

# The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of *S*-allele frequencies in populations at equilibrium

R. J. BROOKS, A. M. TOBIAS & M. J. LAWRENCE†\*

*Operational Research Group, School of Manufacturing and Mechanical Engineering and*

*†Wolfson Laboratory for Plant Molecular Biology, School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.*

The effects of limited pollen and seed dispersal, of overlap between generations and of variation in plant size on the steady-state variance of *S*-allele frequencies have been investigated in a simulated population of size 3840 containing 16 *S*-alleles whose initial frequencies were exactly equal. Simulations were run with each of the 16 possible combinations of these four factors to investigate their effects on the time for the population to reach steady state and on the average variance of *S*-allele frequencies in steady state. The time to steady state appeared to be relatively unaffected by any of the factors and was about 50 generations. However, the steady-state variance was markedly affected, with variation in plant size increasing this variance by an average of 228 per cent and overlapping generations decreasing the variance by an average of 30 per cent. The effects of limited pollen and seed dispersal were individually small, although their combined effect was to increase the steady-state variance by an average of 12 per cent. Limited seed and pollen dispersal, when combined with variation in plant size, caused the alleles to cluster. The four factors together caused a large increase in the average steady-state variance. Furthermore, even when a population is in steady state, the variance for a particular generation can be considerably greater than this average value. Consequently the frequencies of the *S*-alleles of a population in steady state can be very different. It is possible, therefore, that the large variation in *S*-allele frequencies found in samples taken from *Papaver rhoeas* populations is consistent with their being in steady state.

**Keywords:** computer simulation, *Papaver rhoeas*, population genetics, self-incompatibility polymorphism, steady state, variance of *S*-allele frequency.

## Introduction

In a population of a species with a gametophytic system of self-incompatibility, frequency-dependent selection will tend to push the frequencies of the incompatibility alleles towards a value close to their mean frequency (Wright, 1939). In the absence of other kinds of selection, the frequencies of each of

the alleles are expected to fluctuate around this mean. Campbell & Lawrence (1981) and Lawrence & O'Donnell (1981) found, however, a significant variation in *S*-allele frequencies in each of the samples taken from three British natural populations of *Papaver rhoeas*. A preliminary investigation of the polymorphism by computer simulation indicated that, although there was considerable variation in *S*-allele frequencies in populations at equilibrium, its magnitude was apparently insufficient to account

\*Correspondence.

for the variation observed in the poppy samples (Lawrence *et al.*, 1994). The populations simulated, however, were panmictic with nonoverlapping generations. This assumes a single plant size (in that each plant produces the same amount of seed and pollen) and a uniform dispersal distribution for the pollen and seed. None of these conditions is likely to hold in practice. Pollen and seed are unlikely to travel far from their source; the very long-lived seed of the seed bank, from which flowering plants arise, is likely to originate from several generations, not just the immediately previous one; and the empirical evidence indicates quite clearly that there is a marked variation in seed output in this species (Lawrence *et al.*, 1994). It is possible, therefore, that one or other of these four factors, either singly or in combination, could inflate the steady-state variance of *S*-allele frequencies to the extent that they could account for the variation in frequencies observed in the *P. rhoeas* samples.

The aim of the investigation reported here was to quantify the effects of the four factors by a comparison of models containing different combinations of the factors to assess the importance of the factors on the population genetics of the polymorphism. Furthermore, each factor effectively adds a modelling level of detail, so that the relationship between the level of detail and model performance could be investigated. The implications of the results for modelling methodology will subsequently be discussed elsewhere.

## Simulated population

### *Size, density and number of alleles*

The population simulated contained a constant number of 3840 plants arranged on a rectangular grid consisting of 60 rows and 64 columns. The distance between plants within columns and rows was taken to be 0.5 m, which although a little less than the average density of  $9\text{ m}^{-2}$  reported by Lawrence *et al.* (1994), is arguably more typical of natural populations in general. The original intention was that the population should contain 31 alleles, as with the simulations of Lawrence *et al.* (1994), but alleles were lost in the runs of some of the models investigated and so the number of alleles was reduced to 16 to avoid this problem. A population size of 3840 is probably small for this species, but is sufficient for present purposes and allowed more runs to be carried out. The effects of increasing the population size and the number of alleles will be considered elsewhere. Each run was started

with a population containing 32 replicates of each of the 120 possible genotypes, giving 480 copies of each of the 16 *S*-alleles. To ensure an approximately even spatial distribution of genotypes in the first generation, the grid was divided into 32 blocks, each of which contained a single plant of each genotype whose position within the block was assigned at random. A population containing an exactly equal number of alleles and genotypes is not in a steady state, and so this allows the time, in generations, for the population to reach steady state to be measured and compared for the different models. Mutation was not simulated.

### *Dispersal of pollen*

Although the flowers of *P. rhoeas* are pollinated by bees, nothing is known about the parameters of the distribution of the distance travelled by the pollen for this species. However, a number of studies on the movement of pollen by insects have been carried out on other species which indicate that flight distances are highly leptokurtic, because most flights occur between near neighbours (see reviews by Levin & Kerster, 1974; Handel, 1983). In the absence of information about *P. rhoeas*, it has been assumed that the mean flight distance of bees is of the order of 1.5 m (i.e. three plant spaces), that most of the pollen is deposited after one flight and that the dispersal distance of the pollen follows a negative exponential distribution. As values of less than 0.25 m are not valid, because this would result in self-pollination, a negative exponential distribution with a mean of 1.25 m was used to investigate the effect of limited pollen dispersal on the polymorphism, to which 0.25 m has been added, giving a mean dispersal distance of 1.5 m.

### *Dispersal of seed*

A study of seed dispersal for *P. rhoeas* has been carried out by Blattner & Kadereit (1991). Their histogram of dispersal distance has a peak at about 0.5 m and a mean distance of about 1.1 m. The shape of the histogram would appear to match more closely the curve from the tilted Gaussian plume model (Okubo & Levin, 1989) than the more commonly used inverse power law or negative exponential curves. The histogram can be approximated quite closely by a lognormal distribution and so a lognormal distribution based on the histogram has been used to model seed dispersal here. The normal distribution used to generate the lognormal had a mean of 4.4 and a standard deviation of 0.7, giving a

mean of 1.04 m, a standard deviation of 0.83 m and a kurtosis value of 17.8 for the lognormal distribution. A uniform distribution for the direction of travel was used for both the seed and the pollen (i.e. they were equally likely to travel in any direction).

#### *Generation overlap*

The extent of the overlap between generations of *P. rhoeas* is not known, but could, in principal, be quite large, because dormant seed in undisturbed soil can remain viable for many years. Hence, the seed bank, by which the species maintains itself between flowering episodes, could contain seed produced by several previous generations of flowering plants. If the probability of a given seed germinating is a constant, independent of the age of the seed, a geometric distribution is appropriate. It has been assumed that this probability is 0.7, so that the probability of a seed germinating after  $t$  generations of seed rain is  $0.3^{t-1} \times 0.7$ .

#### *Distribution of plant size*

There is good evidence of a large variation in plant size in this and the related species, *P. dubium*, the chief cause of which is that plants which originate from seed which germinates in the autumn and survive the winter are much larger than those which arise from seed which germinates in spring (Arthur *et al.*, 1973 and Mackay, 1981 for *P. dubium*; Ooi, 1970 and Lawrence *et al.*, 1994 for *P. rhoeas*). Large plants bear many more flowers than small plants and, hence, produce much more pollen and seed than the others. Arthur *et al.* (1973), for example, found that autumn plants of *P. dubium* produced 69 times as much seed as spring plants in one year and 21 times as much in the following year. The autumn plants made up 21.5 per cent and 17.6 per cent of the total number of flowering plants in the two years, respectively. For convenience, and in the absence of further data for this species, we have assumed that there are only two sizes of plant in the simulated populations, 20 per cent of which are large and that these large plants produce 20 times as much pollen and seed as small plants.

#### *Models investigated*

The models built contained different combinations of the four factors, overlap of generations (i.e. seed dormancy, denoted D), limited pollen dispersal (P), limited seed dispersal (S), and variable plant size (the existence of large plants, L). Only the presence

or absence of each factor was considered, giving 16 alternative models. Each model is denoted by the letters of the factors it contains (for example, the model containing all the factors is denoted DPSL), with the model containing no factors being denoted 'NIL'. In the models without seed dormancy, the successive populations generated by the program in any run are successive generations in the exact sense, whereas in those with dormancy, a generation is equal to 1.43 flowering episodes; it is convenient, however, to refer to the latter as generations in what follows.

#### *The program*

The simulation program, written in FORTRAN, used the well-established NAG library routine, G05CAF to generate random numbers. The program kept a record in each run of the size and genotype of the plant at each location in each of the ten previous generations. The procedure used to generate the genotype of each plant in each generation of the most complex model, DPSL, began by selecting the seed rain from which the first plant in the grid arose using the dormancy distribution. The maternal parent was then chosen from this generation, using the seed dispersal distribution to determine the origin of the seed relative to the location of the new plant. The paternal parent was chosen in a similar way, using the pollen dispersal distribution, to determine its location relative to the maternal parent. As large plants were assumed to produce 20 times as much pollen and seed as small plants, the probability of selecting such a plant during this process had to be multiplied by 20 and the sum of the probabilities of selecting any plant rescaled so that it equalled unity. This was achieved by rejecting the selected plant if it was small with a probability of 19/20 and repeating the selection. Once the parents had been chosen, one of the paternal alleles, chosen at random, was compared with the maternal alleles for compatibility. If this allele was the same as one of the maternal alleles, the paternal parent was discarded and a new male parent chosen. The original plant was not precluded when the new plant was chosen. The other allele in the original paternal parent was not tested for compatibility, as this would give the paternal alleles in half-compatible crosses twice the probability of being selected as those of fully compatible crosses. The maternal parent, on the other hand, was retained on the assumption that the presence of incompatible pollen on its stigma would not reduce the amount of seed it produced. Once a compatible paternal allele had been found,

one of the maternal alleles was chosen at random to form the new zygote. New plants were designated large or small at random, with a probability of 0.2 of each being large. This entire procedure was then repeated, working systematically through the grid, until a new population of 3840 plants had been produced. Simpler models were run by replacing the appropriate subroutines.

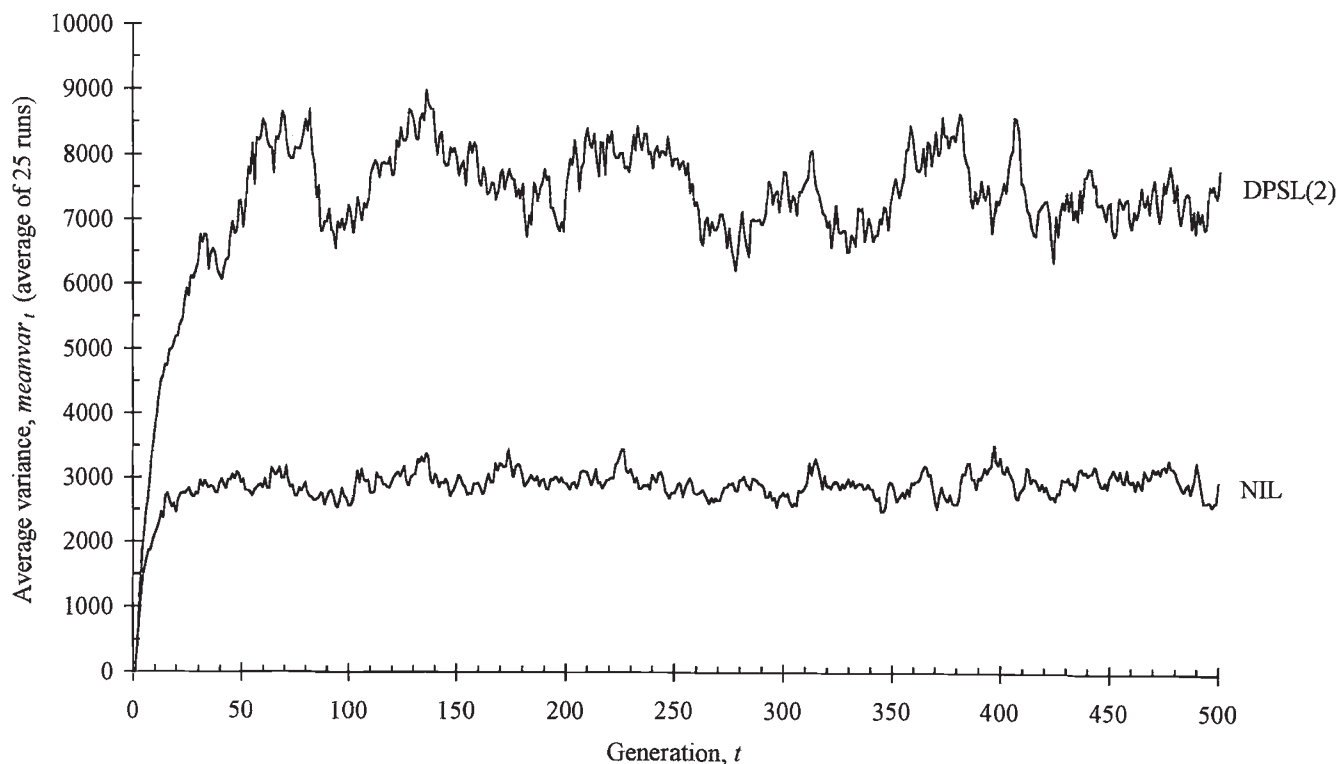
The program calculated the variance of allele numbers,  $var_t$  (the variance of their absolute rather than their relative frequency) in each generation,  $t$ , the divisor of which was 16, because it concerns the population rather than a sample drawn from it. Each of the 16 models was run 25 times for 500 generations. An allele was lost in two of the runs of the PL model and one of the runs of the PSL model which were, therefore, replaced by extra runs in order to ensure comparability overall. The program also issued a message if the frequency of any allele fell below 50 (mean frequency = 480); this occurred in three runs of the PSL, and one each of the PL and SL models, in addition to runs in which an allele was lost. The program was run on an IBM3090 mainframe computer. The most complex models required up to 7 min of c.p.u. time per run, which limited the number of runs that could be carried out.

## Results

### *Estimation of the time to steady state and the average variance of the steady-state allele frequency*

Figure 1 shows the variance of allele frequency, averaged over 25 runs ( $meanvar_t$ ), plotted against the generation,  $t$ , for the NIL and DPSL(2) models (the DPSL model was run twice, as explained below). The curves for all of the models had a similar shape, in that this variance first increases rapidly from its initial value of zero ( $t = 1$  for the first generation) before levelling off, after which it appears to fluctuate about a constant value when the population has achieved a steady state. The usual procedure used to estimate the mean steady-state value is to estimate first the time taken for the model to reach steady state,  $T_{SS}$ , (often called the warm-up period) and then to use the average output value after  $T_{SS}$  as the estimate for the steady-state variance,  $var_{SS}$  (Law & Kelton, 1991).

Estimates of the time to steady state were obtained by using Welch's (1983) graphical procedure, in which a centred moving average is calculated for each generation. This is a conservative procedure in the sense that it identifies the latest



**Fig. 1** Average variance values for the DPSL(2) and NIL models.

time to reach steady state for the 25 runs of each model, rather than the average time. To obtain a reasonably smooth curve, it was found necessary to calculate a moving average of 20 values before and after each generation (40 values in all). The curves of even this broad moving average still showed sizeable fluctuations, and so the estimates of  $T_{SS}$  are rather subjective. Estimates of the time to steady state are shown for each of the 16 models investigated in Table 1.

As the times to steady state for all but one of the models were considerably less than 200 generations, the average steady-state variance of allele number,  $var_{SS}$ , was estimated as the average of all values of  $var_i$  for generations 201–500 over all 25 runs (7500 values in all), which is, of course, equivalent to taking the average of the  $meanvar_i$  values over these generations. However, for the first set of runs of the DPSL model, DPSL(1), this average was calculated on the values of  $var_i$  from generations 301 to 500 (5000 values). Standard deviations of the individual  $var_i$  values in the steady-state period and 95 per cent confidence intervals of the steady-state  $var_{SS}$  values were calculated in the usual way; these are shown, together with the estimates of  $var_{SS}$ , in Table 1.

#### Time to steady state, $T_{SS}$

The entries in Table 1 indicate that the time to steady state for most models is approximately 40–60

generations. Six of the models (L, S, P, PSL, DSL and, in particular, DPSL(1)), however, apparently took considerably longer to achieve steady state. Although the DPSL(1) moving average curve only levelled out after 230 generations, the moving average values were higher for generations 110–230 than the average value for the generations after 230 (i.e. it appeared to overshoot the steady state). Twenty-five further runs of this model, using a different initial random number, were therefore carried out to find out whether this was an intrinsic property of this model or a chance effect. The time to steady state of this second set of simulations, DPSL(2), was only 70 generations and the steady-state values for the second run are comparable with the values of the first run between 70 and 230 generations and generally higher than the original values after 230 generations. It therefore seems likely that the values for the original run are slightly lower than the true steady-state value for generations 230–500 because of chance effects and that the original model had in fact reached steady state after about 110 generations.

As there is no obvious relationship between the factors of the five remaining models that yielded  $T_{SS}$  estimates of 90 generations or more and because the  $T_{SS}$  estimates are not very exact, it is likely that had it been possible to carry out many more runs than 25 for each model their times to steady state would have turned out to be similar to the other models.

**Table 1** Values from the simulation runs for the time to steady state,  $T_{SS}$ , the average steady-state variance,  $var_{SS}$ , the 95 per cent confidence interval for  $var_{SS}$  and the steady-state standard deviation,  $stdev(var_i)$

Model	$T_{SS}$ (no. of generations)	$var_{SS}$	95% confidence interval for $var_{SS}$	$stdev(var_i)$
DPSL(1)	230	7036	± 399 (± 5.7%)	2588
DPSL(2)	70	7427	± 419 (± 5.6%)	2750
DPS	50	2198	± 100 (± 4.5%)	822
DPL	50	6740	± 334 (± 5.0%)	2472
DSL	130	6358	± 273 (± 4.3%)	2337
PSL	110	11683	± 789 (± 6.7%)	4531
DP	40	2102	± 78 (± 3.7%)	771
DS	50	2070	± 91 (± 4.4%)	780
DL	50	6658	± 359 (± 5.4%)	2504
PS	60	3170	± 94 (± 3.0%)	1092
PL	50	9251	± 399 (± 4.3%)	3332
SL	40	9200	± 280 (± 3.0%)	3410
D	40	2080	± 64 (± 3.1%)	749
P	140	2840	± 100 (± 3.5%)	1036
S	110	2991	± 143 (± 4.8%)	1126
L	90	9725	± 318 (± 3.3%)	3438
NIL	50	2914	± 96 (± 3.3%)	1042

See text for further details.

*Average steady-state variance of allele frequency,  $var_{SS}$*

There is no doubt that there are significant differences between the models in respect of their average steady-state allele frequency variance,  $var_{SS}$ . The effect of each of the factors can be shown most clearly by calculating the differences in the  $var_{SS}$  estimates for models which differ by only one factor (i.e. the change in the  $var_{SS}$  estimates when a single factor is added); these values are given in Table 2. The factor that has the greatest effect is variation in plant size (L) which, on average, increases the steady-state variance by 228 per cent. Overlapping generation (D), on the other hand, decreases this variance by an average of 30 per cent. Each of the percentage changes in the  $var_{SS}$  values caused by adding L are similar, as are each of the changes caused by adding D, indicating a systematic effect. On the other hand, there is not a consistent percentage change from adding either P or S and these values are quite small. However, the joint effect of adding both P and S causes a small increase in the steady-state variance in each case with an average increase of 12 per cent. It follows that the PSL model has the largest and the D model the smallest  $var_{SS}$ . In contrast to the estimates of  $T_{SS}$ , those of  $var_{SS}$  are much more precise, with the largest 95 per cent confidence interval (for PSL) being only 13.4 per cent of the estimate.

While the effective population size,  $N_e$ , of the NIL model is the same as its census size,  $N$ , of 3840, this is, of course, not so with the other models. It is possible, therefore, that the differences between the steady-state variances of the models result solely from differences in the effective size of the simulated populations. The calculation of  $N_e$  for a self-incompatible species, when two or more factors are involved, is not a straightforward task and has, therefore, not been attempted here.

*Steady-state standard deviation of  $var_t$*

The steady-state standard deviations,  $stdev(var_t)$ , given in the last column of Table 1 for each model, are the standard deviations of all of the steady state  $var_t$  values (i.e. the standard deviation of the 7500 values of  $var_t$  for generations 201–500 for the 25 runs). These standard deviations are large, which indicates that even when a population has achieved steady state, its  $var_t$  values will vary very considerably over generations (although successive values are correlated). The magnitude of this fluctuation of  $var_t$  values is such that it is evident even when they are

**Table 2** Effect on  $var_{SS}$  of adding the model factors

Previous model	Effect of adding factor D			Effect of adding factor P			Effect of adding factor S			Effect of adding factor L				
	Increase in $var_{SS}$	Increase in $var_{SS}$ (%)	Previous model	Increase in $var_{SS}$	Increase in $var_{SS}$ (%)	Previous model	Increase in $var_{SS}$	Increase in $var_{SS}$ (%)	Previous model	Increase in $var_{SS}$	Increase in $var_{SS}$ (%)	Previous model	Increase in $var_{SS}$	Increase in $var_{SS}$ (%)
NIL	-834	-29	NIL	-74	-3	NIL	77	3	NIL	6811	234	NIL	6811	234
P	-738	-26	D	22	1	D	-10	0	D	4578	220	D	4578	220
S	-921	-31	S	179	6	P	330	12	P	6411	226	P	6411	226
L	-3067	-32	L	-474	-5	L	-525	-5	S	6209	208	S	6209	208
PS	-972	-31	DS	128	6	DP	96	5	DP	4638	221	DP	4638	221
PL	-2511	-27	DL	82	1	DL	-300	-5	DS	4288	207	DS	4288	207
SL	-2842	-31	SL	2483	27	PL	2432	26	PS	8513	269	PS	8513	269
PSL	-4256	-36	DSL	1069	17	DPL	687	10	DPS	5229	238	DPS	5229	238
Average % increase		-30	Average % increase		6	Average % increase		6	Average % increase		6	Average % increase		228

See text for further details.

averaged over 25 runs as in Fig. 1. Figure 2 shows the  $var_t$  values of a single run of the DPSL(2) model and a single run of the NIL model. The most striking feature of these results, however, is that there is a strong linear relationship between the standard deviation of these variances and their mean value,  $var_{SS}$ , such that models with higher  $var_{SS}$  values also have higher values of  $stdev(var_t)$ . Linear regression analysis of  $stdev(var_t)$  against  $var_{SS}$  with the constant fixed at zero gives a regression coefficient of 0.37 and an  $r^2$  value of 0.996. We will show elsewhere, in a theoretical analysis, why this relationship exists.

## Discussion

### *Time to steady state*

Although the estimates of the time to steady state,  $T_{SS}$ , are not very precise, there is little evidence in these simulations that any of the four factors, either singly or in combination with the others, affects the time to steady state, which appears to be about 50 generations (although for each model, the average variance is close to the steady state after 20 generations, as in Fig. 1). Lawrence *et al.* (1994) estimated

that it took 70–80 generations to reach equilibrium in a simulated population of a similar size (3720) to that of the present investigation (3840). Their population, however, contained 31 alleles, rather than 16, as here. They suggested that because the strength of the frequency-dependent selection that maintains the polymorphism becomes attenuated as the number of alleles in a population increases a population containing a small number of alleles would achieve equilibrium more quickly than one of the same size containing a larger number of alleles. A comparison of their results with those of the present investigation is consistent with this expectation. We will show elsewhere that doubling the number of alleles in a population from 16 to 32 approximately doubles the time to steady state.

### *Average steady-state variance of allele frequency*

By contrast to the time to steady state, the average steady-state variance,  $var_{SS}$ , can be estimated quite accurately from the simulations and is considerably affected by the factors. The variation in plant size (L) has the greatest effect, causing a large increase in  $var_{SS}$ . When a population contains two plant sizes,

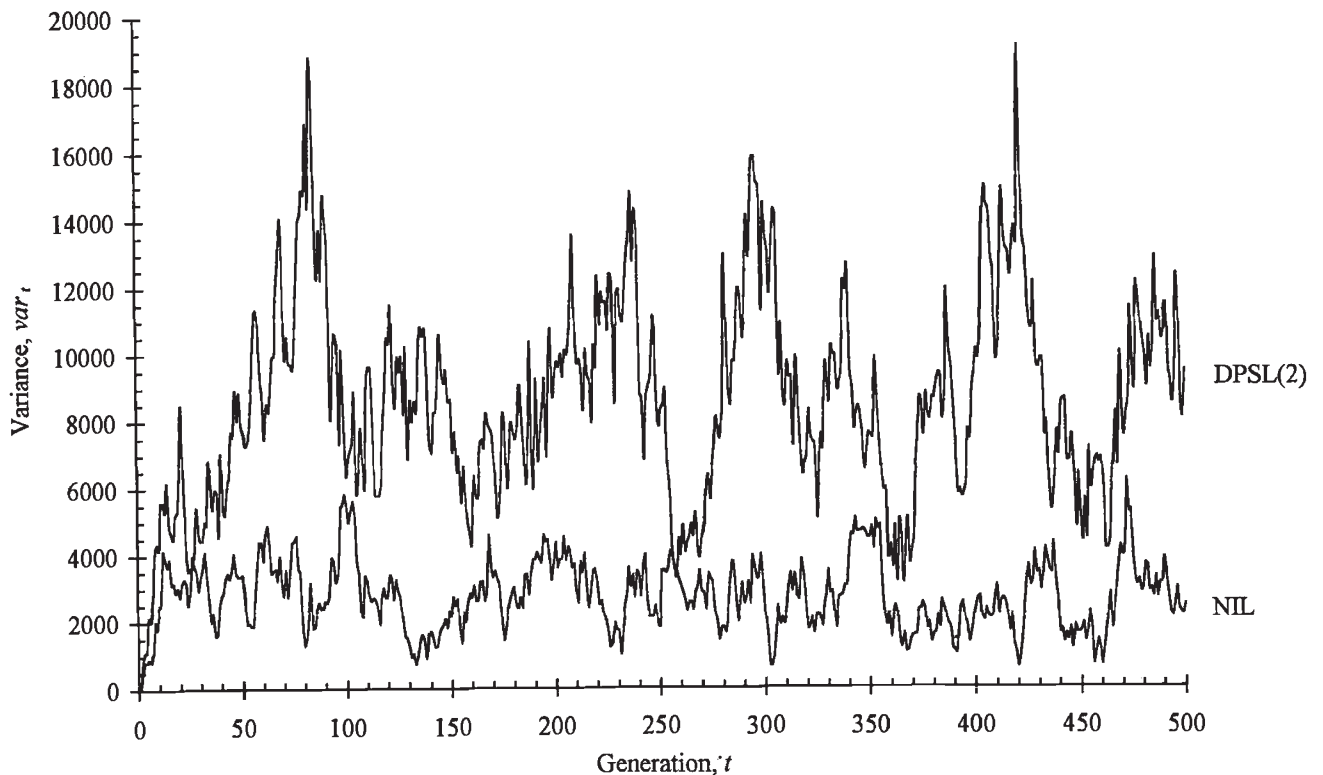


Fig. 2 Variance values for a single DPSL(2) population and for a single NIL population.

the probability of an allele being chosen in any one generation depends not only on its frequency in that generation, but also on the extent to which it occurs in large plants, which introduces an additional source of randomness; it is this that leads to a greater variation in allele frequencies and, hence, to an increase in  $var_{SS}$ . The effect of overlapping generations (D) is that the expected frequency of each allele in the new generation depends upon a weighted average of the allele frequencies in several previous generations. This averaging process will tend to reduce the extent of the variation in the allele frequencies and so reduce  $var_{SS}$ . These effects will be shown more clearly elsewhere in a theoretical analysis. This analysis will also consider the effect of changes in the parameter values for the D and L factors. The magnitude of the effect of the L factor indicates that further field measurements of this factor, particularly, would improve the accuracy of the models.

Although the effects of limited pollen (P) and seed (S) dispersal on  $var_{SS}$  are quite small, they introduce a spatial element to the system. Thus, in their absence, parents are chosen at random from any part of the population (although with an increased probability of selection for large plants). Hence, the selection of alleles is global, rather than local. On the other hand, when seed dispersal is limited, plants near to each other are more likely to be related, although this has no effect on the strength of frequency-dependent selection, which occurs only in the selection of the paternal parent. Limited pollen dispersal, however, results in the paternal parent being more likely to be near the maternal parent, which increases the local strength of frequency-dependent selection because the selection of alleles is now local, rather than global.

Thus, the presence of both P and S allows long-term local effects to develop because not only are the parents of each new plant likely to be close to each other, but the location of the new plant is also likely to be close to that of its parents. This could cause the alleles to be clustered and was investigated by adding a visual display to the model which showed the locations of a selected allele and which of these plants were large. The models with both or neither of the dispersal distributions were run and the location of the  $S_1$  allele in generations 200–210 recorded to investigate the effect of adding both limited seed and pollen dispersal. The displays for the NIL, D, L, DL, DPS and PS models appeared to be very similar with the allele scattered fairly evenly across the population area and, in particular, there were no large areas in which it was absent. There

appeared to be a greater correlation between areas of high and low allele density for successive generations in the PS model than in the other models. These results are as expected; the locations of the allele should be random for the models without P and S (because their selection does not then depend on their location) and the presence of the dispersal distributions means that the locations of the parents are likely to be close to the location of the progeny. The appearance of the PSL and DPSL models, on the other hand, was quite different, in that the populations had sizeable areas in which the allele was absent and other areas in which there were many occurrences of the allele, with the locations of these areas being similar in successive generations. This clustering of the allele was more pronounced in the PSL model than in the DPSL model. Figures 3 and 4 show visual displays of the spatial distribution of the  $S_1$  allele in the 200th generation of one of the runs of the NIL and one of the runs of the PSL model, respectively. Whereas this allele appears to be randomly distributed in the first of these displays, it is clearly not in the second.

Examination of this process shows that clustering of alleles is noticeable after only three generations of a PSL run. The effect of these factors is to create neighbourhoods of related individuals (Wright 1943, 1946), each of which contains a subset of the alleles of the population. Variation in plant size reduces the effective size of these neighbourhoods and so the presence of the L factor was necessary for the clustering to be clearly discernible. The strength of frequency-dependent selection will be stronger within these neighbourhoods than in the population at large and this will tend to decrease the variance of allele frequency. Random variation in the amount of this clustering, on the other hand, will tend to increase this variance. The P and S factors have, therefore, two opposite effects on the average steady-state variance of allele frequency which, as their combined effect in the populations simulated here has resulted in only a relatively small increase in this variance, appear to have approximately cancelled out. The number of alleles that can be maintained locally, however, depends on the effective size of a neighbourhood and not, provided this is not very small, on the number of alleles in the population. Hence, if the number of alleles in the population rises, clustering of these alleles could become more pronounced; i.e. the effect of limited pollen and seed dispersal on the steady-state variance of allele frequency could be greater in populations containing more alleles than our simulated populations.



Inspection of the visual displays of the PSL model suggested that the large areas in which  $S_1$  was absent appeared to be situated more often at the edges of the grid than elsewhere (Fig. 4). Edge effects are a well-known property of models involving the spatial

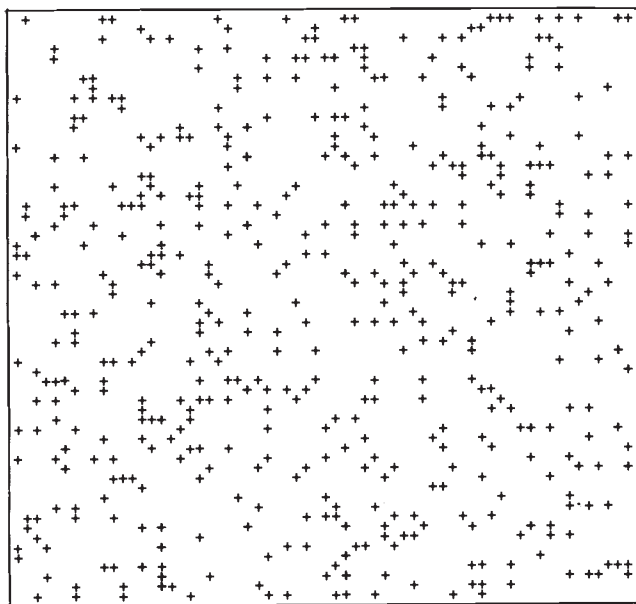


Fig. 3 Location of the  $S_1$  allele in the NIL model in the 200th generation of one of the simulation runs. The + shows the location of the allele.

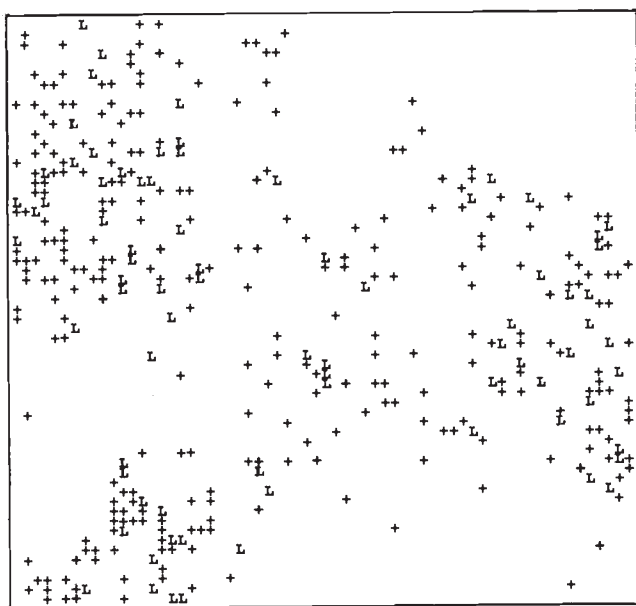


Fig. 4 Location of the  $S_1$  allele in the PSL model in the 200th generation of one of the simulation runs. The + shows the location of the allele on a small plant and L shows the location of the allele on a large plant.

distribution of units (Rohlf & Schnell, 1971; Levin & Wilson, 1978; Van Dijk, 1987). The size of the edge effect was examined in the usual way by joining the opposite edges of the grid to form a torus, thereby eliminating the edges (Maruyama 1972; Felsenstein 1975). This model, PSLWRAP was run 50 times and the PSL model a further 25 times to obtain more precise estimates of their steady-state variance of allele frequency. The PSLWRAP model gave a  $var_{SS}$  estimate of 10411 with a 95 per cent confidence interval of  $\pm 160$  and the revised PSL model, a  $var_{SS}$  estimate of 11521 with a 95 per cent confidence interval of  $\pm 444$ . The revised PSL estimate is similar to the previous one of 11683 (Table 1) with, as expected, a slightly smaller confidence interval. The PSLWRAP estimate is, however, noticeably lower and lies between the PSL and L estimates, which suggests that a considerable part of the combined effects of limited pollen and seed dispersal results from edge effects. On the other hand, a torus can hardly be regarded as a realistic model for a natural population because alleles occurring at one edge in one generation can reappear on the opposite edge in the following generation. Furthermore, although it is unlikely that limited seed dispersal would be subject to edge effects in natural populations, because this is a passive process, limited pollen dispersal almost certainly is, as the foraging activity of bees will distribute pollen only within the area occupied by a flowering stand of poppies. In practice, therefore, the true steady-state variance of the PSL model probably lies somewhere between the revised estimate and that of the PSLWRAP model.

## Conclusions

The average steady-state variance for the DPSL model, the most realistic of those investigated, is considerably greater than that of the idealized population simulated by Lawrence *et al.* (1994). Furthermore, a model with high average steady-state variance also has a high standard deviation of the steady-state variance values, which indicates that the variance for a particular generation could be considerably above the average value, even when the population is in steady state. It is, thus, possible that the unequal allele frequencies observed in samples taken from each of three natural populations of *P. rhoeas* (Campbell & Lawrence, 1981; Lawrence & O'Donnell, 1981) were, in fact, the result of purely stochastic effects operating in steady-state populations and it is hoped to carry out further simulations to investigate this.

## Acknowledgements

We thank Mr A. J. Girling for advice on statistical matters.

## References

- ARTHUR, A. E., GALE, J. S. AND LAWRENCE, M. J. 1973. Variation in wild populations of *Papaver dubium*. VII. Germination time. *Heredity*, **30**, 189–197.
- BLATTNER, F. AND KADEREIT, J. W. 1991. Patterns of seed dispersal in two species of *Papaver* L. under near natural conditions. *Flora*, **185**, 55–64.
- CAMPBELL, J. M. AND LAWRENCE, M. J. 1981. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. II. The number and frequency of S-alleles in a natural population (R106). *Heredity*, **46**, 81–90.
- FELSENSTEIN, J. 1975. A pain in the torus: some difficulties with models of isolation by distance. *Am. Nat.*, **109**, 359–368.
- HANDEL, S. N. 1983. Pollination ecology, plant population structure and gene flow. In: Real, L. (ed.) *Pollination Biology*, pp. 163–211. Academic Press, Orlando, FL.
- LAW, A. M. AND KELTON, W. D. 1991. *Simulation Modeling and Analysis*. McGraw-Hill, New York.
- LAWRENCE, M. J. AND O'DONNELL, S. 1981. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. III. The number and frequency of S-alleles in two further populations (R102 and R104). *Heredity*, **47**, 53–61.
- LAWRENCE, M. J., O'DONNELL, S., LANE, M. D. AND MARSHALL, D. F. 1994. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VIII. Sampling effects as a possible cause of unequal allele frequencies. *Heredity*, **72**, 345–352.
- LEVIN, D. A. AND KERSTER, H. W. 1974. Gene flow in seed plants. *Evol. Biol.*, **7**, 139–220.
- LEVIN, D. A. AND WILSON, J. B. 1978. The genetic implications of ecological adaptations in plants. In: Freyden, A. H. J. and Woldendorp, J. W. (eds) *Structuring and Functioning of Plant Populations*, pp. 75–98. North Holland, Amsterdam.
- MACKAY, I. J. 1981. *Population Genetics of Papaver dubium*. Ph.D. Thesis, University of Birmingham.
- MARUYAMA, T. 1972. Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. *Genetics*, **70**, 639–651.
- OKUBO, A. AND LEVIN, S. A. 1989. A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology*, **70**, 329–338.
- OOI, S. C. 1970. *Variation in Wild Populations of Papaver rhoeas*. Ph.D. Thesis, University of Birmingham.
- ROHLF, F. J. AND SCHELL, G. D. 1971. An investigation of the isolation by distance model. *Am. Nat.*, **105**, 295–324.
- VAN DIJK, H. 1987. A method for the estimation of gene flow parameters from a population structure caused by restricted gene flow and genetic drift. *Theor. Appl. Genet.*, **73**, 724–736.
- WELCH, P. D. 1983. The statistical analysis of simulation results. In: Lavenberg, S. S. (ed.) *Computer Performance Modelling Handbook*, pp. 268–328. Academic Press, New York.
- WRIGHT, S. 1939. The distribution of self-sterility alleles in populations. *Genetics*, **24**, 538–552.
- WRIGHT, S. 1943. Isolation by distance. *Genetics*, **28**, 114–138.
- WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.