

THE POPULATION GENETICS OF THE SELF-INCOMPATIBILITY POLYMORPHISM IN *PAPAVER RHOEAS*. II. THE NUMBER AND FREQUENCY OF *S*-ALLELES IN A NATURAL POPULATION (R106)

J. M. CAMPBELL and M. J. LAWRENCE

Department of Genetics, University of Birmingham, Birmingham, B15 2TT

Received 30.vi.80

SUMMARY

A randomly-drawn sample of 51 plants from a natural population (R106) carried between them 31 different *S*-alleles whose frequencies were, contrary to theoretical expectation, significantly unequal. Two alternative hypotheses are proposed to account for this unexpected observation. One of the eleven S_3 alleles found in the experiment appeared to be a stigma part mutant of the revertant type.

1. INTRODUCTION

IN a preliminary survey of the self-incompatibility polymorphism in *Papaver rhoeas*, Campbell and Lawrence (1981) found 32 different *S*-alleles among the members of eighteen full-sib families that were derived from three natural populations in the West Midlands. Though most of these alleles occurred in one or other of these populations only, an analysis of their distribution failed to indicate any significant differentiation between populations in this respect. However, no more than 24 alleles were sampled from any of these populations, so that the power of the test of differentiation could not have been very high. There was also a little evidence in these data that some *S*-alleles occurred more frequently than others among families that had originated from the same population, though because of the way in which these families were derived from their wild ancestors, such inequalities could have been artefacts of their origin alone. Be this as it may, these results suggest that it would be worthwhile investigating the number, frequency and distribution of *S*-alleles in samples that are considerably larger than those we have examined so far and in circumstances where any inequalities of frequency or distribution could be attributed to natural causes alone.

In the present paper, we consider the number and frequency of *S*-alleles in a sample of 51 plants taken from one of the populations investigated in the previous paper.

2. MATERIALS AND METHODS

The seed from which these plants were raised was obtained from a natural population consisting of several thousand individuals that was found growing in 1967 on a newly-formed roadside verge near Hackmans Gate (R106) in Worcestershire. Because the population, which ran alongside a

fence, was linear in shape, sampling was carried out in a partly systematic, partly random manner, so as to ensure that all parts of the population were visited. This was accomplished by removing open-pollinated capsules containing ripe-seed from a plant chosen at random at each of 68 successive sampling stations that were a fixed number of paces apart, the seed from each plant being placed and stored in a separate and sequentially numbered packet.

In 1977, three plants were raised from the seed of each of these 68 wild plants in a single randomised block experiment grown outdoors. Shortly after flowering commenced, one plant from each of these 68 half-sib families was chosen at random as a candidate for pollination, the other two being discarded. A random sample of 51 of these plants were then classified with respect to their incompatibility phenotype by means of a 51×51 half-diallel crossing scheme using the aniline-blue fluorescence method described in previous papers (Lawrence, 1975; Lawrence, Afzal and Kenrick, 1978; Campbell and Lawrence, loc. cit.). Most of the pollinations that were classified as half-compatible or incompatible in the first round of crosses were repeated in order to check the accuracy of this classification. In all, nearly 1500 pollinations were made during the six weeks that the 51 plants were in full flower. In addition, each plant was crossed reciprocally to another chosen at random in order to produce seed and thereby to preserve the *S*-alleles identified in this experiment for future investigation; and each was also self-pollinated to test its self-incompatibility.

3. RESULTS

The total number of crosses classified in the experiment was 1275, of which 2 (0.16 per cent) turned out to be incompatible, 216 (16.86 per cent) half-compatible and 1058 (82.98 per cent) fully compatible (See Campbell, 1977, for the results in full.) As a result of these crosses it was possible to assign to each plant in the experiment the genotypes shown in table 1.

The total number of *S*-alleles examined was $2 \times 51 = 102$ of which only 31 were different. With $n = 31$ and $m = 102$, the repeatability, $R = 0.72$ (Campbell and Lawrence, loc. cit.), which is second only to the repeatability obtained by Emerson (1939) with the *Oenothera* population. The most striking feature of these data, however, is that the frequencies of the 31 alleles are very unequal (table 2 and fig. 1). Intuitively, if the frequencies of the *S*-alleles in the natural population, from which this sample of 51 plants has been drawn, were approximately equal, we expect an occurrence distribution with a limited spread around a mode of about 3. We find instead that the modal frequency is 1 and that two of the alleles, S_3 and S_{17} occurred no less than eleven times apiece (fig. 1). It is thus most unlikely that the frequencies of the *S*-alleles in the R106 population are in fact equal.

The chief problem here is how to test the null hypothesis of equal allele frequency. If the *S*-alleles that have been found in this sample of 51 plants could be regarded as a random sample of the alleles of the R106 population, the conventional χ^2 goodness of fit procedure would be appropriate, apart, that is, from the rather low expected number in each of the 32 classes (3.29). However, since we have sampled *plants*, rather than pollen, from this population, the alleles have been sampled in pairs. In these circumstances,

TABLE 1

Genotypes of the 51 plants classified in the experiment. (Note that there is no known correspondence between the numbering of the alleles in the present experiment and those of the one discussed in the previous paper in this series.)

Plant No.	Genotype $S_i S_j$	Plant No.	Genotype $S_i S_j$
1	1, 2	37	26, 27
2	3, 4	39	3, 17
4	5, 6	40	16, 28
5	3, 7	41	27, 29
6	3, 8	42	24, 30
7	3, 9	43	7, 10
8	10, 11	44	1, 3
10	12, 13	45	7, 28
11	10, 14	46	5, 11
12	15, 16	47	11, 19
13	17, 18	48	19, 31
16	3, 14	50	10, 16
17	17, 19	51	1, 28
20	10, 17	52	3, 21
21	15, 20	53	17, 21
22	17, 21	54	3, 10
23	7, 22	55	20, 22
24	11, 23	58	7, 10
27	1, 15	59	25, 27
28	8, 17	62	11, 14
30	3, 24	63	14, 24
31	18, 25	64	4, 17
32	12, 14	65	17, 20
33	16, 26	66	3, 11
34	17, 27	68	17, 28
35	18, 27		

TABLE 2

The frequency of the occurrence of the 31 S-alleles in the R106 sample of 51 plants

S_i	Frequency	S_i	Frequency
1	4	17	11
2	1	18	3
3	11	19	3
4	2	20	3
5	2	21	3
6	1	22	2
7	5	23	1
8	2	24	3
9	1	25	2
10	7	26	2
11	6	27	5
12	2	28	4
13	1	29	1
14	5	30	1
15	3	31	1
16	4		

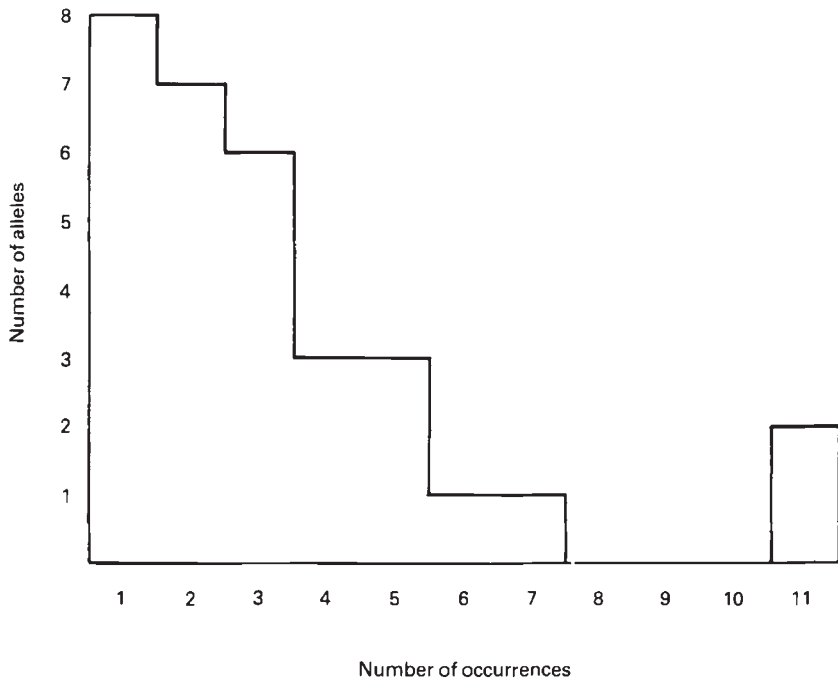


FIG. 1.—Distribution of *S*-alleles in the R106 population ($n = 31$), $m = 102$, $R = 0.72$, $\chi^2_{(30)} = 64.260$, $P = 0.000272$).

the occurrence of the second allele in each zygotic pair cannot be independent of that of the first, because all individuals in the population must be heterozygous for this gene. The conventional procedure cannot be used for testing the null hypothesis of equal allele frequency for this reason.

We are indebted to Dr Paul Davies for a solution to this problem which involves the calculation of a modified χ^2 (see appendix to this paper). In the present case $\chi^2_{(30)} \cong 64.620$ and $P = 0.00027$. We thus need have no hesitation in rejecting the null hypothesis that the allele frequencies in the R106 population are equal.

The number of alleles in the population from which this sample of 51 plants has been drawn, estimated by the conventional procedure (see Appendix) is $\hat{N} = 32.23 \pm 0.93$, which is very close to the number actually found. The equation from which this estimate has been obtained, however, is valid only when the alleles in the population are equally frequent, a hypothesis which we have tested and rejected. It is clear, therefore, that this estimate must be biased. Since those alleles which occur at a low frequency in the population are less likely to turn up in a sample than those which occur at a relatively high frequency, \hat{N} must be regarded, in these circumstances, as an estimate of the minimum number of alleles in the population.

There is one further and final point worth making about these results. The classification of pollinations with the aniline-blue fluorescence technique in this experiment presented few problems apart from one group of crosses all of which involved plant No. 54, which was eventually assigned the genotype S_3S_{10} . Crosses between this plant and others known to contain

S_{10} were half-compatible irrespective of the direction of the cross, as expected. Similarly, crosses between plant No. 54 and others known to contain S_3 were also half-compatible, provided that it was used as the male parent of the cross. When used as female parent, however, this plant appeared to be fully compatible with others containing S_3 . The simplest explanation of this unexpected outcome is that there was a weakening of the expression of S_3 in the stigmas of plant No. 54. Thus when this plant was self-pollinated, while one portion of the pollen exhibited the full incompatible reaction as expected, a second portion grew much longer tubes than normal. Self-pollinations of this plant, therefore, appeared to be half-compatible, the incompatible portion of the pollen containing, presumably, S_{10} and the apparently compatible portion, S_3 . Though the length of the tubes produced by S_3 pollen was less than those of compatible pollen, it was difficult to distinguish from the latter in half-compatible pollinations when plant No. 54 was used as female and it was for this reason that such crosses were initially classified as fully compatible.

4. DISCUSSION

The classification of a group of unrelated plants is, in general, a more hazardous task than the classification of the individuals of a full-sib family, because it is in the latter case possible to exploit the knowledge that every plant must fall into one or other of either two or four incompatibility classes. However, the high incidence of half-compatibility among the 1275 crosses made between the 51 plants of the R106 sample greatly reduces the probability of misclassification, because any plant carrying a particular S -allele must, of course, be half-compatible with *all* those which have been found earlier in the analysis to contain the same allele. Thus the advantage of a high incidence of repeats is not only that the number of S -alleles found in the sample must be near to the number present in the population, but also that the correspondingly high incidence of half-compatibility provides a network of check pollinations with respect to each plant in the sample. Furthermore, in the present case, each plant in the sample of 51 contains at least one S -allele that occurs in at least one other. For these reasons, confidence in the accuracy of our classification of these plants must be high.

Turning now to matters of more general interest, there are four points worth making about these results. Firstly, the unequal frequencies of the alleles found in this investigation cannot be attributed to our experimental procedure as was arguably the case in the previous experiment. Considerable care was taken to sample the whole population by a combination of systematic and random sampling and seed was not taken from closely adjacent plants. In addition, the 51 plants classified in this experiment were sampled at random from the total of 68 natural progenies that were raised in the experimental field for this purpose. For these reasons, the differences in allele frequency that we have observed must be due to natural causes alone.

Secondly, this is the first case of a population in which the frequencies of the S -alleles are significantly unequal. Indeed, there is only one other analysis of the polymorphism in the literature which is sufficiently thorough to yield any worthwhile information on this question and this is, of course, Emerson's (1939) on *Oenothera organensis*. Although all of Emerson's data cannot be used for this purpose, that part of it which can leaves little doubt

that the frequencies of the *S*-alleles in the *Oenothera* population approximate very closely to the expected equality (fig. 2). The present data, therefore, provide a particularly striking contrast to Emerson's findings.

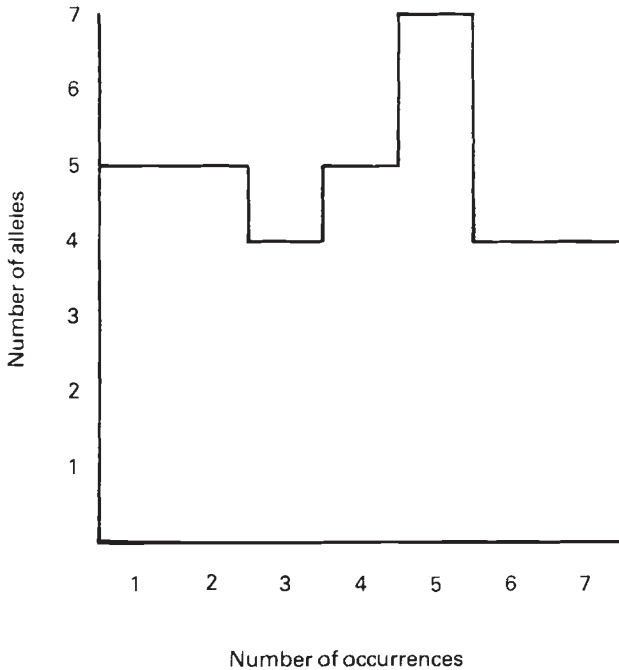


FIG. 2.—Distribution of *S*-alleles in the *Oenothera* population ($n=34$, $m=134$, $R=0.76$, $\chi^2_{(33)}=33.462$, $P=0.4448$). (Data from Table 2, page 532 of Emerson, 1939).

The third point worth making about our results concerns the cause of the unequal allele frequencies in the *P. rhoeas* population. Two alternative hypotheses can be advanced to account for this observation. The first hypothesis supposes that the alleles are subject to natural selection over and above that associated with the incompatibility effect due either to pleiotropic effects of the *S*-gene or to close linkage with other genes that affect fitness. Pollen carrying certain *S*-alleles, for example, may grow more quickly through compatible stigmas than pollen carrying other alleles. Clearly, if this were to be the case, the frequencies of the former in the population would be greater than those of the latter and neither would be equal to the frequencies expected on the assumption that selection is limited to the incompatibility effect.

The second hypothesis supposes that the grossly unequal allele frequencies are due to sampling effects of a kind particularly to be expected in a colonising species in which, by chance, a very small number of plants could make a disproportionate contribution of seed to the seed bank. On this hypothesis, the frequencies of the alleles are unequal because the population is in a state of disequilibrium, whereas on the first hypothesis, these frequencies are expected to be unequal in all circumstances.

Now both hypotheses predict that allele frequencies will be unequal in other populations also. However, while the first hypothesis leads to the prediction that the *same* alleles will occur at a relatively high frequency in different populations, the second predicts that *different* alleles will be common in different populations, those that are common being determined by chance alone. We thus have a means of distinguishing between these two hypotheses and experiments are in hand with this purpose in mind. In the meantime, it is worth pointing out that if, by chance, just one plant made a disproportionate contribution to the seed bank, both of the *S*-alleles carried by this plant would be expected to occur at a relatively high and similar frequency in the next generation. It may not be entirely a coincidence, therefore, that alleles S_3 and S_{17} each occurred eleven times in our sample of 51 plants.

The fourth and final point concerns the anomalous behaviour of S_3 pollen on stigmas of plant No. 54 which, we recall, was eventually assigned the genotype S_3S_{10} . Now if, as was suggested earlier, the excessive growth of S_3 pollen was due to a weakening of expression of S_3 in the stigmas of this plant, we expect to obtain more seed by the self-pollination of this plant than from any other. Usually, very little if any seed is obtained by self-pollinating *P. rhoeas* plants. The quantity of seed obtained by selfing plant no. 54, however, was appreciable, though not as great as that obtained by crossing a pair of compatible plants. Furthermore, if our explanation of the behaviour of S_3 pollen on the stigmas of plant No. 54 is correct, we expect to recover two classes of plants only in its self progeny. Thus if only S_3 pollen functions, half of this progeny should consist of plants whose genotype is S_3S_{10} and the other half of plants whose genotype is S_3S_3 . We are indebted to Miss Jo Moriarty, who examined 17 plants raised from seed obtained by selfing plant No. 54 in the season following the present experiment, for showing that this prediction is correct. Thus 10 of these plants fell into one class and the remaining 7 into another. Crosses between plants of the first and those of the second class were incompatible when the former were used as females, but half-compatible when the cross was made in reciprocal. Plants of the first class are thus heterozygotes, S_3S_{10} and those of the second, homozygotes. Crosses between the latter, used as males, and plants from other families known to be carrying S_{10} were fully compatible, whereas crosses between these homozygotes and other plants known to be carrying S_3 were incompatible when the latter were used as females. The genotype of individuals of the second class is thus S_3S_3 . Finally, the observed ratio of the numbers of S_3S_3 to S_3S_{10} plants in this family is in good agreement with the expected 1:1 ratio for $\chi^2_{(1)} = 0.529$ and $P = 0.5-0.3$.

The S_3 allele of plant No. 54 thus appears to be a naturally occurring and spontaneous stigma part mutant of a kind similar to the stylar part mutants described by Pandey (1956, 1967) in *Trifolium pratense* and *Nicotiana glauca*, respectively. Furthermore, since no difficulty was experienced with the analysis of the self progeny of this plant, this mutant must have been of the revertant type (Lewis, 1951), its function being fully restored in the following generation (though whether changes of this kind are properly referred to as mutants is open to question for they may be caused by the environment and are in any case not heritable). However, it is doubtful whether such mutants have anything but a very small effect on allele frequencies in the population both because they are not common and because it is likely that

pollen carrying these mutants will be successful in fertilisation only when compatible cross pollen is excluded from the stigmas of plants carrying the mutant.

Acknowledgments.—We are indebted to Mrs D. Edgerton for technical assistance and to Dr J. S. Gale for much useful discussion. One of us (J.M.C.) was in receipt of an S.R.C. Studentship whilst the work described in this paper was being carried out.

5. REFERENCES

- CAMPBELL, J. M. 1977. Population genetics of self-incompatibility in *Papaver rhoeas* L. Ph.D. thesis, University of Birmingham.
- CAMPBELL, J. M., AND LAWRENCE, M. J. 1981. The population genetics of the self-incompatibility polymorphisms in *Papaver rhoeas*. I. The number and distribution of S-alleles in families from three localities. *Heredity*, **46**, 68-79.
- EMERSON, S. 1939. A preliminary survey of the *Oenothera organensis* population. *Genetics*, **24**, 524-537.
- LAWRENCE, M. J. 1975. The genetics of self-incompatibility in *Papaver rhoeas*. *Proc. R. Soc. Lond. B.*, **188**, 275-285.
- LAWRENCE, M. J., AFZAL, M., AND KENRICK, J. 1978. The genetical control of self-incompatibility in *Papaver rhoeas*. *Heredity*, **40**, 239-253.
- LEWIS, D. 1951. Structure of the incompatibility gene. III. Types of spontaneous and induced mutation. *Heredity*, **5**, 399-414.
- PANDEY, K. K. 1956. Mutations of self-incompatibility alleles in *Trifolium pratense* and *T. repens*. *Genetics*, **41**, 327-343.
- PANDEY, K. K. 1967. Elements of the S-gene complex. II. Mutation and complementation at the S_I locus in *Nicotiana glauca*. *Heredity*, **22**, 255-238.

APPENDIX BY P. DAVIES. *Department of Mathematical Statistics, University of Birmingham B15 2TT*

A TEST OF THE EQUAL FREQUENCY HYPOTHESIS FOR S-ALLELES

Suppose that there are N different alleles in the population from which a sample of r plants is drawn at random. Suppose further that n different alleles are found in this sample. The data may then be represented as an array by writing 1 for the occurrence of an allele and 0 for its absence as follows:

		Allele				Row
Plant	S_1	S_2	S_3	S_n	Totals	
1	1	1	0	⋯	0	2
2	1	0	1	⋯	0	2
3	1	0	0	⋯	1	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
r	0	0	1	⋯	1	2
Column totals	C_1	C_2	C_3	C_n	$2r$	Grand total

The hypothesis of equal allele frequency is then equivalent to the hypothesis that any random permutation of the 1's in each row is equally

likely. For the case of unrestricted randomisation, this problem has been considered by Mantel (1974), who proposed the following statistic:

$$\chi^2 = (n-1) \left(\sum_{j=1}^n C_j^2 - 4r^2/n \right) / (2r - 4r/n)$$

as having approximately a chi-square distribution with $(n-1)$ degrees of freedom if the hypothesis is true and r is large.

However, as each observed allele is known to occur at least once it follows that the appropriate randomisation for the above problem would be conditional on each of the column totals, $C_1, C_2, C_3, \dots, C_n$, being greater than or equal to 1. Modification of the Mantel procedure conditional on this randomisation is a complex problem which we hope to make the subject of a further communication. However, it will be shown that in this instance the use of Mantel's statistic in its simple form provides a satisfactory though somewhat conservative test.

First, note that when the number of plants r is reasonably large relative to the number of population alleles N nearly all the alleles would occur in the sample at least once even supposing unrestricted randomisation. Further it can be shown that the expected number of different alleles observed in the sample is

$$E(n) = N(1 - (1 - 2/N)^r)$$

a result also obtained by Paxman (1963). This provides a method of estimating N by equating observed to expected:

$$n = \hat{N}(1 - (1 - 2/\hat{N})^r)$$

with $n = 31$ and $r = 51$, this gives $\hat{N} = 32.23$ which is close to n . As it is clear that for large r , \hat{N} and n are approximately equal then they should have roughly the same variance. However a better first order approximation obtained from the estimating equation above is

$$\text{Var}(n) \approx [1 - (1 - 2/N)^r + 2r(1 - 2/N)^{r-1}/N]^2 \text{Var}(\hat{N}).$$

Using the methods of Gittelsohn (1969) it can be shown that

$$\text{Var}(n) = N(N-1) \left[\frac{(N-2)(N-3)}{N(N-1)} \right]^r + N(1 - 2/N)^r - N^2(1 - 2/N)^{2r}.$$

From these two equations a standard error for \hat{N} can be obtained. When $r = 51$ and $\hat{N} = 32.23$, S.E. $(\hat{N}) = 0.93$ so it is unlikely that the true number of alleles differs grossly from the estimate \hat{N} . Though this may indicate that, on the assumption of equal allele frequency, nearly all the population alleles will be observed, the main justification for use of Mantel's statistic here is as follows. Clearly if it were possible for the C_j 's to take zero values they would vary more about their mean value $2r/N$ and hence produce higher values of Mantel's statistic than if restricted to be one or more. Consequently if Mantel's test rejects the equal frequency hypothesis then so would a correct test allowing the restriction on the C_j 's. The use of the test in this way should not therefore be misleading when the hypothesis is rejected but could have

low power in general. As shown below this is not a problem for the data obtained here.

	$n = 31, \quad r = 51$												
	$C_j =$	1	2	3	4	5	6	7	8	9	10	11	Total
Column frequencies		8	7	6	3	3	1	1	0	0	0	2	31

Hence $\sum C_j^2 = 540$

$$\begin{aligned}\chi^2 &= (n-1) \left(\sum_{j=1}^n C_j^2 - 4r^2/n \right) / (2r - 4r/n) \\ &= 30(540 - 102^2/31) / (102 - 204/31) = 64.260\end{aligned}$$

and

$$P(\chi_{(30)}^2 \geq 64.260) = 0.00027$$

Thus the equal frequency hypothesis would be rejected at a very low probability level.

REFERENCES

- GITTELSOHN, A. M. 1969. An occupancy problem. *Amer. Statistician*, **23**, 11-12.
 MANTEL, N. 1974. Approaches to a health research occupancy problem. *Biometrics*, **30**, 355-362.
 PAXMAN, G. J. 1963. The maximum likelihood estimation of the number of self-sterility alleles in a population. *Genetics*, **48**, 1029-1032.