

EFFECT OF POPULATION SUBSTRUCTURING ON ESTIMATES OF OUTCROSSING RATE IN PLANT POPULATIONS

R. A. ENNOS* and M. T. CLEGG

Department of Botany and Department of Molecular and Population Genetics,
University of Georgia, Athens, Georgia 30602, U.S.A.

Received 4.viii.81

SUMMARY

In plants possessing a mixed selfing and random mating system, outcrossing rate t can be estimated from progeny genotype arrays. Computer simulation studies show that if there is heterogeneity of gene frequency between subpopulations, these estimates of t are severely biased downwards, the degree of bias increasing as the true value of t increases. This effect has been demonstrated in the field using experimental populations of *Ipomoea purpurea*. These field studies also provide evidence for pollen "carryover" in *I. purpurea*.

1. INTRODUCTION

THE use of genetic markers and the development of appropriate genetic models have recently made possible precise estimates of plant mating system parameters (Brown and Allard, 1970; Clegg *et al.*, 1978). Quantitative studies of plant mating systems are based on the model of mixed selfing and random outcrossing. Mating is assumed to take place either by selfing, at a rate s , or by random outcrossing at a rate $t = (1 - s)$. For simplicity it is assumed that the rate of outcrossing is constant over maternal genotypes, and that there is random dispersal of outcross pollen genotypes over maternal parents (see Clegg, 1980 for a detailed discussion of the mixed mating model).

Rates of outcrossing may differ between maternal genotypes, however, as a result of differential attractiveness to pollinators (Levin, 1969; Kay, 1976; Brown and Clegg, 1981) or heritable differences in flower morphology (Breese, 1959; Jain, *et al.*, 1979, Ennos 1981). There are also conditions under which there will be non-random dispersal of outcross pollen genotypes over maternal parents. This is most likely to occur when a population is divided into a number of subpopulations, differing in genotype frequency, between which there is very limited pollen flow. Levin and Kerster (1968), Bradshaw (1972) and Schaal (1974) have all argued that substructuring of populations and limited amounts of gene flow will regularly be found in plant populations.

In this report, the effect of population substructuring on estimates of outcrossing rate is investigated using two approaches. The first involves a computer simulation of the mating process in subdivided populations. The second employs experimental populations of *Ipomoea purpurea* within which the spatial arrangement of genotypes has been manipulated. The

* Present address: Genetics Department, University of Newcastle, Ridley Building, Claremont Place, Newcastle Upon Tyne NE1 7RU, U.K.

effect of different spatial arrangements of genotypes on estimates of outcrossing rate has then been assessed.

2. MATERIALS AND METHODS

(i) *Simulation studies*

Each simulated population consisted of ten subpopulations, each with 50 maternal genotypes, yielding a total of 500 maternal genotypes. For each maternal genotype ten progeny were sampled assuming the mixed mating model. The sampling scheme assumed a constant outcrossing rate, t , for all subpopulations. Gene frequencies were assumed to vary among subpopulations. Denote the gene frequency for the i th subpopulation by p_i ($i = 1, 2, \dots, 10$) and assume a single diallelic locus. Then maternal genotype frequencies in the i th subpopulation were calculated assuming inbreeding equilibrium as,

$$\begin{aligned} f_i(AA) &= p_i^2 + p_i(1 - p_i)F \\ f_i(Aa) &= 2p_i(1 - p_i)(1 - F) \\ f_i(aa) &= (1 - p_i)^2 + p_i(1 - p_i)F, \end{aligned}$$

where $F = (1 - t)/(1 + t)$ (Haldane, 1924). The program first selected a maternal parent by drawing a pseudorandom number, y , uniformly distributed on $(0, 1)$. If, for example, $y < f_i(AA)$ a maternal parent of type AA was selected. Next, ten progeny were drawn independently. Again assuming the maternal parent was type AA , the offspring were type AA with probability $s + tp_i$ or Aa with probability $t(1 - p_i)$. Offspring genotypes were determined by successive draws of pseudo-random numbers uniformly distributed on $(0, 1)$. The sampling process was repeated until 50 families had been drawn following which the gene frequency was set to p_{i+1} and the sampling process was repeated for the $i + 1$ st subpopulation. The simulated data were analysed by the procedures given in Clegg *et al.* (1978) to give estimates of the apparent rates of outcrossing for comparison with their true values. Simulations were conducted for $t = 0.1, 0.3, 0.7$, and 1.0 . Values of \bar{p} ranged from $\bar{p} = 0.2$ to $\bar{p} = 0.5$, while values of σ_p^2 lay between $\sigma_p^2 = 0.0150$ and $\sigma_p^2 = 0.1275$.

(ii) *Experimental studies*

Populations of *Ipomoea purpurea*, a bee pollinated self compatible annual vine were employed. Flowers of this species open at dawn but close and wither by noon of the same day. The outcrossing rate in natural populations is approximately $t = 0.7$ (Ennos 1981).

I. purpurea populations are polymorphic at a diallelic, codominant esterase locus which can be scored in both adult and seedling material. Approximately 70 individuals of the two homozygous genotypes E^F/E^F and E^S/E^S were isolated from seed collections. On the evening before the experiment, plants were arranged in a grid outside the Botany greenhouses on the University of Georgia campus. Plants were placed 2 m apart, eight plants to a row. The number of rows varied from four to six depending on the number of individuals possessing flower buds ready to open the next morning. In all populations there were equal numbers of the two homozygous genotypes, but the spatial arrangement of genotypes was varied. In

design 1, genotypes were assigned randomly to a position in the grid. In design 2, they were assigned to opposite halves of the rows so that the population consisted of two adjacent blocks of homozygous genotypes. Plants were trimmed to one flower bud per individual to attempt to ensure equal contribution of maternal parents to the pollen pool. Experimental designs 1 and 2 were each replicated three times in a random sequence during August 1980.

The following morning pollinator visits were observed between 8 a.m. and 10 a.m. The grid position of each flower visited was noted. When flowers had begun to close, plants were transferred to the greenhouse and the flowers involved in the experiment were labelled. The seeds set were collected, and seven day old seedling progeny were scored for their genotype at the esterase locus.

3. RESULTS

(i) Simulation studies

The results are shown graphically in fig. 1. The relative lowering of the outcrossing rate, $(t - \hat{t})/t$, is plotted against the standardised variance in gene frequency between subpopulations $\sigma_p^2/\bar{p}(1 - \bar{p})$ where:

- t = true outcrossing rate
- \hat{t} = estimated outcrossing rate
- σ_p^2 = variance in gene frequency between subpopulations
- \bar{p} = mean gene frequency among populations.

The standardised variance is used because \bar{p} is not constant for all simulations.

Table 1 gives the results of a linear regression analysis for each simulated value of t . Note first that within the range of gene frequency variation

TABLE 1

(a) Regression analysis of relative lowering of outcrossing rate on standardized variance in gene frequency between subpopulations

Source of variation	df	SSQ	MS	F
Among groups	12	1.7880	0.1490	10.64***
Linear regression	1	1.5035	1.5035	58.05***
Deviations from regression	11	0.2845	0.0259	1.85
Within groups	43	0.6028	0.0140	
Total	55	2.3908		

*** $P < 0.001$.

(b) Regression analysis as a function of outcrossing rate. n denotes the number of simulated populations for each value of t

Outcrossing rate t	Slope b	SE b	Intercept a	SE a	n
0.1	0.8840***	0.1220	0.0634	0.0326	11
0.4	1.1779***	0.0427	0.0269*	0.0111	10
0.7	1.5133***	0.1211	-0.0244	0.0815	10
1.0	1.3410***	0.0737	0.0644**	0.0210	10
	Significance of difference from $b = 0$		Significance of difference from $a = 0$		

* $P < 0.05$.
 ** $P < 0.01$.
 *** $P < 0.001$.

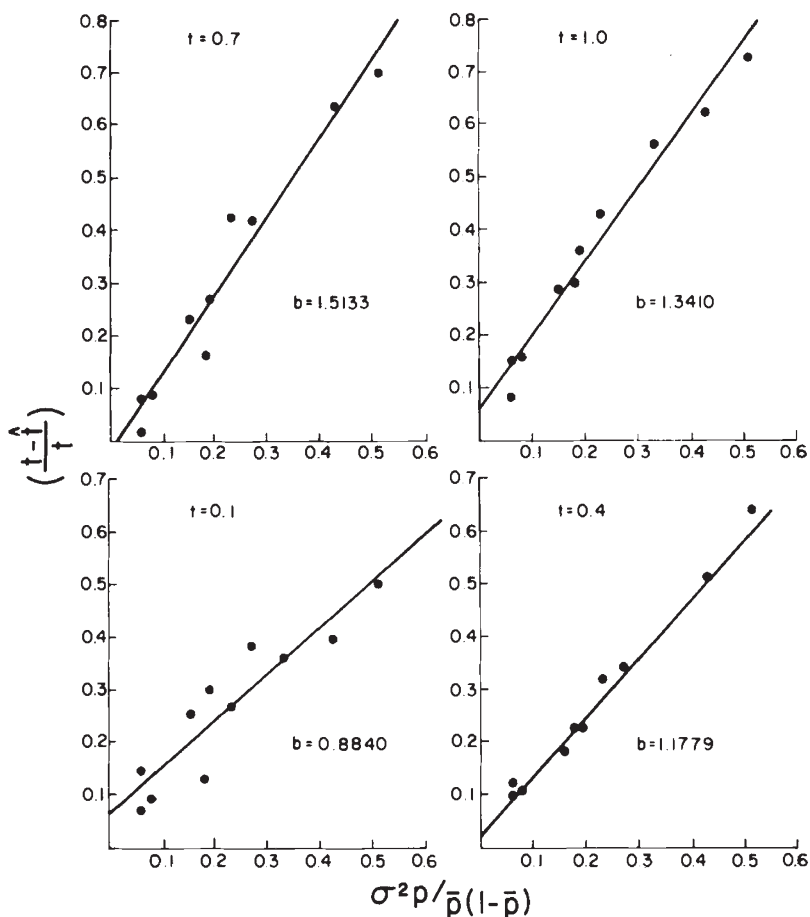


FIG. 1.—Regression of relative lowering of outcrossing rate $\left(\frac{t-f}{t}\right)$ on standardised variance in gene frequency between subpopulations $\left(\frac{\sigma_p^2}{\bar{p}(1-\bar{p})}\right)$ for a number of values of t .

simulated $\left[\frac{\sigma_p^2}{\bar{p}(1-\bar{p})}\right]$, a linear regression model adequately accounts for the data, because the deviation from regression is not significant. Despite this good fit the estimated intercept values deviate significantly from zero for $t=0.4$ and $t=1.0$. This may indicate a slight downward bias in the estimation procedure or it may simply indicate that linear extrapolation back to zero is inappropriate. Some evidence in favour of this latter interpretation comes from 20 additional simulated data sets using $t=0.7$ and $\sigma_p^2=0$ (*i.e.*, no gene frequency variation among subpopulations). In these simulations $\hat{t}=0.7$ within the limits of sampling error in all cases. Eleven of the estimates exceeded 0.7 while nine runs gave estimates slightly below 0.7.

What is readily evident from fig. 1 is that the estimated outcrossing rate is seriously biased downwards by gene frequency variation among subpopulations. Moreover, the degree of bias (slope of the regression line) increases as the true value of t increases (at least up $t=0.7$). In all

cases large variation in gene frequency, $\sigma_p^2/\bar{p}(1-\bar{p}) > 0.5$, can cause the estimated outcrossing rate to deviate from the true value by as much as 50 per cent of the true value.

(ii) *Experimental studies*

(a) *Pollination*

Flowers were visited both by bees of the genus *Bombus* (chiefly *B. pennsylvanicus*) and by butterflies. The distribution of total flight distances is given in fig. 2. There was no significant difference between the distributions of bee and butterfly flight distances ($\chi^2_{(3)} = 0.96$). Some 94 per cent of flights did not exceed 4 m, suggesting that pollen flow within the experimental population is likely to be very restricted.

A comparison of pollinator behaviour in experimental designs 1 and 2 showed that the distributions of pollinator flight distances did not differ significantly ($\chi^2_{(3)} = 5.50$). The mean number of pollinator visits per seed set was 0.2315 for populations with random arrangements of genotypes, and 0.4721 for populations with genotypes arranged in two blocks.

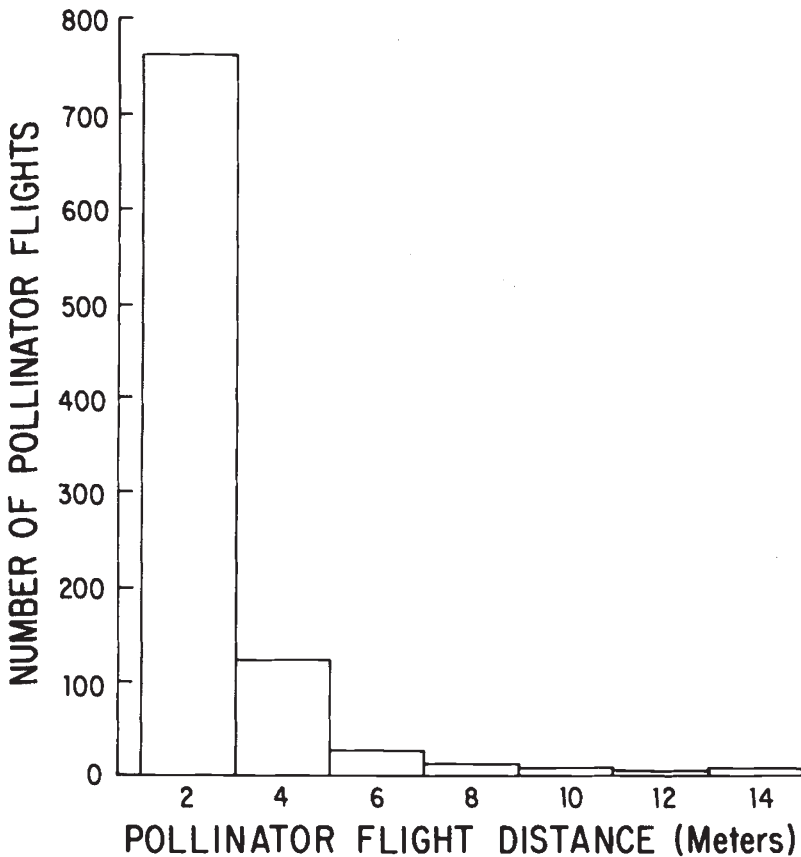


FIG. 2.—Distribution of pollinator flight distances in experimental populations of *I. purpurea*.

(b) *Outcrossing rate*

The total numbers of progeny of each genotype derived from the homozygous parents are given in table 2. The outcrossing rate is found using the maximum likelihood estimator $\hat{t} = h_A/(1-p)$ where h_A is the

TABLE 2
Numbers of each genotype in progeny from experimental populations

Experimental design	Maternal genotype	Progeny genotype			Total
		E^F/E^F	E^F/E^S	E^S/E^S	
Random (design 1)	E^F/E^F	112	50	—	162
	E^S/E^S	—	71	77	148
	Total	122	121	77	310
Blocks (design 2)	E^F/E^F	112	42	—	154
	E^S/E^S	—	48	96	144
	Total	112	90	96	298

proportion of A/a heterozygotes among the progeny of A/A homozygotes, and p is the frequency of allele A in the outcross pollen pool. There are two degrees of freedom for calculating two mating system parameters. We may make either of two assumptions:

(1) The frequency of allele A is the same in the pollen pool as in the maternal parent population *i.e.* 0.5. The outcrossing rate for each of the two homozygous genotypes may then be estimated for each experimental design.

(2) The outcrossing rates for the two homozygous genotypes are equal. Under this assumption the estimators are $\hat{t} = h_a + h_A$ and $\hat{p} = h_a/(h_a + h_A)$ where h_a is the proportion of heterozygous offspring among maternal parents of genotype a/a .

Under assumption (1) the results are accounted for entirely in terms of differences in outcrossing rate between the homozygous genotypes. Under assumption (2) they are due solely to differences in gene frequency between the maternal parent population and the uniting pollen pool. Neither of these processes alone is likely to account for the results obtained.

If we assume equal contribution of maternal genotypes to the pollen pool, the estimated rate of outcrossing is higher for both homozygous genotypes when they are randomly arranged in the population, than when ordered in two blocks (table 3a). For the E^S/E^S homozygote this difference is significant ($P < 0.01$). In both experimental designs the estimated outcrossing rate for the E^S/E^S maternal genotype is greater than that of the E^F/E^F maternal genotype, and this difference is highly significant when the genotypes are randomly arranged in the population ($P < 0.005$).

When it is assumed that the two homozygous maternal genotypes outcross at the same rate within a particular design, the estimated outcrossing rate is again higher for random spatial arrangement of genotypes than for genotypes arranged in two blocks $p < 0.02$ (table 3(b)). A second interesting feature of the random design is that \hat{p} differs significantly from $\frac{1}{2}$, the value expected based on parental genotypic frequencies. This is not true of the blocks design where $\hat{p} = 0.55 \pm 0.0439$. Whatever the assumptions of the estimation procedure therefore, the estimated outcrossing rate

TABLE 3
Estimates of outcrossing rates

(a) *Assuming equal contribution of maternal parents to uniting pollen pool*

Experimental design	Maternal genotype	Outcrossing rate \hat{t}	S.E. \hat{t}
Random (design 1)	E^F/E^F	0.6173	0.0758
	E^S/E^S	0.9595	0.0821
Blocks (design 2)	E^F/E^F	0.5455	0.0718
	E^S/E^S	0.6667	0.0786

(b) *Assuming equal outcrossing rates for homozygous genotypes*

Experimental design	Gene frequency \hat{p}	Outcrossing rate \hat{t}
Random (design 1)	0.6085 ± 0.0352	0.7884 ± 0.0546
Blocks (design 2)	0.5500 ± 0.0439	0.6061 ± 0.0531

is always higher under random design 1 than under the non-random spatial arrangements of genotypes used in design 2. Clearly the spatial arrangement of genotypes has a significant effect on the estimated rate of outcrossing.

(c) *Gene flow*

Gene flow may readily be studied in design 2 by observing the distribution of heterozygous progeny produced at various distances from the boundary between the two blocks of homozygous genotypes. The distribution of distances moved from this boundary by genes is given in fig. 3. Also given is the distribution of distances from the boundary moved by pollinators passing from one block of genotypes to the other. If there is no "carryover" of pollen, the distributions should be identical. They are however very significantly different ($\chi^2_{(3)} = 28.24, P < 0.001$). The number of genes transferred falls off far less sharply with distance from the boundary than does the number of pollinator moves. This suggests that pollen is often carried beyond the next flower which the pollinator visits.

4. DISCUSSION

The simulation studies reported above show that when there is population subdivision and restricted pollen flow in plant populations, estimates of outcrossing rate will not reflect the true rate at which outcross pollen is received by maternal genotypes. The more substructured the population, the lower will be the estimated rate of outcrossing. If there is no pollen flow between subpopulations (as in the simulation studies) the relative lowering of the outcrossing rate due to population subdivision appears to increase linearly with the standardised variance in gene frequency between subpopulations.

These results are borne out qualitatively by the studies on the experimental populations of *Ipomoea*. In the substructured population (design 2) the estimated outcrossing rate is lower than in the population with no spatial heterogeneity of genotype frequencies (design 1). Clearly the effect of spatial arrangement of genotypes on estimates of outcrossing rate can be significant.

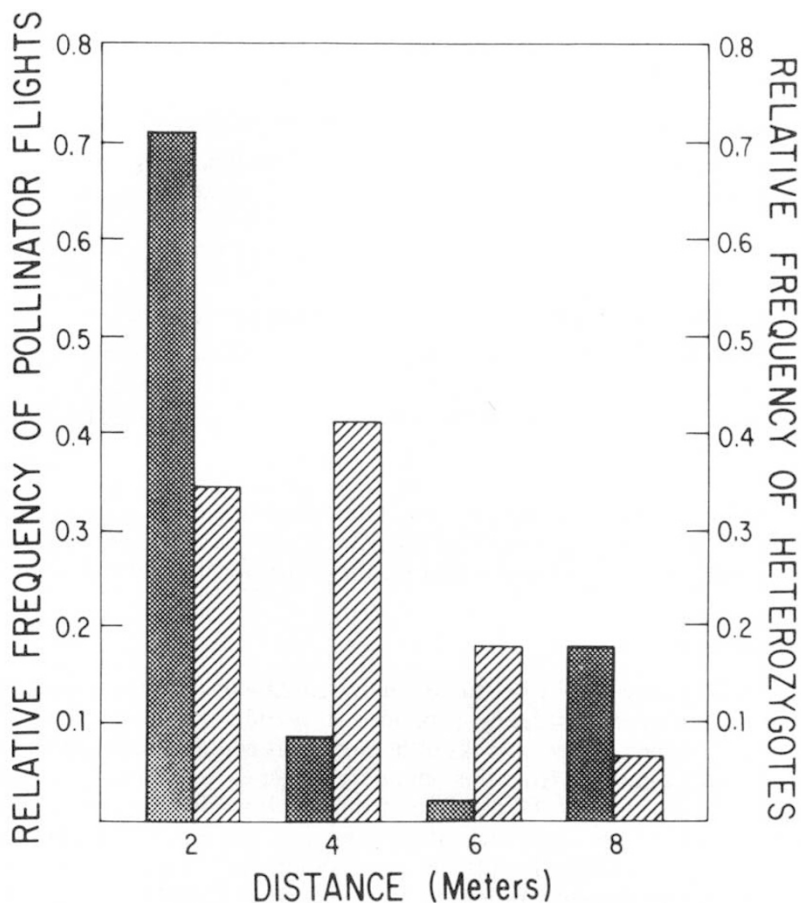


FIG. 3.—Comparison of pollinator flight distances and gene flow in experimental populations of *I. purpurea*.

The influence of spatial arrangement of genotypes on estimates of outcrossing rate may help to account for two results which have emerged from quantitative studies of plant mating systems. The first is that different marker loci may yield different estimates of outcrossing rate for the same population (Brown, *et al.*, 1975; Sanders and Hamrick, 1980). Differences between loci in the spatial arrangement of genotypes could account for this result. The second finding is that self-incompatible species often show a significant deficit of heterozygotes compared with that expected under random outcrossing (Schaal and Levin, 1976; Levin, 1977, 1978; Valizadeh, 1977). If populations of these self incompatible species are substructured and if pollen flow is restricted, the realised rate of outcrossing will fall below $t = 1.0$, resulting in a deficit of heterozygotes compared with that expected under random outcrossing (*i.e.*, the Wahlund effect). In support of this explanation Brown (1979) has pointed out that in agricultural populations, where there is little opportunity for population substructuring to develop, self-incompatible species show no such deficit of heterozygotes.

An interesting point to emerge from the analysis of the experimental populations is that there is a significant difference in the outcrossing rate of the two homozygous genotypes used and/or there are differences in gene frequency between the maternal parents and the pool of outcross pollen. Unfortunately the experimental design does not give us enough information to distinguish between these two alternatives. The homozygous maternal parents were taken from two populations approximately half a mile apart. Differences in outcrossing rate between maternal genotypes may reflect differences in outcrossing rate between the two parental populations. Alternatively there may be differential production of pollen by the two genotypes, or differential selection of pollen genotypes in the pollen pool.

Gene flow via pollen dispersal does not follow the same distribution as pollinator flight distances within the experimental populations. The distance moved by genes tends to be greater than the distance moved between subsequent flowers by pollinators. A likely explanation is that only a fraction of the pollen picked up by a bee is deposited on the stigma of the next flower visited, and consequently, there is significant "carryover" of pollen to flowers subsequently visited. This phenomenon has previously been documented by Schaal (1980) in an experimental population of *Lupinus nanus*. The practice of inferring pollen flow from observations of pollinator movement will, therefore, lead to an underestimate of gene flow and neighbourhood size (Wright, 1946; Levin and Kerster, 1968). Nevertheless pollen flow is sufficiently restricted in the experimental populations of *Ipomoea* for differences in the spatial arrangement of genotypes to have a significant effect on estimates of outcrossing rate.

Acknowledgements.—This work was conducted while R.A.E. was in receipt of a N.A.T.O. postdoctoral fellowship.

5. REFERENCES

- BRADSHAW, A. D. 1972. Some of the evolutionary consequences of being a plant. *Evol. Biol.*, 5, 25–47.
- BREESE, E. L. 1959. Selection for different degrees of outbreeding in *Nicotiana rustica*. *Ann. Bot. N.S.*, 23, 331–344.
- BROWN, A. H. D. 1979. Enzyme polymorphism in plant populations. *Theor. Pop. Biol.*, 15, 1–42.
- BROWN, A. H. D., AND ALLARD, R. W. 1970. Estimation of mating system in open pollinated maize populations using isozyme polymorphisms. *Genetics*, 66, 133–145.
- BROWN, A. H. D., MATHESON, A. C., AND ELDRIDGE, K. G. 1975. Estimation of the mating system in *Eucalyptis obliqua* L'Herit. using allozyme polymorphisms. *Aust. J. Bot.*, 23, 931–949.
- BROWN, B., AND CLEGG, M. T. 1981. The influence of flower color polymorphism on genetic transmission in a natural population of the common morning glory, *Ipomoea purpurea*. *Evolution* (submitted).
- CLEGG, M. T. 1980. Measuring plant mating systems. *Bioscience*, 30, 814–818.
- CLEGG, M. T., KAHLER, A. L., AND ALLARD, R. W. 1978. Estimation of life cycle components of selection in an experimental plant population. *Genetics*, 89, 765–792.
- ENNOS, R. A. 1981. Quantitative studies of the mating system in two sympatric species of *Ipomoea* (Convolvulaceae). *Genetica*, 57, 93–98.
- HALDANE, J. B. S. 1924. A mathematical theory of natural and artificial selection. Part 11. The influence of partial self-fertilisation, inbreeding, assortative mating and selective fertilisation on the composition of Mendelian populations and on natural selection. *Proc. Camb. Phil. Soc., Biol. Sci.*, 1, 158–163.

- JAIN, K. B. L., SCHALLER, C. W., AND JAIN, S. K. 1979. Genetic variation in the outcrossing rates in barley. *Genetica*, 50, 41-49.
- KAY, Q. O. N. 1976. Preferential pollination in yellow-flowered morphs of *Raphanus raphanistrum* by *Pieris* and *Eristalis* spp. *Nature*, 261, 230-232.
- LEVIN, D. A. 1969. The effect of corolla colour and outline on interspecific pollen flow in *Phlox*. *Evolution*, 23, 444-455.
- LEVIN, D. A. 1977. The organisation of genetic variability in *Phlox drummondii*. *Evolution*, 31, 477-494.
- LEVIN, D. A. 1978. Genetic variation in annual *Phlox*: Self compatible versus self incompatible species. *Evolution*, 32, 245-263.
- LEVIN, D. A., AND KERSTER, H. W. 1968. Local gene dispersal in *Phlox*. *Evolution*, 22, 130-139.
- SANDERS, T. B., AND HAMRICK, J. L. 1980. Variation in the breeding system of *Elymus canadensis*. *Evolution*, 34, 117-122.
- SCHAAL, B. A. 1974. Isolation by distance in *Liatris cylindracea*. *Nature*, 252, 703.
- SCHAAL, B. A. 1980. Measurement of gene flow in *Lupinus texensis*. *Nature*, 284, 450-451.
- SCHAAL, B. A., AND LEVIN, D. A. 1976. The demographic genetics of *Liatris cylindracea* Michx. (Compositae). *Amer. Nat.*, 110, 191-206.
- VALIZADEH, M. 1977. Esterase and acid phosphatase polymorphism in the fig tree (*Ficus carica* L.). *Biochem. Genet.* 15, 1037-1047.
- WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics*, 31, 39-50.