

Estimates of gene flow in *Eichhornia paniculata* (Pontederiaceae): effects of range substructure

BRIAN C. HUSBAND* & SPENCER C. H. BARRETT

Department of Botany, University of Toronto, Toronto, Ontario, Canada, M5S 3B2

Gene flow in the annual aquatic plant, *Eichhornia paniculata*, was inferred from estimates of genetic differentiation at 24 isozyme loci among 44 populations from north-east Brazil. Population differentiation, estimated as the correlation among genes of different individuals (ϕ), was 0.45 (range among loci, 0.10–0.69). Based on Wright's island model, this heterogeneity would result from gene flow equivalent to 0.31 immigrants (Nm) entering each population per generation. The distribution of *E. paniculata* in north-eastern Brazil is geographically and genetically subdivided, and therefore, the assumption that migrants are a random draw from all populations is likely to be violated for this range-wide estimate of gene flow. We investigated the importance of range subdivision on indirect estimates of gene flow through computer simulation and through a hierarchical analysis of F_{ST} and Nm in populations of *E. paniculata* from northeastern Brazil.

Simulations indicated that estimates of Nm in the presence of range substructure consistently underestimated the actual values of gene flow. The degree to which Nm was underestimated increased in proportion to the magnitude of genetic differentiation among range subdivisions. In *E. paniculata*, northern and southern regions of the geographical range and local clusters within regions were genetically differentiated ($\phi = 0.10$, $\phi = 0.14$, respectively) and there was a strong negative relationship between Nm estimated for pairs of populations and the geographical distance between them. Average estimates of population differentiation decreased and gene flow increased with successive reductions in spatial scale, from the complete range sampled ($Nm = 0.31$), to regions ($Nm = 0.44$), to local areas within regions ($Nm = 0.64$), and to neighbouring population pairs within local areas ($Nm = 0.58$). Similarity in estimates for neighbouring pairs and local areas suggests that, below the spatial scale of local area, gene flow estimates are not influenced by range substructure and can be considered to occur at random. Our results suggest that range substructure can have a substantial influence on gene flow estimates, and that ecologically relevant rates are likely to be higher than those indicated by range-wide analyses in organisms with geographical subdivision.

Keywords: allozymes, *Eichhornia paniculata*, F_{ST} , gene flow, range substructure, simulation.

Introduction

Gene flow, the movement of genes among populations, is of interest to evolutionary biologists because of its influence on genetic differentiation. Because it is difficult to observe directly, gene flow is often inferred indirectly from the distribution of genetic variation within and among populations (Slatkin,

1985a; 1987). This is made possible by the availability of neutral genetic markers and appropriate genetic models with which to interpret their distribution (Wright, 1951; Barton *et al.*, 1983; Slatkin, 1985b), the most common of these being Wright's island model. The island model comprises an infinite number of discrete populations containing N individuals, a proportion, m , of which are migrants each generation. Wright (1951) showed that at drift-migration equilibrium, the average number of migrants per population per generation (Nm) is

*Correspondence: Department of Botany, University of Guelph, Guelph, Ontario, Canada N1G 2W1.

inversely proportional to the levels of genetic differentiation among populations, measured as the standardized variance in allele frequency among populations (F_{ST}). This relationship is based on the assumption that gene flow is uniform among populations, and that migrants comprise a random sample of genetic variation from all other populations.

Estimates of Nm based on the indirect method represent the level of gene flow in an island model required to produce the level of genetic differentiation observed among populations. As an 'effective' rate of gene flow, Nm is valuable for evaluating the relative importance of gene flow and genetic drift to evolutionary differentiation (Slatkin & Barton, 1989). However, indirect estimates of Nm may differ from the current patterns of gene dispersal among populations (Slatkin, 1987); one reason is that most species do not resemble an island model of population structure. Plant populations may deviate from the island model in two fundamental ways (Barrett & Husband, 1990b). First, gene flow may not be constant among populations, but rather may vary depending on such factors as the availability of dispersal agents, phenology, outcrossing rate, and the distance between populations. Secondly, the genetic composition of migrants may not be constant because migrants are rarely a random sample of all populations (Levin, 1988). The gene pool of the migrants may be heterogeneous among populations when gene flow is strongly distance-dependent (Levin & Kerster, 1974) and the species' range is genetically subdivided, as a result of stepping stone migration patterns or rare colonization events into regions between which there is little subsequent gene flow.

While there are no rigorous statistical techniques for identifying the geographical structure and migration patterns most likely to account for observed genetic differentiation (Felsenstein, 1982), range substructure or hierarchical subdivision is probably common within a species range, particularly among colonizing species, because of spatial heterogeneity of habitats, topographical and geological barriers to gene flow and the importance of repeated colonization events (Erickson, 1945; Carter & Prince, 1988). The consequences of range substructure for indirect gene flow estimates are unclear; however, Slatkin (1985b) and Slatkin & Barton (1989) have shown that with some forms of nonrandom migration, estimates of gene flow may be biased because of the inflated genetic similarity among neighbouring populations, relative to the entire range. In these cases, the scale at which populations are sampled becomes particularly important.

Ecological and genetic studies of the annual, aquatic plant *Eichhornia paniculata* (Spreng.) Solms. (Pontederiaceae) indicate that the distribution of this species is genetically as well as geographically substructured. *Eichhornia paniculata* occurs in temporary pools and ditches in the arid caatinga of north-eastern Brazil and has disjunct populations in Cuba, Jamaica, Nicaragua and western Brazil (Barrett, 1985; Barrett *et al.*, 1989; Husband & Barrett, 1993). Within north-eastern Brazil, the range of *E. paniculata* is separated into northern and southern regions, with populations concentrated in local areas where suitable habitats are most abundant within each region (Husband, 1992; Husband & Barrett, 1993). Patterns of electrophoretic variation among populations in north-eastern Brazil also confirm that the range is genetically substructured (Husband & Barrett, 1993). Northern and southern parts of the range can be distinguished by several endemic alleles, and differentiation among local clusters of populations suggests that many of these local areas were probably colonized by few founding episodes. Because of the ephemeral nature of these habitats, colonization episodes on both a local and long-distance scale appear to be an important feature of the population biology of *E. paniculata*.

The magnitude of gene flow and the degree to which populations are isolated has important ecological and evolutionary implications for *E. paniculata* in north-eastern Brazil. The species has a tristylous breeding system, in which populations consist of up to three mating types, distinguished by their style and stamen lengths (Barrett, 1985). Previous studies have shown that populations are often small enough for genetic drift to result in the loss of mating types and promote a shift in mating system from outcrossing to predominantly selfing (Barrett & Husband, 1990a; Husband & Barrett, 1992). However, the role of dispersal in maintaining or restoring mating types in natural populations is not known. Indirect measures of gene flow may provide some insights in this regard.

In this study, we examined the effect of range subdivision on gene flow estimates among populations of *Eichhornia paniculata* in north-eastern Brazil. First, we estimate gene flow among all 44 populations of *E. paniculata* sampled from throughout its geographical range. Then, using computer simulations, we determine how accurately the island model depicts gene flow in populations with contrasting population migration structures (island, stepping stone and hierarchical). Here, we focus on how the degree of genetic substructure in a hierarchical structure might bias these estimates. Finally,

we examine the significance of range genetic sub-structure for estimates of gene flow in *E. paniculata* through an hierarchical analysis of population genetic structure and gene flow at a range of spatial scales. The implications of these results for genetic drift in *E. paniculata* are also considered.

Materials and methods

Gene flow estimate

Gene flow was estimated indirectly from the distribution of genetic variation among populations using polymorphic isozyme loci as genetic markers. Open-pollinated seed families were collected from plants in 44 populations of *E. paniculata* distributed throughout its geographical range in north-eastern Brazil (Fig. 1). Sample sizes ranged from 4–59 seed families (mean 26), depending on the size of the population, which ranged from eight to approximately 10 000 individuals. About 10 progeny from each family were grown to flowering and assayed electrophoretically at 24 enzyme loci. For more details on sampling and the allozyme frequencies for each population, see Husband & Barrett (1993, Table 1 and Appendix I.)

Gene flow (Nm), the number of migrants entering a population each generation, was estimated from the sample of 44 populations using the island model

(Wright, 1951), based on the equilibrium relationship:

$$Nm \approx \frac{(1 - F_{ST})}{4F_{ST}},$$

where Nm is the product of effective population size (N_e) and the fraction of individuals that are migrants (m), and F_{ST} is described as the standardized variance in allele frequency among populations or as the correlation between genes in different individuals. Nm is equivalent to the parameter M , defined by Slatkin (1993) for applications to DNA markers. Estimates of Nm using the island model are based on the assumption that all populations were initially colonized independently from a common ancestral population, that subsequent gene flow is random among populations and that the populations have reached a migration–drift equilibrium.

Population differentiation was estimated using the estimator ϕ of Weir & Cockerham (1984), which corresponds closely to Wright's F_{ST} , and takes sample size and allele number into account. Standard deviations of ϕ were calculated by a jack-knifing procedure with populations as the sampling unit. Whether ϕ was significantly different from zero was tested by a one-sample t -test, with $n - 1$ degrees of freedom ($n =$ number of loci).

For comparison, we also estimated Nm using Slatkin's private allele method (Slatkin, 1985b), which assumes that the frequency of an allele that occurs in only one population is proportional to the amount of gene flow among populations as described by the equation, $\ln(p(1)) = a \ln(Nm) + b$, where $p(1)$ is the mean frequency of alleles that occur in only one population and a and b are -0.505 and -2.44 , respectively.

Computer simulations

We examined the effects of three different models of population migration structure on estimates of gene flow using Monte Carlo computer simulations: the island model, which is depicted by random gene flow among all populations; the stepping stone model, in which genes flow occurs only between neighbouring populations; and the hierarchical model, in which gene flow occurs within isolated groups (subdivisions) of populations which may be genetically differentiated. Island and hierarchical models represent extremes of a probable continuum of population structure.

Simulated populations comprised 100 subpopulations or demes, organized as a 10×10 two-dimen-

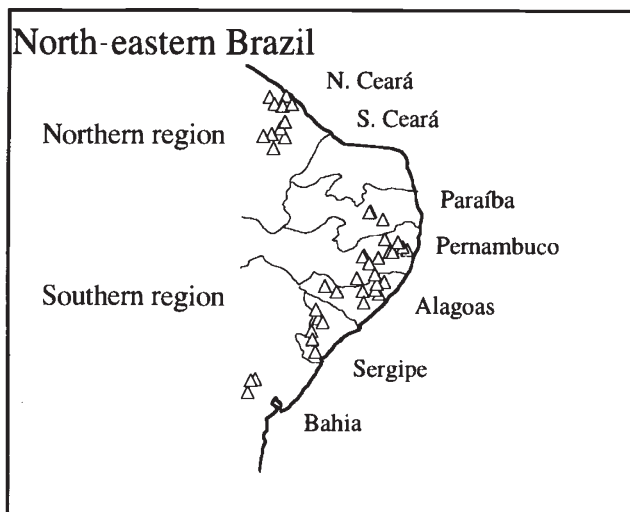


Fig. 1 Geographical distribution of 44 populations of *Eichhornia paniculata* from north-eastern Brazil sampled for electrophoretic variation. Based on their geographical location, populations were subdivided into northern and southern regions, which were in turn divided into two and five local areas, respectively.

sional array. Each subpopulation contained 20 individuals each with a genotype at a single diallelic locus assigned in Hardy–Weinberg proportions. In each subsequent generation, a proportion m of the 20 genotypes was replaced by migrants, selected from other subpopulations, and a proportion $1-m$ was replaced through random mating. The source of migrants depended on which population structure was being simulated. In the island model, migrants were chosen at random from all other populations whereas in the stepping stone model only neighbouring populations exchanged migrants. In the hierarchical model, the array was subdivided into four quarters (subdivisions) of 25 subpopulations each; migration was random within each subdivision but did not occur between them.

After 100 generations, 25 subpopulations were randomly selected, without replacement, to calculate F_{ST} which was then used to infer Nm . Because the genotypes of all individuals are known and sample sizes are equal in all populations, F_{ST} was calculated as:

$$F_{ST} = \frac{s^2}{p(1-p)},$$

where s^2 is the variance in allele frequency among subpopulations and p is the mean allele frequency across all subpopulations (Wright, 1951). Simulations were allowed to run for 100 generations, which exceeds the estimated time required to establish a drift–migration equilibrium (Crow & Aoki, 1984). Values of Nm estimated from the sample were then compared to the actual or parametric values used in each simulation with a one-sample t -test. We conducted simulations for three parameter values of $Nm = 0.10$, 0.80 , and 1.5 ; each was replicated 10 times (Table 1).

A second set of simulations was conducted to explore further the effect of range subdivision in a hierarchical structure on indirect estimates of Nm . As before, a two-dimensional array of 100 subpopulations was divided into four subdivisions between which gene flow was assumed to be negligible. Populations within each subdivision of the array were randomly assigned genotypes. The mean allele frequency in each subdivision was altered for different simulations such that subdivisions were differentiated to different degrees ($F_{ST} = 0.0, 0.04, 0.16, 0.36, 0.64$, or 1.0). The mean allele frequency, however, for all 100 populations remained constant ($p = 0.5$). After 100 generations of mating and migration, populations were sampled and Nm estimated as above. We examined the effect of range subdivision

on deviations between estimates of Nm and the parameter levels of gene flow for three parameter sets: $Nm = 0.1, 0.8$ and 1.5 .

F_{ST} and Nm as a function of spatial scale

We had two objectives in examining gene flow in *E. paniculata* at a range of spatial scales: first to determine whether estimates of F_{ST} and Nm are scale-dependent as would be expected if the range is genetically substructured, and secondly, to identify the scale that most accurately reflects actual patterns of gene flow in *E. paniculata*. This scale was identified as that scale below which further range substructuring was undetectable, a kind of genetic neighbourhood at the level of the metapopulation. Estimates of gene flow at this scale would be the product of differences among populations, without the additional influence of genetic subdivisions among groups of populations.

If the range of *E. paniculata* is genetically substructured, estimates of gene flow should vary depending on the spatial scale at which populations are sampled (Slatkin & Barton, 1989). To explore this component of gene flow in *E. paniculata* we (1) examined heterogeneity in F_{ST} and Nm among pairs of populations separated by increasing distance, (2) conducted a hierarchical analysis of population differentiation which partitions the total F_{ST} among all populations into components among regions, among areas within regions, among populations within areas and among neighbouring pairs of populations within areas, and (3) examined the effect of scale-dependent variation in F_{ST} on estimates of gene flow among populations.

All population pairs separated by 1–31, 32–94, 95–188, 189–375, 376–625, and 626–875 km were identified. These distance classes were chosen so that the number of population pairs within each class was as even as possible. Average Nm among pairs was calculated within each distance class; however, to control for nonindependence, a population was allowed to appear only once within each distance class.

To determine the effect of spatial scale on gene frequencies, we conducted a hierarchical analysis of population differentiation (F_{ST}). For this purpose, the geographical range of *E. paniculata* sampled in this study was divided into two 'regions', representing the northern and southern portions of the range in north-eastern Brazil (Fig. 1). Within regions, 'areas' were designated representing smaller concentrations of populations that were separated geographically by 50–100 km. Areas closely correspond to the different states in north-eastern Brazil. There

were two areas within the northern region (N. Ceará and S. Ceará) and five areas within the southern region (Bahia, Alagoas, Paraíba, Pernambuco and Sergipe). We estimated mean F_{ST} as ϕ among all populations, among regions within the entire range, among areas within regions, among populations within areas and among neighbouring pairs within areas.

Gene flow (Nm) among populations was also calculated for all populations within geographical boundaries of different size, including: whole range, region, area, and neighbouring pair. Mean ϕ at the region level was calculated as the average of all individual regions, weighted by the number of populations within each region. A similar weighting approach was used for all lower spatial scales. Gene flow (Nm) was then estimated from the ϕ values at each spatial scale using the island model.

Finally, to illustrate the impact of range substructure, we calculated F_{ST} and Nm for each 'area' using two different measures of F_{ST} : (1) the variance in allele frequencies among populations within an area, relative to the mean allele frequencies within a given area; (2) the variance in allele frequencies among populations with an area relative to the mean allele frequencies for the range (all 44 populations). The two estimates of Nm , which differ with respect to the populations used to form the migrant pool are compared using Spearman's rank correlation coefficient.

Results

Gene flow

Progeny from 1100 maternal plants from the 44 populations of *Eichhornia paniculata* were assayed

for variation at 24 isozyme loci. Of the 24 loci, 20 were polymorphic (common allele <0.95) in at least one population. The average level of differentiation among populations (ϕ) was 0.45, and ranged from 0.10 to 0.69 among loci (Table 2). These estimates were based on an analysis of maternal families and therefore not all genotypes were independent. However, ϕ was also calculated for a sample consisting of one progeny, drawn at random, per maternal plant and found to be 0.47. Both measures are significantly different from zero based on a chi-square test ($P < 0.05$). Based on the island model, such a level of differentiation ($\phi = 0.45$) would result from an average level of gene flow (Nm) of 0.31 migrants per generation.

In the sample of 44 populations, only three private alleles were detected (*PGI-2a*; *AAT-3a*; *PGD-1a*). Their respective frequencies were 0.658, 0.065 and 0.061. The average Nm based on the private allele method was 0.11, about one third of the estimate based on the island model.

Computer simulations

With random gene flow among populations (island model), estimates of Nm based on 25 randomly chosen populations were not significantly different from the parameter values of Nm (Table 1). This result was consistent for all levels of Nm used in the simulations. With a stepping stone pattern of migration, Nm estimates also did not differ significantly from parameter values although the disparities were larger than for the island model. Only one of the Nm estimates in simulations of hierarchical population structure (no genetic subdivision added) deviated significantly from the parameter value of gene

Table 1 Estimates of gene flow (Nm) among 25 randomly chosen populations from simulated data

Population structure	Expected Nm		
	pv (Nm) [†] = 0.1	pv (Nm) = 0.8	pv (Nm) = 1.5
Island	0.14 (0.016)	0.82 (0.079)	1.32 (0.113)
Stepping stone	0.16 (0.017)	0.58 (0.040)	1.00 (0.109)
Hierarchical	0.35 (0.016)*	0.84 (0.054)	1.14 (0.114)

[†]pv (Nm) are the parameter values of Nm .

Each simulation consisted of 100 populations in a 10×10 array and 20 individuals per population (N) with either an island, stepping stone or hierarchical population structure.

Each value is the mean and SE of 10 simulations at a diallelic locus ($p = 0.5$).

Three different parameter values for Nm were used: 0.1, 0.8 and 1.5.

Based on an experimentwise error rate of $\alpha' = 0.05$, only one estimate differed significantly from the parameter value for that simulation (indicated by *).

Table 2 Heterogeneity in allele frequencies at 20 polymorphic isozyme loci among 44 populations of *Eichhornia paniculata* in north-eastern Brazil

Locus	Mean ϕ	SD	Locus	Mean ϕ	SD
<i>AAT-1</i>	0.3616	0.1890	<i>ACP-1</i>	0.4832	0.0693
<i>AAT-2</i>	0.2096	0.0372	<i>ACP-2</i>	0.3716	0.0752
<i>AAT-3</i>	0.4625	0.0647	<i>MDH-1</i>	0.1200	0.0321
<i>TPI-2</i>	0.5241	0.0897	<i>MDH-2</i>	0.2603	0.0753
<i>TPI-3</i>	0.1008	0.0076	<i>MDH-3</i>	0.2748	0.0872
<i>PER-1</i>	0.4186	0.0800	<i>DIA-1</i>	0.2260	0.0412
<i>GDH-1</i>	0.5224	0.0758	<i>IDH-1</i>	0.5058	0.0903
<i>PGI-2</i>	0.6479	0.0386	<i>IDH-2</i>	0.3384	0.1150
<i>PGD-1</i>	0.6884	0.1332	<i>ACO-1</i>	0.2172	0.0860
<i>PGD-2</i>	0.3039	0.0803	<i>ACO-2</i>	0.2897	0.0380

Mean over all loci = 0.4493 (SD = 0.0753).

Population differentiation was reported using the statistic ϕ ($\approx F_{ST}$) which is the extent of inbreeding in a population arising from correlation among alleles caused by substructuring.

Values of ϕ , estimated using Weir & Cockerham's method (1984), range from 0, if there is no differentiation among subpopulations, to 1, when subpopulations are fixed for different alleles.

Standard deviations for each estimate were generated by jackknifing over populations (Weir & Cockerham, 1984).

flow and that was at the lowest simulated level of gene flow.

The degree to which a hierarchical population structure affected estimates of Nm depended on the degree of genetic differentiation among the four subdivisions of the range. When each of the subdivisions, each containing 25 populations, had the same average allele frequency, the hierarchical model behaved as a combination of four island models and estimates of Nm were similar to the parameter values of Nm (Table 1; Fig. 2). Estimates of Nm were in fact slightly larger than the parameter values for the lowest levels of gene flow simulated ($Nm = 0.1$). As the variance among subdivisions increased, estimates of Nm showed increasing deviation from the expected values. With an F_{ST} of 0.16 among subdivisions, estimates of Nm were underestimated by an average of 44 per cent and as much as 96 per cent with an F_{ST} of 0.36. In virtually all simulations, Nm estimates were lower than the expected value.

Spatial scale and gene flow in Eichhornia paniculata

Estimates of Nm among pairs of populations of *E. paniculata* ranged from 0.17 to 10.16. The mean Nm was significantly correlated with geographical dis-

tance (Spearman rank correlation $r_s = -0.94$, $P < 0.05$, $n = 6$), when calculated for six distance classes (Fig. 3). Mean Nm for populations separated by less than 30 km was 1.02 but decreased rapidly with increasing separation to a minimum of 0.45.

The ϕ value of 0.45, which was based on all populations, can be partitioned into genetic structure occurring at a range of spatial scales. In our hierarchical analysis of ϕ , there were marked differences in gene frequency among regions ($\phi = 0.10$), areas within regions ($\phi = 0.14$), populations within areas ($\phi = 0.26$) and neighbouring populations within areas ($\phi = 0.30$). The degree of differentiation between regions, and among areas within regions is within the range of values of substructure that cause substantial biases in estimates of gene flow in simulations (see above).

Mean ϕ and Nm among populations varied as a function of the spatial scale at which they were estimated (Table 3). When all 44 populations were included in a single estimate, ϕ was 0.45 and Nm was equal to 0.31, as discussed above. In successive reductions of the scale at which estimates were made, ϕ decreased and Nm increased (Table 3) up to the level of the 'area', at which $\phi = 0.28$ and $Nm = 0.64$. There were only small differences between levels of differentiation and gene flow estimated at the area and neighbour level, and neigh-

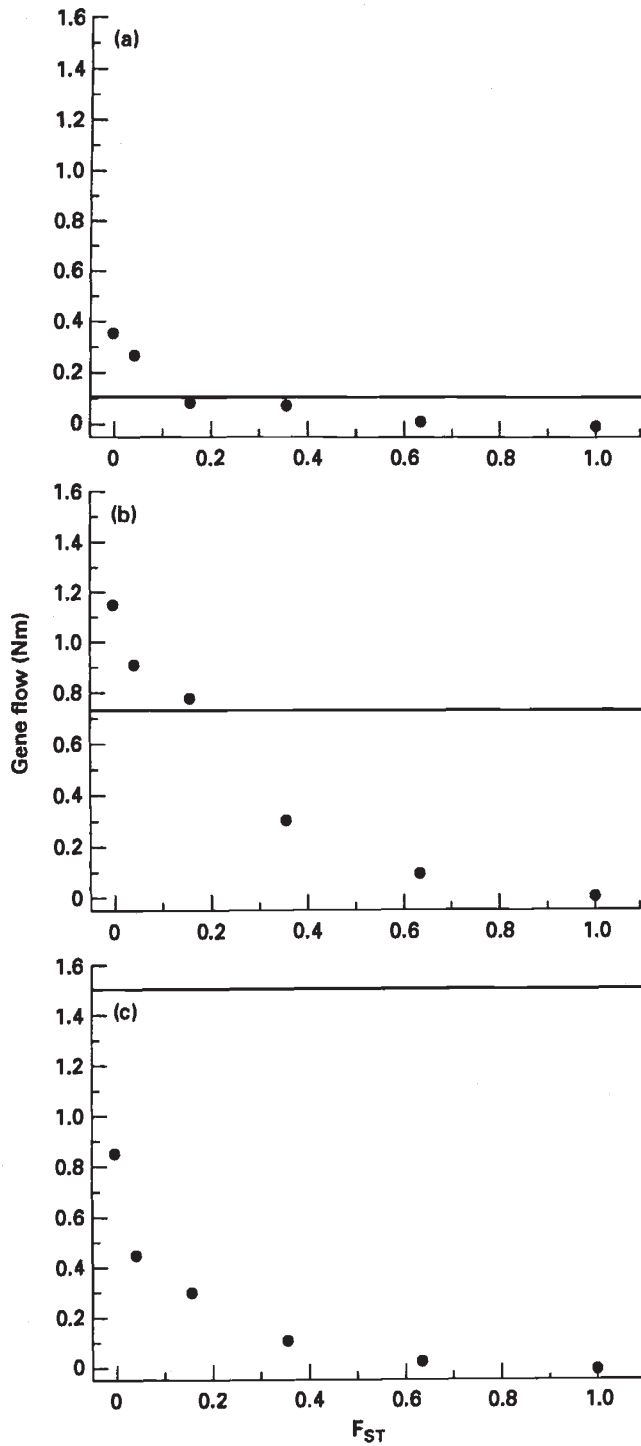


Fig. 2 The effect of range substructure (F_{ST} among four subdivisions) on estimates of gene flow. Estimates of Nm were made by sampling 25 populations from a 10×10 array of simulated populations which was divided into four genetically differentiated subdivisions. The solid line represents the expected level of gene flow among simulated populations. (a) $Nm = 0.10$, (b) $Nm = 0.80$, (c) $Nm = 1.5$.

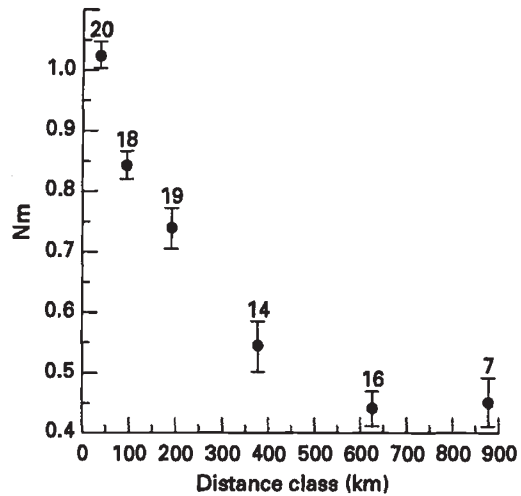


Fig. 3 Estimates of Nm between pairs of sample populations of *Eichhornia paniculata* as a function of distance. Mean Nm was calculated for all independent pairs within each distance class (n indicated above each mean).

Table 3 Hierarchical analysis of differentiation (ϕ) and gene flow (Nm) in populations of *Eichhornia paniculata* from north-eastern Brazil

Spatial scale	ϕ	Nm
Whole range	0.45	0.305
Region	0.36	0.444
Area	0.28	0.643
Neighbouring populations	0.30	0.583

Genetic differentiation among populations (ϕ) was calculated at the levels of region, area and neighbouring populations as the average, which was weighted by the number of populations for the analysis of regions and areas.

bours actually showed slightly higher levels of differentiation.

Estimates of F_{ST} and Nm differed substantially among the seven local areas sampled in north-eastern Brazil (Table 4). However, the estimates differed depending on whether, in the calculation of F_{ST} , variance among populations was expressed relative to allele frequencies from the entire range or just from the local area. When range-wide allele frequencies were used, mean Nm among the 'areas' was 0.335 and ranged from 0.092 in N. Ceará to 0.610 in Alagoas state. The estimate of Nm for each area was correlated with the difference in allele frequencies between that local area and the range ($r_s = -0.79$, $P = 0.057$, Fig. 4). When F_{ST} was expressed relative to the local allele frequencies,

Table 4 Estimates of genetic differentiation (F_{ST}) and gene flow (Nm) for populations of *Eichhornia paniculata* from seven spatially separated areas in north-eastern Brazil

Area	n	F_{ST} -constant	Nm	F_{ST} -local	Nm
Alagoas	8	0.29	0.610	0.25	0.750
Bahia	3	0.43	0.330	0.15	1.417
N. Ceará	6	0.73	0.092	0.20	1.000
Paraíba	4	0.35	0.464	0.26	0.712
Pernambuco	9	0.33	0.508	0.33	0.508
S. Ceará	6	0.60	0.167	0.17	1.221
Sergipe	8	0.59	0.174	0.47	0.282
Mean		0.47	0.335	0.26	0.841

F_{ST} was calculated two ways: the variance in allele frequencies among populations within each area relative to (1) the maximum variance, based on mean allele frequencies of all 44 populations (F_{ST} -constant), and (2) the maximum variance, based on the mean allele frequencies of populations within the particular scale (F_{ST} -local).

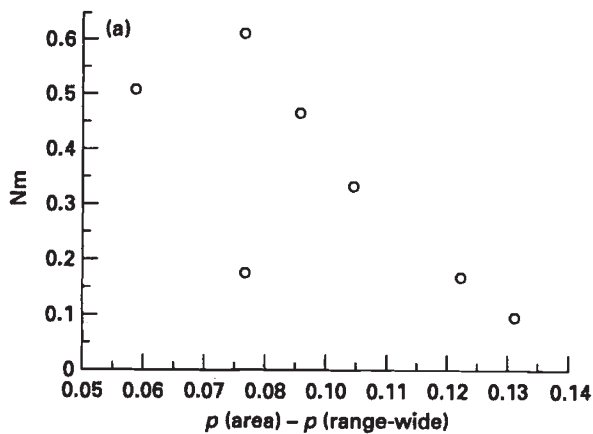


Fig. 4 Gene flow (Nm) estimates in *Eichhornia paniculata* from seven areas in north-eastern Brazil as a function of the average deviation in allele frequencies between that area and the range-wide average. The negative correlation illustrates the dependence of Nm estimates on range substructuring, when average allele frequencies are based on populations from the complete range.

values of Nm were much higher. The mean Nm was 0.841, and ranged from 0.28 to 1.42. Areas such as N. Ceará, which had a low level of gene flow in the previous analysis, now had among the highest levels, while Sergipe state had the lowest. There was no correlation between the two estimates of Nm for each region ($r_s = -0.32$, $P > 0.25$).

Discussion

Determining the magnitude of gene flow among populations is a common and controversial topic among evolutionary biologists (Hamrick, 1987; Slatkin, 1987). Direct observations of dispersal as well as inferential methods using marker genes have been used to estimate gene flow but differences among estimates have yet to be completely reconciled (Ellstrand & Marshall, 1985; Hamrick, 1987; Slatkin, 1987; Campbell, 1991; Godt & Hamrick, 1993). Since the application of electrophoretic markers in population biology, indirect inferences of gene flow from genetic structure have been conducted using several different approaches, of which Wright's island model has been shown to be the most robust (Slatkin & Barton, 1989). However, it is important to understand under what circumstances discrepancies between the estimates and the actual magnitude of gene flow may arise.

The average estimated level of gene flow among populations of *Eichhornia paniculata* in north-eastern Brazil was 0.31 migrants per generation. The estimate for *E. paniculata* was substantially lower than estimates for other plants that are outcrossing and insect-pollinated (mean $Nm = 1.02$, from G_{ST} values in Hamrick & Godt, 1990) and lower than estimates for animals with mobility such as the butterfly *Euphydryas editha* ($Nm = 7.8$; Ehrlich *et al.*, 1975), the bivalve *Mytilus edulis* ($Nm = 42$; Slatkin, 1985b) and species of *Drosophila* ($Nm = 1.09$ – 2.66 ; Singh & Rhomberg, 1987; $Nm = 1.0$ – 9.9 in Slatkin, 1985b). Gene flow in *E. paniculata* was comparable

to estimates for self-fertilizing plants ($Nm = 0.24$, from G_{ST} values in Hamrick & Godt, 1990) and for organisms known for their site fidelity such as the salamander *Batrachoseps campi* ($Nm = 0.16$; Slatkin, 1985b) and cave arthropods (Caccone, 1985). This suggests that the actual level of gene flow in *E. paniculata* is quite low. In fact, the estimate of Nm for this species is substantially less than the critical value of 1, below which drift plays an important role in population differentiation (Wright, 1931, 1969).

The low estimate of gene flow for *E. paniculata* may reflect the habitat type and natural history of the species in north-eastern Brazil. Populations occur in small pools and ditches which are often separated by large tracts of arid, uninhabitable land. Since suitable habitats are ephemeral and reliant on the availability of water, populations are often isolated not only in space but also in time. This reduces opportunities for movement of pollen among populations. Moreover, *E. paniculata* is insect-pollinated, and therefore, pollen dispersal will be limited by the distance between populations (Levin & Kerster, 1974). Seed dispersal is similarly restricted as the seeds are not small enough to be air-dispersed and most fall directly into the water in which the parent plant grows. The most likely mechanism of gene flow by seed is through transport on waterfowl or livestock. As a result, recurrent gene flow between populations by gamete or seed dispersal is probably limited in frequency as well as distance.

While our estimates of Nm represent the genetically effective levels of migration among *E. paniculata* populations, they may not represent the actual levels of gene flow currently occurring among populations of *E. paniculata*. Slatkin (1987) has shown that in several species where direct observations and indirect marker-based estimates of gene flow have been made, the two estimates differ substantially. This discrepancy may arise in part from difficulties in estimating and interpreting gene flow using indirect methods. Discrepancies may be most likely when the population structure deviates significantly from that of an island model and, hence, genetic differentiation is the result of factors other than the balance between genetic drift and recurrent migration.

The ecology and habitat distribution of *E. paniculata* in north-eastern Brazil probably promotes migration that is different from an island model. Its geographical range comprises a number of regions, and local areas within, that are separated by uninhabitable territory which probably acts as a barrier to dispersal (Husband & Barrett, 1993). These regions and areas are genetically differentiated and

relatively homogeneous within (Husband & Barrett, 1993). Genetic differentiation was positively related to geographical distance (Husband & Barrett, 1993; Fig. 3). Such isolation by distance may arise with either a stepping stone (Golenberg, 1987; Slatkin, 1993) or a hierarchical migration structure. In *E. paniculata*, such patterns probably arise through founder effects during occasional long-distance dispersal events to a region and subsequent colonization from a few founding populations. Population genetic structure of *E. paniculata* in Jamaica also suggests that the distribution of the species is the product of few geographically separate colonization events followed by subsequent expansion (Husband & Barrett, 1991). Therefore despite a wide distribution, individual populations regularly exchange genes with only a small subset of all populations, most of which are genetically similar. Because estimates of Nm based on the island model assume a constant allele frequency in the migrant pool, range substructure may cause significant biases in estimates of gene flow in *E. paniculata* when natural populations are sampled from throughout the range.

Simulations and analytical studies have shown that, under certain sampling regimes, deviations from an island population structure may provide spurious estimates of the actual levels of gene flow. Slatkin (1985a) suggested that Nm values would be slightly underestimated when the true pattern of gene flow is restricted to adjacent populations, although Crow & Aoki (1984) showed that average estimates of Nm from populations with stepping stone structures were quite similar to populations in an island structure with the same levels of gene flow. Our simulations also indicated little bias in a stepping stone structure; however, estimates became increasingly low with levels of $Nm > 1$ (Table 1). Slatkin & Barton (1989) extended this type of analysis to show that in an island structure, estimates are independent of the location of samples. They found, however, that when the patterns of gene flow are between adjacent populations (stepping stone structure), sampling populations located near to one another will cause an overestimate of actual values of gene flow. This is largely because of the genetic correlation between neighbouring populations generated through nonrandom migration among populations (Kimura & Weiss, 1964).

There have been few theoretical investigations into the effects of hierarchical population structure on estimates of gene flow using the island model. Slatkin (1985b) studied such effects on estimates using the private alleles method. However, he did not examine the possibility that divisions within the

hierarchy may become genetically differentiated. Not surprisingly, he found that the hierarchical model behaved like an island model, which in fact it resembled, because all populations experienced the same migrant pool. In the present study, we examined the effect of differentiation among range subdivisions on estimates of gene flow. We found that the larger the variance in allele frequency among subdivisions, the greater the disparity between estimated and expected levels of gene flow, when populations were sampled randomly. In almost all cases, the effect was to underestimate markedly the level of gene flow. This occurs because gene flow estimates are based on genetic variances among populations, relative to allele frequencies in the entire sample. Combining populations from genetically different regions into a single sample results in population differentiation being quantified relative to an allele frequency that are representative of neither individual region. As a result, estimates of ϕ (or F_{ST}) are a product of differentiation among local populations, as well as differentiation at higher spatial scales. An analogous situation exists in the measurement of mating-system parameters in plant populations using the mixed-mating model. If one assumes the allele frequencies in the pollen are constant among maternal plants, population subdivisions can inflate the estimates of effective selfing (Ennos & Clegg, 1982; Ritland & Ganders, 1985).

The results from this study suggest that regional differentiation or substructuring of the geographical range (e.g. among regions and areas within the range of *E. paniculata*) can influence indirect estimates of gene flow. Estimates at gene flow in *E. paniculata* varied depending on the spatial scale from which populations were drawn (Table 3), from 0.31 using range-wide average allele frequencies to 0.64 when gene flow was calculated as the mean of estimates at the level of the area. Estimates at the level of the area and neighbouring populations were similar, indicating that no further differentiation arising from substructuring occurred within local areas. It is at this spatial scale that gene flow among populations could be considered to occur randomly or nearly so. The progressive increase in Nm estimates at smaller spatial scales results from the elimination of contributions to effective gene flow by regional patterns of genetic differentiation. The finding that Nm at the range level is an underestimate confirmed predictions made by the simulations of hierarchical population structure (Fig. 2).

The variation in Nm values among the seven local areas sampled within the range of *E. paniculata* provided additional evidence that hierarchical popula-

tion structure has an important influence on estimates of gene flow. Values ranged from 0.09–0.61 among areas when differences among populations were quantified relative to range-wide allele frequencies (Table 4). Interestingly, these values of Nm were related to the degree to which allele frequencies in the particular area deviated from the range-wide allele frequencies (Fig. 4). This suggests that estimates of Nm are influenced primarily by the extent of differentiation among areas. To remove the component of gene flow arising from genetic differentiation among areas, Nm was recalculated for each area, where differences among populations were estimated relative to mean allele frequencies of the local area, rather than the range. Because there was no detectable structuring below the area level, these estimates should reflect more closely the true levels of gene flow between *E. paniculata* populations. Local areas show a large degree of variation in Nm , but have no relationship with estimates based on a constant migrant pool. Without the influence of range subdivision, the mean Nm (0.84) was 2.5 times larger than the original estimate (0.31) of effective gene flow among populations. Both values are less than the critical value of 1, below which drift plays an important role in population differentiation (Wright, 1931, 1969). Furthermore, from the independent estimate of N_e (mean of 10 populations is 15.8, Husband & Barrett, 1992) and the estimate of Nm of 0.84, we calculate that m is on average approximately 0.05. In other words, 5 per cent of the individuals in each population are migrants.

A feature of *E. paniculata* that may also deflate estimates of gene flow is the wide fluctuation in population size and apparently frequent extinction and colonization observed in Brazilian populations. Wade & McCauley (1988) showed that such population dynamics can either enhance or diminish differentiation among populations, although greater differentiation occurs under a broader range of circumstances. While this process may not influence range substructure, the focus of this study, it may help to explain our generally low estimates of Nm within local areas. To understand further the role of local extinction and recolonization on gene flow estimates in *E. paniculata*, we require empirical estimates of the rates of population extinction and the magnitude of genetic diversity of population colonists and migrants (Wade & McCauley, 1988).

Our results suggest that estimates of gene flow in *E. paniculata* are strongly influenced by differentiation among different parts of the range sampled. This genetic substructure is probably established

through colonization of new regions and the subsequent spread and differentiation of populations. The impact of historical colonizations and dispersal barriers is to reduce estimates of gene flow inferred from data on population genetic structure and cause spurious estimates of variation in Nm among areas throughout the range. Estimating Nm from Wright's F_{ST} (1951) and the island model is considered a method that does not depend on population structure (Slatkin & Barton, 1989) and thus would be suitable for the common situation in which the investigator has no knowledge of the population structure. However, if the objective is to quantify, not just genetically effective levels of gene flow, but actual patterns of gene flow among natural populations, our results indicate that indirect estimates should be interpreted with caution, particularly when species exhibit hierarchical structure. Clearly, a knowledge of population structure (stepping stone, random migration, hierarchical) is important for interpreting the ecological significance of gene flow estimates in natural populations.

Acknowledgements

We thank M. Morgan, J. Conn and W. Cole for field assistance, W. Cole for help with the analyses and the Natural Sciences and Engineering Research Council of Canada for financial support.

References

- BARRETT, S. C. H. 1985. Floral trimorphism and monomorphism in continental and island populations of *Eichhornia paniculata* (Spreng.) Solms. (Pontederiaceae). *Biol. J. Linn. Soc.*, **25**, 41–60.
- BARRETT, S. C. H. AND HUSBAND, B. C. 1990a. Variation in outcrossing rate among populations of *Eichhornia paniculata*: the role of demographic and reproductive factors. *Pl. Sp. Biol.*, **5**, 41–55.
- BARRETT, S. C. H. AND HUSBAND, B. C. 1990b. The genetics of plant migration and colonization. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (eds) *Plant Population Genetics, Breeding and Genetic Resources*, pp. 254–277. Sinauer Associates, Sunderland, MA.
- BARRETT, S. C. H., MORGAN, M. T. AND HUSBAND, B. C. 1989. The dissolution of a complex genetic polymorphism: the evolution of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae). *Evolution*, **43**, 1398–1416.
- BARTON, N. H., HALLIDAY, R. B. AND HEWITT, G. M. 1983. Rare electrophoretic variants in a hybrid zone. *Heredity*, **50**, 139–146.
- CACCONE, A. 1985. Gene flow in cave arthropods: a qualitative and quantitative approach. *Evolution*, **39**, 1223–1235.
- CAMPBELL, D. R. 1991. Comparing pollen dispersal and gene flow in a natural population. *Evolution*, **45**, 1965–1968.
- CARTER, R. N. AND PRINCE, S. D. 1988. Distribution limits from a demographic viewpoint. In: Davy, A. J., Hutchings, M. J. and Watkinson, A. R. (eds) *Plant Population Ecology*, pp. 165–184. Blackwell Scientific Publications, Oxford.
- CROW, J. F. AND AOKI, K. 1984. Group selection for a polygenic behavioral trait: Estimating the degree of population subdivision. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 6073–6077.
- EHRlich, P. R., WHITE, R. R., SINGER, M. C., McKECHNIE, S. W. AND GILBERT, L. E. 1975. Checkerspot butterflies: a historical perspective. *Science*, **188**, 221–228.
- ELLSTRAND, N. C. AND MARSHALL, D. L. 1985. Interpopulation gene flow by pollen in wild radish, *Raphanus sativus*. *Am. Nat.*, **126**, 606–616.
- ENNOS, R. A. AND CLEGG, M. T. 1982. Effect of population substructuring on estimates of outcrossing rate in plant populations. *Heredity*, **48**, 283–292.
- ERICKSON, R. O. 1945. The *Clematis fremontii* var. *riehlii* population in the Ozarks. *Ann. Mo. Bot. Gard.*, **32**, 413–460.
- FELSENSTEIN, J. 1982. How can we infer geography and history from gene frequencies? *J. Theor. Biol.*, **96**, 9–20.
- GODT, M. J. W. AND HAMRICK, J. L. 1993. Patterns and levels of pollen-mediated gene flow in *Lathyrus latifolius*. *Evolution*, **47**, 98–110.
- GOLENBERG, E. M. 1987. Estimation of gene flow and genetic neighborhood size by indirect methods in a selfing annual, *Triticum dicoccoides*. *Evolution*, **41**, 1326–1334.
- HAMRICK, J. L. 1987. Gene flow and the distribution of genetic variation in plant populations. In: Urbanska, K. M. (ed.) *Differentiation Patterns in Higher Plants*, pp. 53–67. Academic Press, London.
- HAMRICK, J. L. AND GODT, M. J. 1990. Allozyme diversity in plant species. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (eds) *Plant Population Genetics, Breeding, and Genetic Resources*, pp. 43–63. Sinauer Associates, Sunderland, MA.
- HUSBAND, B. C. 1992. *Stochastic Processes and the Evolution of Self-fertilization in Eichhornia paniculata* (Spreng.) Solms. (Pontederiaceae). Ph. D. Thesis, University of Toronto.
- HUSBAND, B. C. AND BARRETT, S. C. H. 1991. Colonization history and population genetic structure of *Eichhornia paniculata* in Jamaica. *Heredity*, **66**, 287–296.
- HUSBAND, B. C. AND BARRETT, S. C. H. 1992. Effective population size and genetic drift in tristylous *Eichhornia paniculata* (Pontederiaceae). *Evolution*, **46**, 1875–1890.
- HUSBAND, B. C. AND BARRETT, S. C. H. 1993. Multiple origins of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae): Inferences from style morph and isozyme variation. *J. Evol. Biol.*, **6**, 591–608.
- KIMURA, M. AND WEISS, G. H. 1964. The stepping stone model of population structure and the decrease in genetic correlation with distance. *Genetics*, **49**, 561–576.

- LEVIN, D. A. 1988. Consequences of stochastic elements in plant migration. *Am. Nat.*, **132**, 643–651.
- LEVIN, D. A. AND KERSTER, H. W. 1974. Gene flow in seed plants. *Evol. Biol.*, **7**, 139–220.
- RITLAND, K. AND GANDERS, F. R. 1985. Variation in the mating system of *Bidens menziesii* (Asteraceae) in relation to population substructure. *Heredity*, **55**, 235–244.
- SINGH, R. S. AND RHOMBERG, L. R. 1987. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. *Genetics*, **115**, 313–322.
- SLATKIN, M. 1985a. Gene flow in natural populations. *Ann. Rev. Ecol. Syst.*, **16**, 393–430.
- SLATKIN, M. 1985b. Rare alleles as an indicator of gene flow. *Evolution*, **39**, 53–65.
- SLATKIN, M. 1987. Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.
- SLATKIN, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- SLATKIN, M. AND BARTON, N. H. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**, 1349–1368.
- WADE, M. J. AND McCAULEY, D. E. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution*, **42**, 995–1005.
- WEIR, B. S. AND COCKERHAM, C. C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- WRIGHT, S. 1951. The genetical structure of populations. *Ann. Eugen.*, **15**, 323–354.
- WRIGHT, S. 1969. *Evolution and the Genetics of Populations*, vol. 2, *The Theory of Gene Frequencies*. University of Chicago Press, Chicago, IL.