to quantify genetic structure at different spatial scales. Gene flow was then estimated by a recently developed indirect method based on a continuous population model, using pairwise kinship coefficients between individuals. The indirect estimates of gene flow, quantified in terms of neighborhood size, with an average value on the order of 150 individuals, were concordant among different spatial scales (subpopulation, population, metapopulation). This gene-flow value lies within the range of direct estimates previously documented from observations of pollen and seed dispersal for the same metapopulation. Monte Carlo simulations using the direct measures of gene flow as parameters further demonstrated that the observed spatial pattern of allozyme variation was congruent with a model of isolation by distance. Combining previously published estimates of pollen dispersal distances with kinship coefficients from this study, we quantified biparental inbreeding relative to either a single subpopulation or the whole metapopulation. At the level of a neighborhood, little biparental inbreeding was observed and most departure from Hardy-Weinberg genotypic proportions was explained by self-fertilization, whereas both selfing and biparental inbreeding contributed to nonrandom mating at the metapopulation level. Gene flow was also estimated from indirect methods based on a discontinuous population structure model. We discuss these results with respect to the effect of a patchy population structure on estimation of gene flow.

Key words.—Gene flow, inbreeding, inbreeding coefficient, isolation by distance, neighborhood size, population structure, Wright's *F*-statistics.

Received February 4, 2002. Accepted January 16, 2003.

Gene flow determines the scale of local adaptation (Endler 1977) and the role of population structure in the evolutionary process (Wright 1977; Wade 1992; Fenster et al. 1997). Therefore, much effort has been devoted to the estimation of the magnitude of gene flow using direct or indirect methods (Slatkin 1985). Direct methods quantify the movement of genes by monitoring vectors of gene flow (Kerster and Levin 1968; Beattie and Culver 1979; Fenster 1991a), following the movement of marker genes (Schaal 1980; Fenster 1991a), or by performing paternity analyses (Ellstrand and Marshall 1985; Meagher 1986; Hamrick et al. 1995). Indirect estimation of gene flow (reviewed in Slatkin 1985) focuses on gene flow's impact on local differentiation for neutral genetic markers. Indirect estimates are most frequently based on the relationship between either Wright's *F*-statistics (Wright 1951) or spatial autocorrelation parameters (Sokal and Wartenberg 1983) with migration rates (Slatkin 1993; Rousset 1997; Epperson 1990, 1995) as specified by particular models of population structure. Although direct methods in plants have been applied mostly at a very local scale, within continuously distributed populations (but see Hamrick et al. 1995), methods to infer gene flow from local genetic differentiation have been developed mainly for models of subdivided populations (Slatkin 1993).

Recently, several new indirect methods adapted to a continuously distributed population have been proposed to quantify the extent of gene dispersal (Tufto 1996; Epperson and Li 1997; Hardy and Vekemans 1999; Rousset 2000). Some of these new indirect methods consist of regressing a measure of genetic distance (Rousset 2000) or relatedness (Hardy and Vekemans 1999) on the spatial distance between individuals. These approaches are derived from an analytical model of isolation by distance demonstrating that the genetic distance between populations, as measured by pairwise $F_{ST}/(1 - F_{ST})$ ratios, increases approximately linearly with the spatial distance in one-dimensional space, and with its logarithm in two-dimensional space (Rousset 1997). This method was later adapted to the individual level to infer gene dispersal distances among individuals (rather than among populations) using a genetic measure of interindividual distance (Rousset 2000) or relatedness (Hardy and Vekemans 1999). The ''relatedness'' approach uses statistics similar to those of spatial autocorrelation analysis (Hardy and Vekemans 1999). In sum, the indirect gene flow estimates based on interindividual distance, deduced from the rate of change of genetic distance/relatedness between individuals with the spatial distance, can be expressed in term of Wright's neighborhood size. Thus, with these methods, it is possible to compare direct and indirect estimates of the same parameter (e.g., neighborhood size) in the context of isolation by distance.

Direct and indirect methods to estimate gene flow can complement each other by providing opportunities to estimate parameters that are difficult to assess with either method alone. For instance, direct methods estimate the distribution of distances between mates (pollen dispersal probabilities)

[3] Present address: Laboratorie de Génétique et Evolution des Populations Végétales, UMR CNRS 8016, Universite de Lille 1, F-59655 Villeneuve d'Ascq, France; E-mail: xavier.vekemans@univlille1.fr.

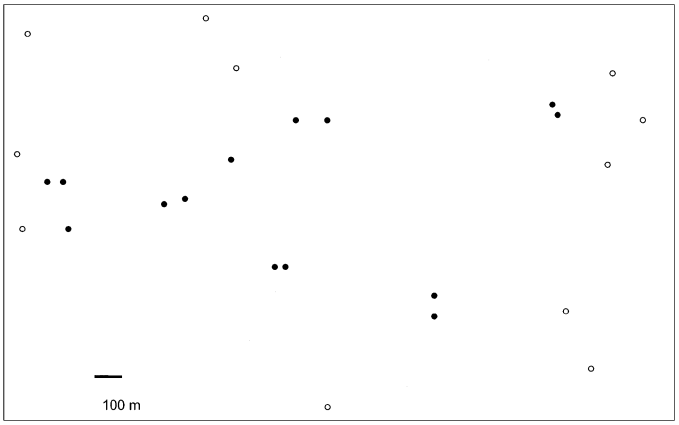

FIG. 1. Distribution of populations of *Chamaecrista fasciculata* in the Gooselake Prairie metapopulation. Filled circles and unfilled circles refer to populations included and not included, respectively, in the allozyme survey. Genotypes were sampled at either end of 12 of the 14 populations. These groups of individuals sampled at either end of the population are referred to as subpopulations. The bar in the lower left-hand corner of the figure is the scale and corresponds to 100 m.

and indirect methods provide the distribution of relatedness between individuals according to the spatial distance. Putting both distributions together, we will demonstrate that it is possible to estimate the average relatedness between mates, and thus assess the contribution of biparental inbreeding (i.e., the inbreeding associated with mating between relatives, selfing excluded) to the overall-inbreeding coefficient of descendant individuals. The most frequent method used to quantify biparental inbreeding relies on the difference in outcrossing rates based on single versus multilocus estimates (Ritland and Jain 1981). There are two advantages of our combined approach. First, biparental inbreeding is quantified in terms of an *F* (kinship or inbreeding) coefficient, and second, we can directly quantify the spatial scale that contributes most to inbreeding.

Gene dispersal is limited in the patchily distributed annual North American native legume, *Chamaecrista fasciculata*, as assessed by direct methods that quantified pollinator movement, pollen dispersal dynamics and seed dispersal (Fenster 1991a). Weighting the dispersal data by the relative probability of gene establishment (Fenster and Sork 1988; Fenster 1991b, c) substantially increased the magnitude of gene flow. However, estimates of gene flow were still limited, with neighborhood sizes estimated on the order of 100–166 individuals over four years and a harmonic mean of 120 individuals based on the yearly variation in population size. Preliminary analyses based on an allozyme survey (Fenster 1988; Fenster and Dudash 1994) indicated substantial population genetic structure, congruent with extremely limited gene flow in this species. Here we greatly expand the analyses of spatial patterns of allozyme genetic variation to (1) obtain indirect estimates of localized gene dispersal distances (neighborhood size) and compare them with previously published direct estimates (Fenster 1991b); (2) combine information from direct and indirect methods to estimate the contribution of biparental inbreeding to the overall inbreeding coefficient of descendant individuals; and (3) assess the consistency between gene dispersal estimates from indirect approaches at different spatial scales, to evaluate the significance of spatial discontinuities to genetic differentiation.

MATERIALS AND METHODS

Study Site

We quantified gene flow in *Chamaecrista fasciculata* using both direct and indirect approaches at Gooselake Prairie Nature Preserve (GLP), Grundy County, Illinois. GLP is a 700 ha disturbed mesic prairie located on the flood plain of the Illinois River, approximately 130 km southwest of Chicago. *Chamaecrista fasciculata* is patchily distributed at GLP. Given the spatial discontinuities of its distribution at GLP, the greater amount of gene flow within patch versus between patches (Fenster 1991a) and frequent extinction and colonization of patches (Fenster 1991b, pers. obs.) we refer to *C. fasciculata* at GLP as a metapopulation. The metapopulation at the time of the study was subdivided into approximately 25 discrete populations (patches), each separated from one another by 50–200 m. Each of the 25 populations consisted of a continuous distribution of 100s–1000s of adult flowering individuals. From these, 14 populations were chosen for study using a stratified random sampling method (Fig. 1). The clustering of the 14 sampled populations into pairs and in the interior of GLP reflected an attempt to sample genotypes on a roughly exponential scale of increasing distance from one another and to sample within the more pristine prairie habitiat, perhaps corresponding to more natural processes of gene flow.

Study Organism

Chamaecrista fasciculata Michx., partridge pea (Fabaceae), is a highly outcrossing (mean outcrossing rate = 80%, Fenster 1991a), self-compatible annual legume of old field, disturbed prairie, and savanna. It is distributed from southern New England to Florida and westward to Texas and Kansas and Minnesota in the eastern United States. Pollination is almost exclusively by large bees (Lee and Bazzaz 1982; Fenster 1991a). Seed dispersal, through explosive pod dehiscence, is limited to several meters (Fenster 1991a). To reiterate, a direct estimate of gene flow was quantified in terms of neighborhood size, resulting in an estimated neighborhood area corresponding to a circle of radius of 3–4 m and consisting of 120 individuals (Fenster 1991a, b).

Sampling

The following hierarchical design was used to determine the spatial pattern of genetic variation among and within the 14 sampled populations of *C. fasciculata* within GLP. Leaf tissue from 13 to 114 adult individuals (range determined by density of individuals) was collected at either end of each of the 14 populations from quadrats with areas of 20 to 100 m^2. Quadrat size was chosen based on preliminary observations of pollinator flight movement and corresponded to both a priori expectations and final estimates of neighborhood area (Fenster 1991a). Henceforth, quadrats will be referred to as subpopulations, which are the lowest spatial subdivision of sampling. Subpopulation size variation reflected variation in

plant density and an expected inverse relationship of pollinator flight distance. Hence neighborhood area is expected to be negatively correlated with plant density (Bateman 1947; Levin and Kerster 1969a, b; Schmitt 1983), which was observed at our study site (Fenster 1991a). Thus, in populations of higher plant density, smaller subpopulation areas were used. The distance between the closest edges of each subpopulation within a population was 40 m. In two of the populations, individuals were sampled from only one end or one subpopulation. The resulting number of subpopulations totaled 26 (12 populations × 2 subpopulations/population + 2 populations × 1 subpopulation/population). For 21 of the 26 subpopulations, all individuals were mapped to the nearest 10 cm. Altogether, leaf tissue from 1729 individuals was collected. Of these, 1365 were precisely mapped to the nearest 10 cm, the remaining 364 individuals were assigned to a position in the middle of the subpopulation. All leaf tissue was collected on ice during August and September 1984, and stored in a freezer at −76°C until protein extraction.

Genetic Markers

Samples were prepared using a modification of the phosphate grinding buffer-PVP solution in Soltis et al. (1983): 0.01 M germanium dioxide, 10% DMSO, and 0.5% 2-phenoxyethanol (Kelley and Adams 1977). Starch (from Sigma) gel concentrations varied from 10.6–11% depending on the lot. The following six enzymes, representing Mendelian, codominant loci, were surveyed and used in the analyses presented here: aconitase (Acn, two loci), diaphorase (Dia, one locus), leucine amino-peptidase (Lap, one locus), phosphoglucomutase (Pgm, one locus), and 6-phosphogluconate dehydrogenase (6-Pgd, one locus). Acn and Dia were resolved on a modified gel and electrode buffer designated system 8 of Soltis and Soltis (1987) and Pgm, 6-Pgd and Lap were resolved on system 11 of Soltis et al. (1983). Known heterozygous standards were run on every gel for each enzyme system to ensure proper interpretation of the banding patterns. A photographic record of every gel was made to ensure that rare alleles could be identified. The inheritance of all loci was verified either by controlled crosses or by observation of the segregation of banding patterns within family arrays. Staining methods for all enzymes followed Soltis et al. (1983), except for Lap, which followed Soltis and Soltis (1987).

Data Analyses

To check whether the different loci could be considered as providing independent replicates of population structure, linkage disequilibrium between all pairs of loci within each subpopulation and across all subpopulations was tested using the computer program GENEPOP 3.2a (Raymond and Rousset 1995). This program was also used to compute individual inbreeding coefficients (F_I) at the subpopulation, population, and metapopulation levels following the ANOVA approach of Weir and Cockerham (1984).

We then characterized the genetic structure of the metapopulation at different spatial scales, first by a hierarchical decomposition of genetic variance among subpopulations and populations (hierarchical *F*-statistics), second by two different spatial analyses where pairwise kinship estimates between individuals, or pairwise genetic distance measures between subpopulations, were compared to geographical distance. Under the theoretical framework of isolation by distance, these spatial analyses provide indirect estimates of gene dispersal distances (neighborhood area sizes). Combined with published direct estimates of pollen dispersal distances, pairwise kinship estimates were also used to assess the contributions of biparental inbreeding and selfing to the overall inbreeding. All of these analyses are detailed in turn below.

Hierarchical F-statistics

The partitioning of genetic variance within and among subpopulations and populations was assessed using *F*-statistics (Weir and Cockerham 1984) in the context of an island model, computed with the software ARLEQUIN (Schneider et al. 2000). The average inbreeding coefficient within subpopulations (F_{IS}) was computed, and the proportion of genetic variance was determined for the following components: among subpopulations within a population (F_{SP}), among populations within the total Gooselake Prairie metapopulation (F_{PT}), and among subpopulations independently of populations (F_{ST}). Populations with only one subpopulation sampled were removed from this analysis. Thus, the analysis was carried out on the 12 populations each containing two subpopulations. Significance levels of the statistics were determined with appropriate permutation tests (Schneider et al. 2000).

Spatial analysis of genetic differentiation between (sub)populations

Under isolation by distance in a two-dimensional space, pairwise $F_{ST}/(1 - F_{ST})$ ratios between (sub)populations are expected to increase approximately linearly with the logarithm of spatial distance (Rousset 1997). Therefore, $F_{ST}/(1 - F_{ST})$ ratios were computed for each pair of the 26 subpopulations using the computer program GENEPOP 3.2 (Raymond and Rousset 1995). Pairwise $F_{ST}/(1 - F_{ST})$ ratios were regressed on the logarithm of the spatial distance, providing a regression slope (*Blog*). The significance of the *Blog* was tested by a randomization procedure whereby subpopulations were permuted among locations 10,000 times to assess the distribution of *Blog* values under the null hypothesis of no correlation between geographic and genetic distances. *P*-values were estimated as the proportion of this distribution lying higher than the observed *Blog* (equivalent to performing a Mantel test). The same analysis was carried out at the population level, computing pairwise $F_{ST}/(1 - F_{ST})$ ratios for each pair of the 14 populations.

Spatial analysis of kinship coefficients between individuals

Each subpopulation, each population, and the metapopulation as a whole were analyzed separately by a method related to spatial autocorrelation, where relative kinship coefficients were computed between all pairs of individuals. The relative kinship coefficients are defined as $f_{i,j} \equiv (Q_{i,j} - \bar{Q})/(1 - \bar{Q})$, where Q represents the probabilities of identity in state between genes: $Q_{i,j}$ for random genes between individuals i and j, and $\bar{Q}$ for random genes within the reference

population sampled (either the metapopulation, a population, or a subpopulation). The pairwise kinship coefficients thus measure the correlation in the frequencies of homologous alleles between two individuals and were computed in a way similar to Loiselle et al. (1995) and Kalisz et al. (2001): for each allele and each pair of individuals, i and j, $F_{i,j} = \hat{f}_{i,j} = (p_i - \bar{p})(p_j - \bar{p})/(\bar{p}(1 - \bar{p})) + 1/2(n - 1)$, where p_i, p_j are the allele frequencies of i and j (taking the following possible values: 0, 0.5, 1), $\bar{p}$ is the average allele frequency of the reference population, and n is the sample size used to estimate $\bar{p}$. The second term corrects for a sample bias effect, ensuring that the average kinship coefficient over all pairs of individuals equals zero. Average multiallelic and multilocus estimators were obtained by weighting the $F_{i,j}$ values per allele by $\bar{p}(1 - \bar{p})$.

We used two related approaches, based upon analysis of kinship, to characterize the spatial genetic structure at different hierarchical scales. First, average kinship coefficients were computed for the following distance classes as in a spatial autocorrelation analysis: 0–0.75, 0.75–1.25, 1.25–1.75, 1.75–2.25, 2.25–3, 3–4, 4–5, 5–20, 20–50, 50–100, 100–200, 200–500, 500–1000, 1000–2000 m (not all distance classes were represented when the analysis was carried out within a population or subpopulation). Thus, the spatial genetic structure can be characterized without a priori assumptions on its pattern. Second, the regression slopes (*blog*) of pairwise kinship coefficients on the logarithm of pairwise geographical distances between individuals were estimated. Note that *b* and *B* refer to the relationship of distance with individuals and subpopulations, respectively. Under isolation by distance in a two-dimensional space, kinship is expected to decrease approximately linearly with the logarithm of the spatial distance (Maruyama 1977; Rousset 1997; Hardy and Vekemans 1999), at least within a restricted distance range ($\approx\sigma$ to 20σ, where σ is the axial standard deviation of gene dispersal distances, Rousset 1997, 2000). Hence, the regression slope approach assumes a priori that the spatial genetic structure results from isolation by distance in our study metapopulation. When $F_{i,j}$ values actually decrease linearly with distance, the regression slope provides a single parameter characterizing the extent of spatial genetic structuring. As explained below, regression slopes can also be used to estimate neighborhood sizes, but should then be computed over the restricted distance range defined above to avoid biased estimators. Thus, we will refer to ''truncated'' and ''global'' regressions when $F_{i,j}$ values are regressed over a restricted distance range or the whole distance range, respectively. For this purpose, the σ considered to define the truncated distance range is the direct estimate made by Fenster (1991a,b): $\sigma \approx$ 1.5 m. We used a wider range, that is, 1.5–60 m (σ–40σ), to include most pairs of individuals belonging to different subpopulations within a given population.

Approximate confidence intervals for average kinship coefficients per distance class and the regression slopes were obtained as twice the standard error estimates, as calculated by a jackknife procedure over loci. Because only six loci are available, these confidence intervals should be interpreted with caution. Consequently, we also conducted randomization tests, consisting of 10,000 permutations of individuals among locations to check the significance of the average kinship coefficients per distance class (Kalisz et al. 2001) and of the regression slopes (as described above for the $F_{ST}/(1 - F_{ST})$ ratios). For analyses carried out within (sub)populations, individuals were permuted only with those from the same (sub)population.

These analyses were carried out as follows: (1) at the metapopulation level, considering the overall set of 1365 fine-scale mapped individuals, irrespective of their subpopulation or population; (2) within each of eight populations for which individuals were mapped in two subpopulations; and (3) within each of 20 subpopulations mapped at the individual level (one subpopulation containing only 13 individuals was not used). At the subpopulation level, loci with less than 20 individuals genotyped because of missing values were discarded to avoid substantial bias in the estimators resulting from small sample size (Ritland 1996). All analyses were performed with the software SPAGeDi developed by O. Hardy and X. Vekemans (E-mail: ohardy@ulb.ac.be).

To assess if the extent of spatial genetic structure varied with respect to a particular enzymatic locus or subpopulation, nonparametric Kruskal-Wallis ANOVA by ranks tests were performed on the F_{IS} and *blog* values obtained per locus (using subpopulations as replicates), and per subpopulation (using loci as replicates). In these tests, a statistic H is computed which is the among-groups variance of the sums of ranks, and is tested against a Chi-square distribution with $a - 1$ degrees of freedom according to Sokal and Rohlf (1995), where a is the number of groups compared (either loci or subpopulations).

Indirect estimates of gene dispersal distances

Assuming isolation by distance, the extent of gene dispersal can be quantified by the product $D\sigma^2$, where σ^2 is the variance of gene displacements (half the mean squared physical distance between parent and offspring), and D is the effective density of individuals. As a first approximation, D is the product of the density with the ratio Ne/N where Ne and N are the effective and census population sizes, respectively (Crawford 1984). In the context of a two-dimensional continuous population, $Nb \equiv 4\pi D\sigma^2$ can be interpreted as a neighborhood size in terms of numbers of individuals, expressing the strength of local genetic drift according to Wright (1943). According to a theoretical analysis of the isolation by distance model (Rousset 1997) and its extension to a continuous population model (Hardy and Vekemans 1999), indirect estimates of $D\sigma^2$ can be obtained from the regression of the pairwise $F_{i,j}$ values between individuals on geographical distance: $Nb = -(1 - F_I)/blog$, where F_I is an estimate of the inbreeding coefficient relative to the reference population considered. This relationship holds best within the distance range σ–20σ, and is valid only in the absence of selfing. When selfing occurs, a good approximation is $Nb = -(1 - F_{(1)})/blog$, where $F_{(1)}$ is the the average kinship coefficient between adjacent individuals (F. Rousset, pers. comm.; O. Hardy, unpubl. data). Based on analyses either at the subpopulation, population, or metapopulation level, and considering either a global or a truncated (i.e., within 1.5–60 m, corresponding roughly to σ–40σ) regression for each level, we computed the slope *blog*, and quantified six indirect

estimates of neighborhood size. Note that expected values for kinship or inbreeding coefficients depend on a sample (their definition includes $\bar{Q}$) and are thus scale dependent. However, $\bar{Q}$ vanishes for the expression of the ratios $-(1 - F_I)/blog$ or $-(1 - F_{(1)})/blog$. Thus, these ratios do not depend on the sampling scale as long as *blog* is computed on the same distance range.

In the context of a subdivided population, $4\pi D\sigma^2$ should not be interpreted as a neighborhood, and the immigration rate per subpopulation (m) is at least as important as $D\sigma^2$ in characterizing gene flow (Rousset 2001a). Indirect estimates of $D\sigma^2$ can nevertheless be obtained from the regression of the pairwise $F_{ST}/(1 - F_{ST})$ values between subpopulations on geographical distance: $4\pi D\sigma^2 = 1/Blog$ (Rousset 1997). Two estimates of $D\sigma^2$ were obtained from analyses of differentiation among either subpopulations or populations.

Assessment of biparental inbreeding and selfing rate

Biparental inbreeding represents the impact of matings between related individuals on the inbreeding coefficient, exclusive of selfing. It can be quantified by the inbreeding coefficient of truly outcrossed individuals (individuals resulting from cross-pollination events), a quantity we call fx (the *biparental inbreeding coefficient*). In terms of probabilities of identity in state of genes, we define it as $fx \equiv (Qo^* - \bar{Q})/(1 - \bar{Q})$, where Qo^* represents identity between homologous genes within outcrossed individuals. fx is relative to the average identity of genes in the reference population and thus may differ among subpopulation, population and metapopulation levels. Knowing the distribution of pollen dispersal distances for cross-pollinations, $P(d)$, and having estimates of relative kinship coefficients between individuals according to distance, $F(d)$, it is possible to compute an estimator of fx (called Fx), by noting that the inbreeding coefficient of an individual is equal to the kinship coefficient between its parents: $Fx = \int_{d>0}^{\infty} F(d) \times P(d)\, \delta d$. Fenster (1991a,b) assessed the distribution of pollen dispersal distances, $P(d)$, for *C. fasciculata* in GLP. The distribution of *effective* pollen dispersal distances (i.e. accounting for inbreeding depression) followed an approximate normal distribution with a standard deviation of 3.11 m. As $F(d)$ was assessed at three scales (within a subpopulation, within a population, and at the whole metapopulation level), biparental inbreeding coefficients were computed at each scale, summing the products $F(d) \times P(d)$ over the all discrete distance intervals d.

Biparental inbreeding estimates can be used to obtain indirect estimates of the selfing rate. Both selfing and biparental inbreeding can contribute to positive inbreeding coefficients. When selfing is the sole contributor, the selfing rate, s, may be estimated using Wright's (1951) formula: $\hat{s} = 2F_I/(1 + F_I)$. However, Wright's (1951) formula provides an overestimate in the presence of biparental inbreeding. As shown in the appendix, a corrected estimator can be obtained using an estimate of the biparental inbreeding coefficient: $\hat{s} = 2(F_I - Fx)/(1 + F_I - 2Fx)$. This formula was applied to estimate the selfing rate at the subpopulation and metapopulation levels.

Simulations

Monte Carlo simulations were used to check if the spatial genetic structure observed within subpopulations was consistent with expectations for neutral markers using the gene dispersal parameters measured by direct methods (Fenster 1991a,b). To mimic the processes occurring within each subpopulation, we simulated a lattice model (Hardy and Vekemans 1999) having the form of a square population with 25 $\times$ 25 individuals. For spatial analysis, individuals were sampled within the 15 $\times$ 15 central square to avoid border effects. From Fenster (1991a,b), we considered the following gene dispersal parameters: effective selfing rate of 10% (i.e., accounting for the lower fitness of inbred individuals), standard deviations of seed and pollen displacements of 0.31 m and 2.2 m, respectively, and mean effective density (harmonic mean) of 4.34 individuals/m^2. With this density, the average distance between adjacent ''effective'' individuals was 0.48 m, so that standard deviations of seed and pollen displacements used in our simulation model were 0.64 and 4.58 lattice units, respectively (one lattice unit being the distance between adjacent individuals). The simulation model assumed a normal distribution of pollen displacements and a leptokurtic distribution for seed displacements, in accordance with the observations made by Fenster (1991a,b). Because the available positions are discrete in this lattice model, the continuous pollen and seed dispersal distributions are transformed into discrete ones, choosing the discrete position closest to the pointing continuous dispersal vector. No migrant seeds or pollen arrived from outside. Each simulated individual was characterized at one diallelic locus with alleles randomly distributed initially. The simulation was run for 20 generations, a sufficient time to reach a quasi-equilibrium genetic structure (Hardy and Vekemans 1999). Spatial genetic structure was then assessed as described above to obtain F_I, $F(d)$ and *blog* values. The simulations were run a thousand times to obtain accurate average values.

RESULTS

Allozyme Polymorphism and Level of Inbreeding

Over the sample of 1727 individuals, we found two alleles at locus Dia, three alleles at Acn-2, four alleles at Acn-1, Lap, and 6-Pgd, and five alleles at Pgm. Tests of linkage disequilibrium for each locus pair within each subpopulation were all nonsignificant after Bonferroni correction for multiple tests (Rice 1989), and tests across subpopulations were also nonsignificant using Fisher's method (Raymond and Rousset 1995). Thus, the different loci can be considered as providing independent information on population structure. Measures of allozyme polymorphism are reported in Table 1 at the subpopulation, population, and whole GLP metapopulation levels. The average inbreeding coefficient (F_I) within subpopulations was 0.090 $\pm$ 0.015 (SE over loci), whereas slightly higher values are found at the population and metapopulation levels (Table 1). Values of the inbreeding coefficient within subpopulations are given for each locus in Table 2. We obtained significantly positive inbreeding coefficients at the subpopulation level for all but one locus. No significant differences were observed among subpopulations

TABLE 1. Allozyme polymorphism within subpopulations and populations of a *Chamaecrista fasciculata* metapopulation from Gooselake Prairie Preserve: *A*, number of alleles per locus; H_o, observed heterozygosity; H_e, gene diversity; F_I, inbreeding coefficient. Statistics are given as multilocus estimates averaged over replicates, and standard errors are within parentheses.

Level of analysis	*A*	H_o	H_e	F_I
Within subpopulations	2.52 (0.27)	0.261 (0.052)	0.279 (0.041)	0.090 (0.015)
Within populations	2.67 (0.25)	0.256 (0.032)	0.282 (0.029)	0.098 (0.016)
Metapopulation	3.67	0.257	0.291	0.121

(Kruskal-Wallis ANOVA by ranks test: $H = 29.05$, $P = 0.218$, df = 24, $n = 141$), nor among loci ($H = 9.9$, $P = 0.076$, df = 5, $n = 141$).

Differentiation among Subpopulations

Differentiation among subpopulations was low but highly significant (mean $F_{ST} = 0.034$; Table 2), with a higher contribution due to differentiation among subpopulations within population (mean $F_{SP} = 0.022$) than to differentiation among populations (mean $F_{PT} = 0.013$), but both components are statistically significant for the multilocus averages (Table 2).

$F_{ST}/(1 - F_{ST})$ ratios for pairs of subpopulations increased linearly with the logarithm of the geographical distance (Fig. 2), demonstrating a typical pattern of isolation by distance (Rousset 1997). The regression of $F_{ST}/(1 - F_{ST})$ on the logarithm of the distance gives a slope *Blog* = 0.0057 (jackknife over loci: SE = 0.0040; Mantel test: $P = 0.011$). There is much scatter of these ratios around the regression line due to the high sampling variance inherent to pairwise $F_{ST}/(1 - F_{ST})$ ratios, so that the amount of variance explained by the regression analysis is small ($r^2 = 0.034$). Note that there is no abrupt increase in $F_{ST}/(1 - F_{ST})$ ratios when switching from comparisons between pairs of subpopulations within a population (+ symbol) to comparisons between pairs from different populations (× symbol). The regression of $F_{ST}/(1 - F_{ST})$ ratios for pairs of populations on distance also gives a significant slope: *Blog* = 0.0074 (jackknife over loci: SE = 0.0059; Mantel test: $P = 0.021$; $r^2 = 0.081$). In our case, jackknifing is probably not as reliable as randomization to determine statistical significance of *Blog* for two reasons. Jackknifing over just six loci likely does not provide as accurate an estimate of the standard errors and using 2 SE to define a 95% confidence interval is even less accurate because the distribution is not exactly normal. Thus, randomization is much more reliable as it provides an exact test of the departure from random spatial distribution of genotypes and is the method by which we determine statistical significance.

Spatial Analyses of Kinship Coefficients between Individuals

Within the distance range 1–200 m, the average kinship coefficients between pairs of individuals demonstrate an approximate linear relationship with the logarithm of the spatial distance (Figs. 3–5). Below 1 m, somewhat higher kinship coefficients are observed. Beyond 200 m, kinship coefficients do not seem to decrease any longer. Within the 20 subpopulations examined, the average kinship coefficients were significantly positive at the first distance class (0–0.75 m; $P < 0.001$), and significantly negative at the last distance class (5–20 m; $P = 0.042$). The global regression slope (i.e., over the full distance range) was significantly negative: *blog* ± SE (jackknife over loci) = −0.0055 ± 0.0009 ($P < 0.001$). The variance explained by the regression line is very weak (average $r^2 = 0.0004$), as expected from the very high sampling variance for pairwise kinship coefficients (Ritland 1996). The truncated regression slope (i.e., within the distance range 1.5–60 m) does not significantly differ from zero: *blog* ± SE = 0.0004 ± 0.0037 ($P = 0.9$), suggesting that near-neighbor relatedness is largely responsible for the global regression slopes within subpopulations.

Within population, both global and truncated average regression slopes were significantly negative: *blog* ± SE = −0.0072 ± 0.0024 ($P < 0.001$; $r^2 = 0.0016$), and *blog* ± SE = −0.0065 ± 0.0026 ($P < 0.001$; $r^2 = 0.0009$), for global and truncated regression, respectively. Average kinship coefficients were significantly positive for distance classes >0–0.75 m ($P < 0.001$), >0.75–1.25 m ($P < 0.001$), >1.75–2.25 m ($P < 0.001$), >2.25–3.0 m ($P < 0.023$), >3.0–4.0 m ($P < 0.033$), >5.0–20.0 m ($P < 0.027$), and significantly negative for the >20.0 m distance class ($P < 0.001$), but distance classes 1.25–1.75 m and 4–5 m were not significantly different from zero.

At the metapopulation level, regression slopes were also significantly negative: *blog* ± SE = −0.0044 ± 0.0013 ($P < 0.001$; $r^2 = 0.0006$), and *blog* ± SE = −0.0064 ± 0.0021

TABLE 2. *F*-statistics at different hierarchical levels for 12 populations of *Chamaecrista fasciculata* in Gooselake Prairie Preserve, each containing two subpopulations. Tests of significance (null hypothesis: value = 0): ns, non significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

Locus	Sampling level Reference level F_{IS}	Individuals Subpopulations F_{SP}	Subpopulations Populations F_{PT}	Populations Metapopulation F_{ST}
Acn-1	0.080***	0.026***	0.006 ns	0.032***
Acn-2	0.119***	0.016**	0.003 ns	0.019***
Dia	0.001 ns	0.007*	0.016*	0.023***
Lap	0.116***	0.031***	0.005 ns	0.036***
Pgm	0.098***	0.031***	0.029 ns	0.060***
6-Pgd	0.114***	0.015***	0.002 ns	0.017***
Multilocus	0.090***	0.022***	0.013*	0.034***

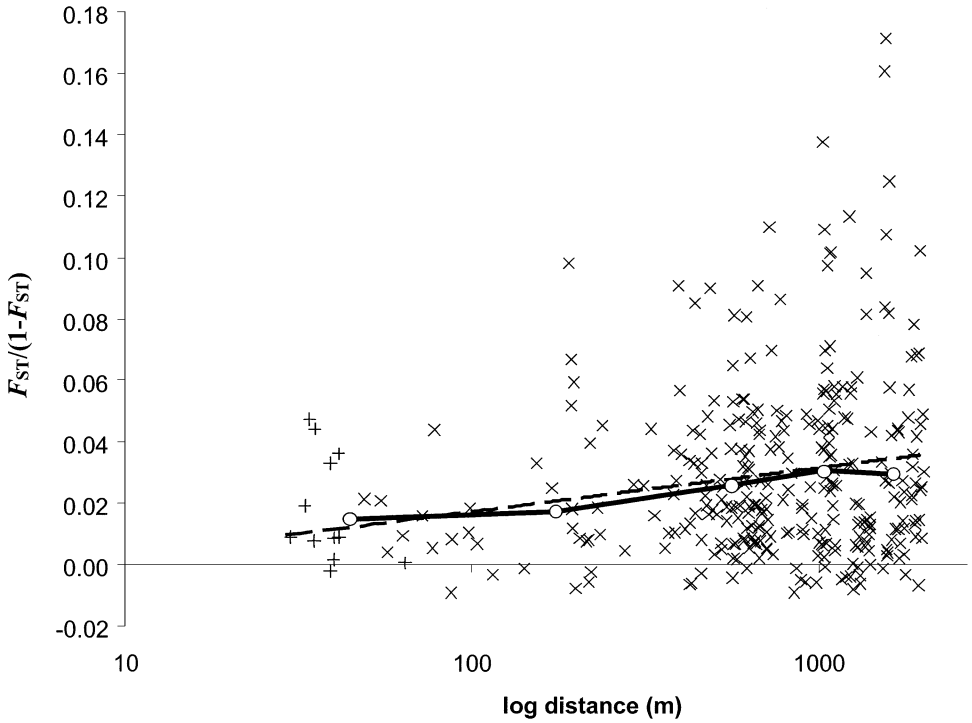

FIG. 2. $F_{ST}/(1 - F_{ST})$ ratios between pairs of 26 subpopulations of *Chamaecrista fasciculata* within Gooselake Prairie (+ symbols for comparisons within population, × symbols for comparisons between populations), regression line (stippled line), and average values for five distance classes (open circles).

($P < 0.001$; $r^2 = 0.0010$), for global and truncated regression, respectively. Average kinship coefficients were significantly positive for each distance class through 75 m ($P < 0.001$).

The *blog* values obtained by the global regression for each locus and each subpopulation did not differ significantly among subpopulations (Kruskal-Wallis ANOVA by ranks test: $H = 21.274$, $P = 0.322$, df = 19, $n = 111$) nor among loci ($H = 0.57$, $P = 0.98$, df = 5, $n = 111$). This suggests that average estimates over all subpopulations are meaningful.

Indirect Estimation of Neighborhood Size

Different estimates of neighborhood size relying on isolation by distance models were obtained according to the spatial scales of genetic differentiation investigated (Table

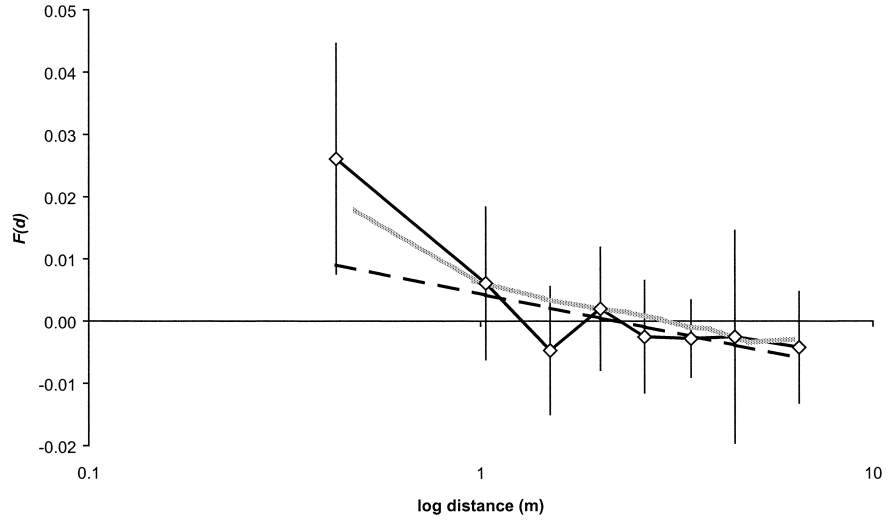

FIG. 3. Observed and simulated kinship coefficients within subpopulations of *Chamaecrista fasciculata* individuals in Gooselake Praire. Full line with diamond symbols: observed average over six loci and 20 subpopulations and confidence intervals. Stippled line: global regression line. Gray line: average values over 1000 replicates of simulations.

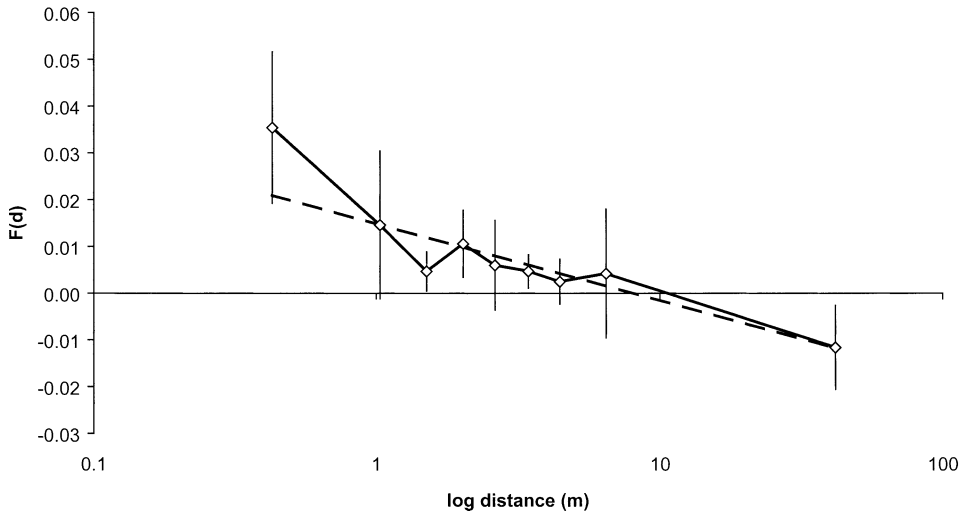

FIG. 4. Observed and simulated kinship coefficients within populations of *Chamaecrista fasciculata* individuals in Gooselake Praire. Full line with diamond symbols: observed average over six loci and eight populations and confidence intervals. Stippled line: global regression line.

3), that is, from the kinship coefficients between individuals either within subpopulations ($Nb = 177$ and ∞), within populations ($Nb = 134$ and 147), and within the whole metapopulation ($Nb = 214$ and 147), where the first and second estimates are based on global and truncated regressions regarding the distance range, respectively. With the exception of the estimate based on the truncated regression within subpopulations ($Nb = \infty$, corresponding to a nonnegative regression slope), these estimates are similar to one another. Their confidence intervals are widely overlapping (Table 3) and also encompass the direct *Nb* estimates previously obtained by Fenster (1991b), where *Nb* ranged from 100 to 166 across four years, with a harmonic mean of 120 individuals.

The confidence intervals are lower for the estimates based on the global rather than the truncated regression analyses, perhaps reflecting the loss of information in the latter estimate. However, following expectations from theory, the estimates based on global regression should suffer higher bias (Rousset 1997). Thus there appears to be a balance between bias and variance in these different estimates. Within subpopulations, 80% of individual pairs are individuals separated by less than 5 m, resulting in the truncated regression being based on a very narrow distance range, which may explain the large error observed for that *Nb* estimate. Thus, at this scale, no reliable estimate of *Nb* can be obtained, although we observe that the estimate based on the global regression is more realistic. In contrast, at the metapopulation level, the available distance range is very wide but bias rather than large standard errors may be a concern, so that the truncated regression is expected to give a better estimate at this scale. In conclusion, the best indirect *Nb* estimates should be obtained from the truncated regression analyses at the population or metapopulation level ($Nb = 147$ in both cases), which lie within the range of the direct *Nb* estimates.

The rate of change of pairwise differentiation ($F_{ST}/(1 - F_{ST})$ ratios) between subpopulations and between populations with distance also gives estimates of gene dispersal: $4\pi D\sigma^2 = 175$ and 135, respectively (Table 3). Here we do not refer to these estimates as neighborhood size because they are not in the context of a continuous population. Thus, σ^2 is expected to refer to gene dispersal among subdivisions rather than among individuals, and D would correspond to a global rather than local density. Nonetheless, these estimates are similar to the *Nb* estimates obtained at the local scale, as would be expected if individuals were distributed continuously throughout the whole metapopulation.

TABLE 3. Comparison of indirect estimates of gene dispersal across different scales of observation using global and truncated regression for populations and subpopulations of *Chamaecrista fasciculata.* Approximate confidence intervals are given under parentheses. These are computed as ±2 SE, where SE is the standard error of *blog* or *Blog* values estimated by jackknifing over loci (see text).

	Distance range of the regression analyses	
	Global regression ($d > 0$)	Truncated regression ($1.5 < d < 60$)
Nb estimates based on average $F(d)$ values		
Within subpopulation	177 (133–262)	∞ (139–∞)
Within population	134 (81–396)	147 (82–766)
Within metapopulation	214 (136–509)	147 (89–421)
$4\pi D\sigma^2$ estimates based on $F_{ST}/(1 - F_{ST})$ ratios		
Among subpopulations	175 (73–∞)	
Among populations	135 (52–∞)	

Simulation Results

The regression approach to estimate *Nb* is inappropriate at the subpopulation scale because it precludes analysis across

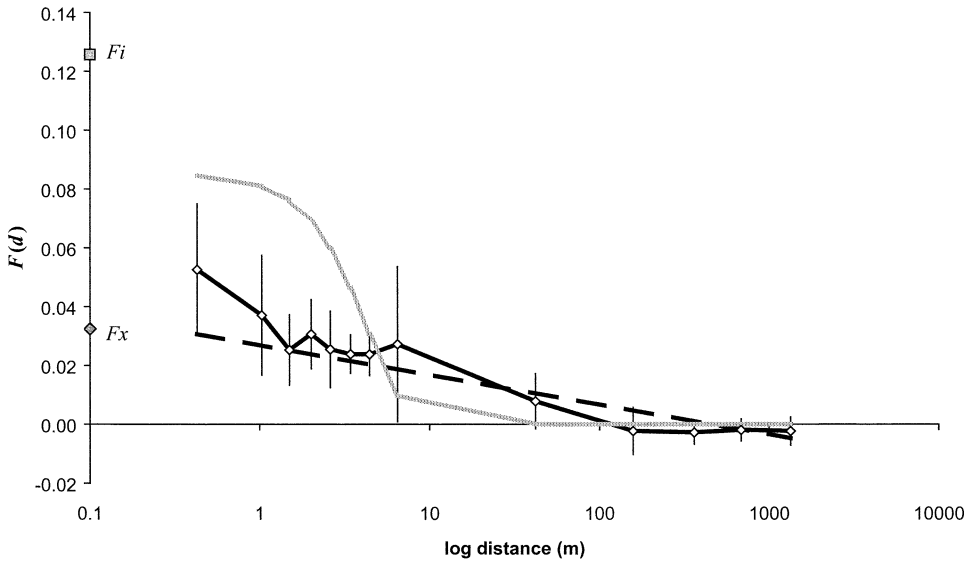

FIG. 5. Kinship coefficients between *Chamaecrista fasciculata* individuals within and across subpopulations in Gooselake Prairie (plain line with diamond symbols): mean values and confidence intervals (twice the SE assessed by jackknifing over the loci), global regression line based on all pairwise kinship coefficients between individuals (stippled line), and probability distribution of pollen dispersal distances (gray line). Values of the inbreeding coefficient (F_I) and the biparental inbreeding coefficient (*Fx*) are also shown on the vertical axis.

the optimal range of inter-individual distances ($\sigma - 20\sigma$). However, consistency between the observed genetic structure and direct gene dispersal estimates (including the details of the dispersal distribution and not just the σ^2) can be inspected at this scale through the simulation results (Fig. 3). A very good agreement between observed and expected spatial patterns of genetic structure is obtained. Kinship coefficients at the shortest distance class were higher than predicted from the regression line for both the simulated and observed curves (Fig. 3). This discrepancy from linearity is expected at very short distances because the rate of change of kinship coefficients for distances inferior to σ depends on the details of the dispersal distribution and not just on $D\sigma^2$ (Rousset 1997). Additional exploratory simulations have shown that the kinship curve bends towards higher values at short distances whenever seed dispersal is substantially lower than pollen dispersal (M. Heuertz, O. Hardy, X. Vekemans, unpubl. results), a situation encountered in *C. fasciculata* (Fenster 1991a).

Estimation of Biparental Inbreeding and Selfing Rates

The biparental inbreeding coefficient (*Fx*) is obtained by integrating the product $P(d) \times F(d)$ over distances, where $P(d)$ is the frequency distribution of pollen dispersal distances (excluding selfing) and $F(d)$ describes the relative kinship coefficients between individuals (Fig. 5). At the metapopulation level, $Fx = 0.033$ (SE = 0.007, jackknife estimate over loci). At the subpopulation level, considering the average $F(d)$ curve over all subpopulations, $Fx = 0.004$ (SE = 0.002).

In a genetically structured population, selfing rates can be estimated when both the inbreeding and biparental inbreeding coefficients are known. At the metapopulation level, the inbreeding coefficient was $F_I = F_{IT} = 0.121$ (SE = 0.019) (Table 1). Because *Fx* expresses the contribution of biparental inbreeding to the inbreeding coefficient, the difference between F_I and *Fx* is the contribution due to selfing. Using the formula developed in the appendix, the indirect estimate of the selfing rate at the metapopulation level was $\hat{s} = 0.167$ (SE = 0.029), whereas $\hat{s} = 0.216$ (SE = 0.030) when Wright's formula was used (i.e., neglecting biparental inbreeding). Similarly, at the subpopulation level, $F_I = F_{IS} = 0.090$ (SE = 0.018), and $\hat{s} = 0.159$ (SE = 0.032), whereas $\hat{s} = 0.165$ (SE = 0.031) when applying Wright's formula.

DISCUSSION

Comparison between Direct and Indirect Estimates

Indirect estimates of neighborhood size (Table 3) of *C. fasciculata* from GLP were consistent with the direct estimates obtained by Fenster (1991a,b). The slightly higher estimates of neighborhood size based on the indirect measures of gene flow may reflect an underestimate of long dispersal events using direct measures. Although the regression approach to estimate *Nb* gave less satisfactory results within subpopulation, simulations based on the gene flow parameters quantified by direct methods confirmed the consistency between spatial genetic structure and gene dispersal at this scale. These results suggest the regression approach will perform well if interindividual distances span a wide range of distances, such that the regression can be applied within an optimal range (approximately $\sigma - 20\sigma$). In the present study, σ was known from direct measures (Fenster 1991a). Without a priori knowledge of σ, one can first make an estimate of *Nb* from the global regression, deduce σ from knowledge of the density (from $Nb = 4\pi D\sigma^2$), estimate *Nb* again from a truncated regression, and repeat the process until convergence. For the present dataset, this approach converges quickly to a meaningful estimate when applied at the population and metapopulation scales, but it fails at the subpopulation scale (results not shown, but can be deduced from Table 3).

The overall consistency between direct and indirect estimates of gene dispersal suggests that the observed local genetic structure is representative of an equilibrium state, a condition of prime importance for valid indirect estimates (Whitlock and McCauley 1999). Thus, we conclude that limited gene flow is primarily responsible for the spatial pattern of genetic variation at the allozyme loci in *C. fasciculata*. In further simulations (Hardy, Vekemans, and Fenster, unpubl. data) we demonstrate that the same gene dispersal parameters may produce a wide range of single locus spatial structure that largely reflects variation inherent to the stochastic process of spatial structuring, which we refer to as stochastic variation. Hence, data from many loci and/or replicate samples are necessary to quantify gene flow parameters with indirect methods.

A similar good agreement between demographic and genetic estimates of neighborhood size was demonstrated by Rousset's (2000) reanalysis of Waser and Elliot (1991) and Waser (unpubl. data) genetic data for kangaroo rats (*Dipodomys spectabilis*). Earlier studies with snails also demonstrate a consistency between limited mobility and local differentiation in allele and phenotype frequencies (Jones et al. 1977; Selander 1975). In the several other studies that have compared direct with indirect estimates of gene flow (e.g., Campbell and Dooley 1992; Godt and Hamrick 1993), direct approaches have underestimated gene flow relative to the indirect approaches. The two approaches may not give identical results for several reasons. Direct methods estimate current gene flow, whereas indirect methods provide estimates of past gene flow or the average level of gene flow (Slatkin 1985). Indirect methods provide ''effective'' gene flow estimates, taking into account gene establishment, which is rarely quantified using direct methods (Levin 1981, but see Fenster 1991a,b,c). Rare long distance gene dispersal events may have substantial impact on the extent of local differentiation, but are difficult to detect with direct methods. Finally, indirect methods to estimate gene flow are based on several assumptions that may be violated in nature (e.g., equilibrium hypothesis, Whitlock and McCauley 1999). Consistency between direct and indirect estimates of gene flow may also depend on the geographical scale on which surveys quantifying patterns of genetic variation are conducted. At a large scale, the spatial patterns of genetic differentiation require more time to reach an equilibrium state, as compared to a local scale (Slatkin 1993; Hardy and Vekemans 1999), and the impact of past colonization or disturbance events may be substantial (Austerlitz et al. 1997).

Isolation by Distance at Different Spatial Scales

The spatial genetic structure that we quantified conforms to theoretical expectations for an isolation by distance process. Kinship coefficients decreased linearly with the logarithm of the distance within a distance range of σ to 20 σ (i.e., approximately 1.5–30 meters according to Fenster's 1991a, b direct σ estimate), as predicted by an isolation by distance process. At shorter distances, expected kinship coefficients depend on the details of the gene dispersal distribution (not only on its variance parameter σ^2), and at larger distances mutation rate may also have an effect (Rousset 1997).

Note that the distance at which the kinship curve crosses the axis (zero kinship) depends strongly on the scale of observation (1.5 m, 10 m, 150 m at the subpopulation, population, and metapopulation scales, respectively, Figs. 3–5). This occurs because kinship coefficients are relative to a given sample (the average kinship among all pairs of sampled individuals is set to zero). Thus negative coefficients can occur, reflecting that individuals are less related on average than random individuals from the sample. Thus, the distance of zero kinship does not characterize spatial genetic structure and depends mostly on the sampling scheme. However, spatial genetic structure is characterized by the rate of decrease of kinship with distance (slopes *blog*), which is essentially scale independent.

Isolation by distance was not limited to the levels of subpopulation or population, because we also observed a negative linear relationship between differentiation among (sub)populations and the log of distance. These results are consistent with a stepping stone model of very limited gene flow (Kimura and Weiss 1964) and correspond to the observation that interpopulation pollinator movement was limited to adjacent populations of *C. fasciculata* at GLP (Fenster 1991a). Others (Bos et al. 1986; Epperson and Clegg 1986; Knight and Waller 1987; Waser 1987) have also observed isolation by distance in populations that are continuously distributed but did not find that differentiation among populations was correlated with distance at higher geographic scales, where populations are apt to be patchily distributed. Studies focusing on larger geographic distances often fail to detect any association between geographic and genetic distance (e.g., Fischer et al. 2000; also summarized in Waser 1993). Where isolation by distance is detected across broader geographic scales, there is often only a weak association between geographic and genetic distances (Ritland 1989; Godt and Hamrick 1993; Wolf et al. 2000).

There are a number of possible explanations for why we were able to detect isolation by distance across all spatial scales, whereas many studies did not or only detected a weak association between genetic divergence and geographic distance. An important issue is that we sampled populations on a much smaller scale than most of the above studies. Hence, it is much less likely that the isolation by distance that we detect is a product of vicariance, that is, historical processes leading to geographic isolation of a group of populations, which was put forward as an alternative explanation for observed patterns of significant spatial genetic structure (Bossart and Prowell 1998). Other studies may have documented different patterns of gene flow operating at different spatial subdivisions, with gene flow following more of an island model at higher hierarchical levels. Furthermore, because genetic relatedness is expected to decrease roughly exponentially with distance under limited gene flow, the amount of differentiation among higher levels of population subdivision is expected to be smaller. Thus, the inability to detect a pattern of isolation by distance at higher spatial scales may reflect the limited extent to which we expect spatially associated divergence, a point also made by Rousset (1997, 2001b). We also expect a greater variance of expected di-

vergence at longer distance intervals because of drift effects (Hutchison and Templeton 1999).

Indirect *Nb* estimates based on small scale differentiation were close to the $4\pi D\sigma^2$ estimates based on large scale differentiation between (sub)populations (Table 3). Moreover, spatial discontinuities in the distribution of individuals were not associated with a significant increase in the level of genetic differentiation (Figs. 2, 5). In theory, such results would be expected if individuals were continuously distributed throughout the area, provided that long distance gene dispersal events were rare (Rousset 1997, pers. comm.). This result is not necessarily expected under a model in which individuals are grouped in discrete populations because $4\pi D\sigma^2$ estimates depend mostly on local density and within population gene dispersal at the within population scale, and on global density and interpopulation gene flow at larger scale. *Chamaecrista fasciculata* at GLP does not fit a continuous population (habitat is continuous but individuals are grouped), nor a classical model of discrete populations (groups of individuals are ephemeral and their potential locations are not fixed). Because we lack theoretical models predicting spatial genetic differentiation under isolation by distance in such a context, it is difficult to interpret the apparent congruence between $4\pi D\sigma^2$ estimates within and among subdivisions. Clearly, further empirical studies are needed to quantify metapopulation parameters, that is, population turnover rates, origin and relatedness of colonists (Whitlock and McCauley 1990; McCauley 1993).

Estimation of Biparental Inbreeding

Our analyses allow us to partition the factors causing the observed level of inbreeding (overall $F_{IT} = 0.121$). The largest amount of inbreeding is associated with the lowest hierarchical level, that is, the individual, as a result of self-fertilization. Indirect estimates of the selfing rate obtained at the subpopulation or metapopulation level are similar ($\hat{s} \approx 0.16$), and furthermore are consistent with direct estimates obtained by progeny genotype analysis (average $\hat{s} = 0.20$, 0.27, Fenster 1991a, 1995, respectively) and pollinator observations (average $\hat{s} = 0.14$, Fenster 1991a). The other source of inbreeding comes from the effect of matings between related individuals, that is, biparental inbreeding, at different levels. We quantified biparental inbreeding and found very low values within each subpopulation. Although limited seed dispersal of *C. fasciculata* likely causes the clumping of siblings within a meter, pollen dispersal patterns within subpopulations appear to result in most matings occurring between nonsiblings. This follows from the situation that adult density is high and descendants from different parents will overlap each other in space, corresponding to results documented by Kalisz et al. (2001) for a population of *Trillium grandiflorum*. At the level of the whole metapopulation, however, a higher value of biparental inbreeding was obtained. This is caused by most matings occurring between individuals of the same subpopulation, as suggested by the observation that differentiation among subpopulations ($F_{ST} = 0.034$) is close to the estimate of biparental inbreeding at the metapopulation level ($Fx = 0.033$). Although mating events at the subpopulation level may not reflect matings between siblings, they do reflect matings between individuals that share recent common ancestry, certainly more so than mating across higher levels of population subdivision. This analysis demonstrates that insights on the mating system may be gained by reporting the inbreeding coefficient on a graph of $F(d)$ values as in Figure 5. Even in the absence of knowledge about the pollen dispersal curve, we conclude that biparental inbreeding could not alone explain the inbreeding level at the individual level in *C. fasciculata*, as F_I is larger than $F(d)$ even at short distances.

Differentiation among higher levels, for example, among populations, was not found to contribute substantially to the level of inbreeding. The observation that most of the inbreeding appears to result from selfing suggests that if there is selection for inbreeding avoidance, then most of the response will be directed to the minimization of selfing and not the avoidance of biparental inbreeding, that is, selection for traits which increase the likelihood for longer distance crosses. The importance of the subpopulation level (corresponding roughly to a genetic neighborhood) to differentiation in terms of the accumulation of deleterious alleles was also observed in several crossing studies. Progeny fitness increased with increasing interparent distance up through the distance of several neighborhoods and then plateaued through distances of 1–2 km (Fenster 1991b; Sork and Schemske 1992). Furthermore, crosses among populations separated on the scale of 100 m–3000 km also almost uniformly resulted in the expression of heterosis in the F_1 offspring compared to progeny of within-subpopulation crosses, whereas the F_3 of the same interpopulation crosses exhibited hybrid breakdown (Fenster and Galloway 2000a, b, c). Overall, the results from the crossing studies are highly consistent with the pattern of limited gene flow documented using both direct and indirect methods.

Wright's Concept of Neighborhood Size

The usefulness of Wright's notion of neighborhood size and its common interpretation as a panmictic unit has been questioned (e.g., Rousset 1997, 2001a, b). Here we define the neighborhood size (*Nb)* as $4\pi D\sigma^2$, (neighborhood area defined as $4\pi\sigma^2$) and restricted this denomination to Wright's context of a continuous spatial distribution of individuals. However, *Nb* does not provide a complete characterization of gene dispersal and does not fully account for the spatial genetic structure (Slatkin 1985; Rousset 1997, 2001a, b). In particular, when rare long-distance dispersal events occur (distribution extremely leptokurtic with large σ^2 but mostly limited dispersal), then *Nb* is a poor descriptor of gene flow. However, if rare long dispersal events are ignored in the σ^2 term (e.g., the distribution is truncated to contain 99% of dispersal events), *Nb* should be a good predictor of the genetic structure, at least within some distance range. At distances inferior to σ, spatial genetic structure depends on the full dispersal distribution, not just σ^2, and at large distances, mutation as well as rare long dispersal events may become important. Thus, the importance of *Nb* should not be overestimated, but if the focus is on the bulk of pollen and seed dispersal, we suggest that *Nb* remains a useful way to synthesize the balance between drift and gene flow at a local

scale. The indirect regression approach provides an *Nb* estimate that corresponds to a truncated dispersal distribution, ignoring rare long-distance dispersal events (Rousset 2001a).

Treating *Nb* as the size of a panmictic breeding unit should be avoided as isolation by distance occurs along a continuum and selfing (beyond random mating) is incompatible with the notion of panmixy. An alternative interpretation, retaining the intuitive idea behind the ''panmictic breeding unit'' metaphor, is to treat the neighborhood area as a circular area within which biparental inbreeding remains insignificant or, equivalently, where the expected heterozygosity according to the local allele frequencies equals the observed heterozygosity of outbred individuals (i.e., not resulting from selfing). Our observation of very low level of biparental inbreeding within subpopulations of *C. fasciculata*, corresponding roughly to the size of a neighborhood, supports such interpretation. However, it is unclear whether this alternative definition of the neighborhood area matches the $4\pi\sigma^2$ definition in general. Moreover, such interpretation doesn't apply to asexual or pure selfing organisms for which biparental inbreeding is a meaningless notion.

Acknowledgments

We thank N. Barton and B. Charlesworth for organizing an ESF workshop which led to our collaboration. We are also grateful to D. Erickson, C. Murren, F. Rousset, M. Schierup, J. Tufto, two anonymous reviewers, and the Associate Editor, S. Tonsor, who provided valuable comments. D. Charlesworth, A. Friedman, E. Garber, C. Goodnight, R. Lande, T. Miller, J. Teeri, M. Wade, C. Walters, and especially C. Fenster's Ph. D. advisor, D. Schemske, provided helpful guidance during early phases of this project. This work was supported in part by funds from a National Institutes of Health Genetics Training Grant, The University of Chicago Hinds Fund, Sigma Xi, National Science Foundation BSR-8501229 and DEB-9815780 to CBF. It was also supported by grant no. 2.4512.97 from the Belgian National Fund for Scientific Research (F.N.R.S.) where OJH is a research assistant.

Literature Cited

Austerlitz, F., B. Jung-Muller, B. Godelle, and P.-H. Gouyon. 1997. Evolution of coalescence times, genetic diversity and structure during colonization. Theor. Popul. Biol. 51:148–164.

Bateman, A. J. 1947. Contamination of seed crops. I. Insect pollination. J. Genet. 48:257–275.

Beattie, A. J. and D. C. Culver. 1979. Neighborhood size in *Viola*. Evolution 33:1226–1229.

Bos, N., H. Harmens, and K. Vrieling. 1986. Gene flow in *Plantago* I. Gene flow and neighborhood size in *P. lanceolata*. Heredity 56:43–54.

Bossart, J. L., and D. P. Prowell. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. Trends Ecol. and Evol. 13:202–206.

Campbell, D. R., and J. L. Dooley. 1992. The spatial scale of genetic differentiation in a hummingbird-pollinated plant: comparison with models of isolation by distance. Am. Nat. 139:735–748.

Crawford, T. J. 1984. What is a population? Pp. 135–173 *in* B. Shorrocks, ed. Evolutionary ecology. Blackwell, Oxford, U.K.

Ellstrand, N. C., and D. L. Marshall. 1985. Interpopulation gene flow by pollen in wild radish, *Raphanus sativus*. Am. Nat. 126: 606–616.

Endler, J. A. 1977. Geographic variation, speciation and clines. Princeton Univ. Press, Princeton, NJ.

Epperson, B. K. 1990. Spatial patterns of genetic variation within plant populations. Pp. 229–253 *in* A. H. D. Brown, M. T. Clegg, A. L. Kahler, and B. S. Weir, eds. Population genetics and germplasm resources in crop improvement. Sinauer Associates, Sunderland, MA.

———. 1995. Spatial distributions of genotypes under isolation by distance. Genetics 140:1431–1440.

Epperson, B. K., and M. T. Clegg. 1986. Spatial autocorrelation analysis of flower color polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). Am. Nat. 128: 840–858.

Epperson, B. K., and Li, T. K. 1997. Gene dispersal and spatial genetic structure. Evolution 51:672–681.

Fenster, C. B. 1988. Gene flow and population differentiation in *Chamaecrista fasciculat* (Leguminosae). Ph. D. diss. The University of Chicago, Chicago IL.

———. 1991a. Gene flow in *Chamaecrista fasciculata* (Leguminosae). I. Gene dispersal. Evolution 45:398–409.

———. 1991b. Gene flow in *Chamaecrista fasciculata* (Leguminosae). II. Gene establishment. Evolution 45:410–422.

———. 1991c. Effect of seed parent and pollen donor on the allocation of resources to developing seeds and fruit in *Chamaecrista fasciculata*. Am. J. Bot. 78:13–23.

———. 1995. Mirror image flowers and mating system in *Chamaecrista fasciculata* (Leguminosae). Am. J. Bot. 82:46–50.

Fenster, C. B., and M. R. Dudash. 1994. Genetic considerations in plant population conservation and restoration. Pp. 34–62 *in* M. L. Bowles and C. Whelan, eds. Restoration of endangered species: conceptual issues, planning and implementation. Cambridge Univ. Press, Cambridge, U.K.

Fenster, C. B., and L. F. Galloway. 2000a. Population differentiation in an annual legume: genetic architecture. Evolution 54: 1157–1172.

———. 2000b. Inbreeding and outbreeding depression in natural populations of *Chamaecrista fasciculata* (Fabaceae): consequences for conservation biology. Conserv. Biol. 14:1406–1412.

———. 2000c. The contribution of epistasis to the evolution of natural populations: a case study of an annual plant. Pp. 232–244 *in* J. B. Wolf, E. D. Brodie III, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford Univ. Press, New York.

Fenster, C. B., and V. Sork. 1988. Effect of crossing distance and male parent on in vivo pollen tube growth in *Chamaecrista* (=*Cassia*) *fasciculata* (Leguminoseae). Am. J. Bot. 75: 1898–1903.

Fenster, C. B., L. F. Galloway, and L. Chao. 1997. Epistasis and its consequences for the evolution of natural populations. Trends Ecol. Evol. 12:282–286.

Fischer, M., R. Husi, D. Prati, M. Peintinger, M. van Kleunen, and B. Schmid. 2000. RAPD variation among and within small and large populations of the rare clonal plant *Ranunculus reptans* (Ranunculaceae). Am. J. Bot. 87:1128–1137.

Godt, M. J. W., and J. L. Hamrick. 1993. Patterns and levels of pollen-mediated gene flow in *Lathyrus latifolius*. Evolution 47: 98–110.

Hamrick, J. L., M. J. W. Godt, and S. L. Sherman-Broyles. 1995. Gene flow among plant populations: evidence from genetic markers. Pp. 215–232 *in* P. C. Hoch and A. G. Stephenson, eds. Experimental and molecular approaches to plant biosystematics. MO. Bot. Gard., St Louis, MO.

Hardy, O. J., and X. Vekemans. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83: 145–154.

Hartl, D. L., and A. G. Clark. 1989. Principles of population genetics, second edition. Sinauer Associates, Sunderland, MA.

Hutchison, D. F., and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53:1898–1914.

Jones, J. S., B. H. Leith, and P. Rawlings. 1977. Polymorphism in *Cepaea*. Annu. Rev. Ecol. Syst. 8:109–143.

Kalisz, S, J. D. Nason, F. M. Hanzawa, and S. J. Tonsor. 2001. Spatial population genetic structure in *Trillium grandiflorum*: the

roles of dispersal, mating, history, and selection. Evolution 55: 1560–1568.
Kelley, W. A., and R. P. Adams. 1977. Preparation of extracts from juniper leaves for electrophoresis. Phytochemistry 16:513–516.
Kerster, H. W., and D. A. Levin. 1968. Neighborhood size in *Lithospermum caroliniense*. Genetics 60:577–587.
Kimura, M., and G. H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576.
Knight, S. E., and D. M. Waller. 1987. Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. Evolution 41:969–978.
Lee, T. D., and F. A. Bazzaz. 1982. Regulation of fruit and seed production in an annual legume, *Cassia fasciculata*. Ecology 63: 1361–1373.
Levin, D. A. 1981. Dispersal versus gene flow in plants. Ann. MO. Bot. Gard. 68:233–253.
Levin, D. A., and H. W. Kerster 1969a. Density dependent gene dispersal in *Liatris*. Am. Nat. 103:61–74.
———. 1969b. The dependence of bee mediated pollen dispersal on plant density. Evolution 23:560–571.
Loiselle, B. A., V. L. Sork, J. Nason and C. Graham. 1995. Spatial genetic structure of a tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). Am J. Bot. 82:1420–1425.
Maruyama, T. 1977. Stochastic problems in population genetics. Springer-Verlag, Berlin.
McCauley, D. E. 1993. Evolution in metapopulations with frequent local extinction and recolonization. Oxf. Surv. Evol. Biol. 9: 109–134.
Meagher, T. R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum* I. Identification of most likely male parents. Am. Nat. 128:199–215.
Raymond, M., and F. Rousset. 1995. GENEPOP (vers. 1.2): population genetics software for exact tests and eucumenism. J. Hered. 86:248–249.
Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 434:223–225.
Ritland, K. 1989. Genetic differentiation, diversity, and inbreeding in the mountain monkeyflower (*Mimulus-caespitousus*) of the Washington Cascades. Can. J. Bot. 67:2017–2034.
———. 1996. A marker-based method for inferences about quantitative inheritance in natural populations. Evolution 50: 1062–1073.
Ritland, K., and Jain, S. 1981. A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. Heredity 47:35–52.
Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145: 1219–1228.
———. 2000. Genetic differentiation between individuals. J. Evol. Biol. 13:58–62.
———. 2001a. Inferences from spatial population genetics. Pp. 239–269 *in* D.J. Balding, M. Bishop, and C. Cannings, eds. Handbook of statistical genetics. John Wiley and Sons, Chichester, U.K.
———. 2001b. Genetic approaches to the estimation of dispersal rates. Pp. 18–28 *in* J. Clobert, E. Danchin, A.A. Dhondt, and J.D. Nichols, eds. Dispersal. Oxford Univ. Press, New York.
Schaal, B. A. 1980. Measurement of gene flow in *Lupinus texenis*. Nature 284:450–451.
Schmitt, J. 1983. Density-dependent pollinator foraging, flowering phenology, and temporal pollen dispersal patterns in *Linanthus bicolor*. Evolution 37:1247–1257.
Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin. Vers. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
Selander, R. K. 1975. Stochastic factors in the genetic structure of populations. Pp. 284–331 *in* G. F. Estabrook, ed. Proceedings of the eight international conference on numerical taxonomy, Freeman, San Francisco.
Slatkin, M. 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16:393–430.
———. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.
Sokal, R. R., and F. J. Rohlf. 1995. Biometry. 3rd ed. Freeman and Company, New York.
Sokal, R. R., and D. E. Wartenberg. 1983. A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics 105:219–237.
Soltis, D. E., C. H. Haufler, D. C. Darrow, and G. J. Gastony. 1983. Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers, and staining schedules. Am. Fern J. 73:9–27.
Soltis, P. S., and D. E. Soltis. 1987. Population structure and estimates of gene flow in the homosporous fern *Polystichum munitum*. Evolution 41:620–629.
Sork, V. L., and D. W. Schemske. 1992. Fitness consequences of mixed-donor pollen loads in the annual legume *Chamaecrista fasciculata*. Am. J. Bot. 79:508–515.
Tufto, J., S. Engen, and K. Hindar. 1996. Inferring patterns of migration from gene frequencies under equilibrium conditions. Genetics 144:1909–1919.
Wade, M. J. 1992. Sewall Wright: gene interaction and the shifting balance theory. Oxf. Surv. Evol. Biol. 8:35–62.
Waser, N. M. 1987. Spatial heterogeneity in a population the montane perrenial plant *Delphinium nelsonii*. Heredity 58:249–256.
———. 1993. Population structure, optimal outbreeding, and assortative mating in angiosperms. Pp. 173–199 *in* N. W. Thornhill, ed. The natural history of inbreeding and outbreeding. The University of Chicago Press, Chicago, IL.
Waser, N. M., and L. F. Elliott. 1991. Dispersal and genetic structure in kangaroo rats. Evolution 45:935–943.
Weir, B. S. and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.
Whitlock, M. C., and D. E. McCauley. 1990. Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution 44:1717–1724.
———. 1999. Indirect measures of gene flow and migration: F_{ST} not equal 1/(4Nm+1). Heredity 82:117–125.
Wolf, A. T., R. W. Howe, and J. L. Hamrick. 2000. Genetic diversity and population structure of the serpentine endemic *Calystegia collina* (Convolvulaceae) in Northern California. Am. J. Bot. 87: 1138–1146.
Wright, S. 1943. Isolation by distance. Genetics 28:114–138.
———. 1951. The genetical structure of populations. Ann. Eugen. Soc. 15:323–354.
———. 1977. Evolution and the genetics of populations. Vol. 4. Variability within and among populations. The University of Chicago Press, Chicago, IL.

Corresponding Editor: S. Tonsor

Appendix

To obtain an estimator of the selfing rate that accounts for biparental inbreeding, we need an expression for the expected inbreeding coefficient, f'_I, as a function of the selfing rate, s, and the biparental inbreeding coefficient, fx (i.e., the inbreeding coefficient of truly outcrossed individuals). The inbreeding coefficient of a progeny (f'_I) is equal to the kinship coefficient between its parents. Hence, $f'_I = fx$ for outcrossing events, and $f'_I = (1 + f_I)/2$ for selfing events (Hartl and Clark 1989), where f_I is the inbreeding coefficient of the parental generation. Thus, on average, $f'_I = (1 - s) \times fx + s(1 + f_I)/2$. At equilibrium, $f_I = f'_I$, so that $f_I = (2(1 - s)fx + s)/(2 - s)$. Replacing parameters ($f_I$, fx) by their estimates (F_I, Fx), we obtain an estimator of the selfing rate: $\hat{s} = 2(F_I - Fx)/(1 + F_I - 2Fx)$. The latter equation reduces to Wright's formula when $Fx = 0$.