

THE EXPECTED NUMBER OF SELF-INCOMPATIBILITY ALLELES IN FINITE PLANT POPULATIONS

SHOZO YOKOYAMA and LIANE E. HETHERINGTON

Department of Psychiatry, Washington University School of Medicine and Jewish Hospital of St. Louis, 216 South Kingshighway, St. Louis, Missouri 63110, U.S.A.; Department of Technology and Human Affairs, Washington University, St. Louis, Missouri 63130, U.S.A.

Received 10.viii.81

THERE have been many studies of the numbers of self-incompatibility alleles in different plant species (*e.g.*, Emerson, 1938, 1939; Bateman, 1947). In many species, self-incompatibility is controlled by a single multiple-allelic gene and pollen does not function on a style that carries the same allele. The numbers of these multiple alleles that can be maintained in a finite population have also been considered by theoretical population geneticists (Wright, 1939, 1960, 1964, 1969; Fisher, 1958; Ewens, 1964*a, b*; Ewens and Ewens, 1966; Kimura, 1965; Yokoyama and Nei, 1979).

Despite these extensive theoretical studies, the distribution of self-incompatibility alleles in a sample has not been studied fully. This sampling distribution is of importance because experimentalists can deal only with samples from a population and not the underlying population itself. In this note, we shall consider the expected number of alleles in a sample and the results obtained will be compared to the observations made by Emerson (1939) and by Campbell and Lawrence (1981) and Lawrence and O'Donnell (1981).

In the following, we assume that the population has achieved equilibrium under mutation, selection, and random genetic drift. We also consider an infinite allele model, that is, new mutations arising in a population are always different from the extant alleles (Wright, 1939; Fisher, 1958; Kimura and Crow, 1964). Let v and x_i be the mutation rate per generation and the frequency of the i th allele, respectively.

Before going into details of samples, we shall consider the number of alleles in the underlying population. There are two measures of the number of alleles: the effective number of alleles and the actual number of alleles. The effective number of alleles (n_e) is defined as the reciprocal of expected homozygosity $J = \sum_i x_i^2$:

$$n_e = 1/J \quad (1)$$

(Kimura and Crow, 1964). It should be noted that J is a mathematical device and, in practice, no homozygotes with respect to self-incompatibility will occur. The value of the homozygosity will be determined shortly. The actual number of alleles (n_a) is defined as

$$n_a = \int_{1/2N}^1 \Phi(x) dx, \quad (2)$$

where N is the effective population size and $\Phi(x) dx$ is the expected number of alleles whose frequencies are from x to $x + dx$ in an equilibrium population. The quantity $\Phi(x)$ is given by

$$\Phi(x) = 4Nve^{2Na_x} (1-2x)^{2Nb-1} x^{-1}, \quad 0 < x < 1/2, \quad (3)$$

where $a = 1/[(1-J)(1-2J)]$, $b = 1/[2(1-J)] + v$ (see formulae (29), (34), and (35) in Yokoyama and Nei, 1979). In this formulation, it has been assumed that no homozygote for self-incompatibility alleles is formed. Furthermore, J is treated as a constant (for its validity see Wright, 1939; Fisher, 1958; Yokoyama and Nei, 1979).

The expected homozygosity, J , can be obtained by using the relationship $\int_0^{1/2} x\Phi(x) dx = 1$. Hence, using (3), J is a solution of

$$v\sqrt{8\pi N} \exp(2NJ/[(1-J)(1-2J)]) = (1-J)^{-1/2} (1-2J)^{-N[1/(1-J)+2v]} \quad (4)$$

When $4NJ \gg 1$, equation (4) is simplified and

$$J = [-\ln(v\sqrt{8\pi N})/2N]^{1/2} \quad (5)$$

approximately.

Let us now consider a sampled population. In the sample, the expected number of alleles whose frequency lies in the range p to q ($0 \leq p < q \leq 1/2$) in a sample of m genes ($m/2$ individuals) drawn from a steady-state population is given by

$$n(p, q) = \sum_{i=[mp]+1}^{[mq]} \int_0^{1/2} \binom{m}{i} x^i (1-x)^{m-i} \Phi(x) dx \quad (6)$$

where $[n]$ is the largest integer contained in n (Chakraborty *et al.*, 1980). Considering the whole range of gene frequencies, when m alleles are sampled the expected number of alleles in the sample is given by

$$n_s = \int_{1/2N}^1 [1 - (1-x)^m] \Phi(x) dx. \quad (7)$$

Using these equations and assuming that $v = 10^{-5}$, some examples are shown in table 1. From the table, we can see that the numbers of alleles (n_e and n_a) increase with increasing values of N and that the number of alleles in the sample is affected strongly by sample size (see also Yokoyama and Nei, 1979).

In fig. 1, the distribution of expected number of alleles in the sample (histogram) and in the population (curve) are shown for $N = 2000, 3000,$

TABLE 1

Effective (n_e) and actual (n_a) numbers of alleles in the population and the number of alleles in a sample of m genes (n_s). It is assumed that $v = 10^{-5}$

N	n_e	n_a	$m = 50$	n_s $m = 100$
2000	27.39	31.06	24.23	28.89
3000	33.67	36.60	25.63	32.44
4000	39.06	43.71	28.07	37.23
6000	48.26	59.13	35.20	48.16

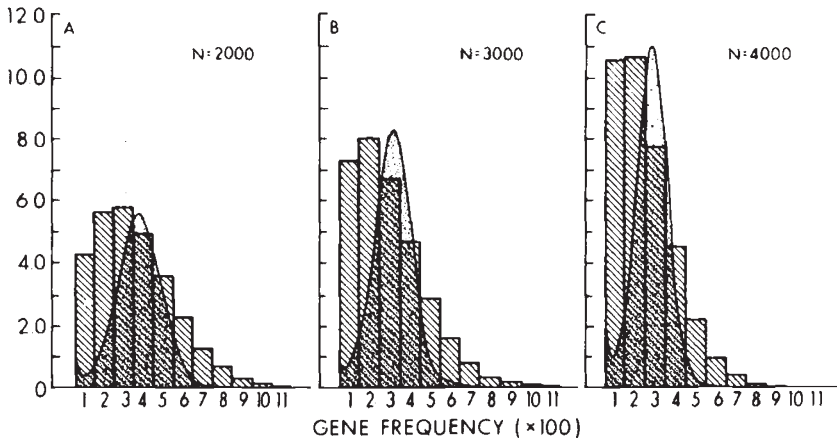


FIG. 1.—Gene frequency distribution of self-incompatibility alleles. The expected number of alleles in a sample and that in the population are shown by histogram and curve respectively. Three population sizes are considered: $N = 2000$ (A), $N = 3000$ (B), and $N = 4000$ (C).

and 4000, assuming that $v = 10^{-5}$ and $m = 100$. The expected number of alleles in the population is given by $\Phi(x)/200$. Both distributions show that as the effective population size increases the mode of the distribution shifts to the left, *i.e.*, the number of alleles increases. Clearly, the two distributions are very different.

Recently, Campbell and Lawrence (1981) made an extensive diallel cross using the Field Poppy, *Papaver rhoeas*. Sampling 51 plants at random from a natural population (R106) consisting of several thousand individuals, they found 31 different self-incompatibility alleles: eight alleles occurred once, seven alleles twice, six alleles three times, three alleles four times, three alleles five times, one allele six times, another allele seven times, and two alleles eleven times. Lawrence and O'Donnell (1981) have extended the analyses to two other natural populations (R102 and R104). For both populations, the numbers of genes sampled were 72. The observed number of alleles was 30 and 26 for the population R102 and R104, respectively.

Using equation (7), the expected number of alleles in a sample can be calculated easily. Assuming that $v = 10^{-5}$ and $m = 102$, the values of n_s are 28.97, 32.58, and 37.44 for $N = 2000$, 3000, and 4000, respectively. For $v = 10^{-5}$ and $m = 72$, the respective values are 27.12, 29.63, and 33.30.

It may be of interest to study the distribution of alleles for the cases $N = 2000$ (for R104) and $N = 3000$ (for R102 and R106), simply because the expected and observed values are very close. Such comparisons are shown in table 2. Table 2 shows that there are some differences between the observed and expected values; namely, that the three observed distributions have larger values in both tails than expected, especially for the right hand tails.

Several factors can account for this discrepancy. In the present theoretical model, it is assumed that all heterozygotes have exactly the same fitness. This may not be the case for the population and the alleles may be subject to selection either directly or via linkage with other genes, as Campbell and Lawrence (1981) note. The existence of partial incompatibility of a

TABLE 2

Distributions of observed and expected number of self-incompatibility alleles

Observations	Number of Occurrences											
	1	2	3	4	5	6	7	8	9	10	11	
<i>P. rhoeas</i>												
R102	12	9	5	0	1	0	2	1				
(<i>N</i> = 3000, <i>m</i> = 72)	9.95	8.64	5.64	3.04	1.42	0.59	0.22	0.13				
R104	9	7	3	3	1	1	0	1	1			
(<i>N</i> = 2000, <i>m</i> = 72)	6.74	7.24	5.83	3.77	2.04	0.94	0.38	0.14	0.04			
R106	8	7	6	3	3	1	1	0	0	0	2	
(<i>N</i> = 3000, <i>m</i> = 102)	7.06	7.94	6.74	4.75	2.94	1.65	0.84	0.39	0.17	0.06	0.04	
<i>O. organensis</i>												
(<i>N</i> = 2000, <i>m</i> = 74)	6.52	7.16	5.89	3.9	3.83							

certain allele may also contribute to the elevation of the allele frequency (see Campbell and Lawrence, 1981). Theory also assumes that the population is in equilibrium under mutation, selection, and random genetic drift. This assumption may not be valid for the Field Poppy populations. If that is the case, some alleles can be much more common than expected from an equilibrium theory. It should be noted that this deviation of gene frequency can occur with or without selection among heterozygotes. Also note that all three Field Poppy populations have the same characteristic: both rare and common alleles have larger values than expected. This uniformity among the populations also supports the notion that *P. rhoeas* may not be a colonizing species (Lawrence and O'Donnell, 1981).

Emerson's (1939) data on *Oenothera* was based on a mixture of half-sib families raised from seed taken from some wild plants and cuttings from others. It seems to be appropriate to use cuttings data only (Lawrence and O'Donnell, 1981). Then from Emerson's table 2, 28 different alleles can be detected among the 74 genes sampled: seven alleles (occurred once), six (twice), eight (three times), four (four times), and three (five times). Assuming that $v = 10^{-5}$ and $m = 74$, the values of n_s are 27.30, 29.90, and 33.66 for $N = 2000$, 3000, and 4000, respectively. These expected values are close to the observed value of 28, especially for $N = 2000$ (see table 2). In this case the equilibrium theory explains Emerson's data.

In conclusion, the present equilibrium theory can explain the *Oenothera* data of Emerson (1939), but it is not compatible with the Field Poppy data of Campbell and Lawrence (1981) and Lawrence and O'Donnell (1981). In the latter, differential selection may be operating among heterozygotes. It is also conceivable that the Field Poppy population has not yet attained its equilibrium state.

Acknowledgements.—The reviewer's comments are greatly appreciated. This work was supported by USPHS grants GM-28672 and MH-31302 with Washington University.

REFERENCES

- BATEMAN, A. J. 1947. Number of S-alleles in a population. *Nature*, 169, 337.
 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.

- CHAKRABORTY, R., FUERST, P. A., AND NEI, M. 1980. Statistical studies on protein polymorphism in natural populations. III. Distribution of allele frequencies and the number of alleles per locus. *Genetics*, *94*, 1038-1063.
- EMERSON, S. 1938. The genetics of self-incompatibility in *Oenothera organensis*. *Genetics*, *23*, 190-202.
- EMERSON, S. 1939. A preliminary survey of the *Oenothera organensis* population. *Genetics*, *24*, 524-537.
- EWENS, W. J. 1964a. On the problem of self-sterility alleles. *Genetics*, *50*, 1433-1438.
- EWENS, W. J. 1964b. The maintenance of alleles by mutation. *Genetics*, *50*, 891-898.
- EWENS, W. J., AND EWENS, P. M. 1966. The maintenance of alleles by mutation. Monte Carlo results for normal and self-sterility alleles. *Heredity*, *21*, 371-378.
- FISHER, R. A. 1958. *The Genetical Theory of Natural Selection*, Second Revised Edition. Dover, New York.
- KIMURA, M. 1965. Simulation studies on the number of self-sterility alleles maintained in a small population. *Ann. Rev. Natl. Inst. Genet. Japan*, *16*, 86-88.
- KIMURA, M., AND CROW, J. F. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, *49*, 725-728.
- 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 natural populations (R102 and R104). *Heredity*, *47*, 53-61.
- WRIGHT, S. 1939. The distribution of self-sterility alleles in populations. *Genetics*, *24*, 538-552.
- WRIGHT, S. 1960. On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in the population of a given size: A re-examination. *Biometrics*, *16*, 61-85.
- WRIGHT, S. 1964. The distribution of self-incompatibility alleles in populations. *Evolution*, *18*, 609-619.
- WRIGHT, S. 1969. *Evolution and the Genetics of Populations*. vol. 2. The University of Chicago Press, Chicago and London.
- YOKOYAMA, S., AND NEI, M. 1979. Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. *Genetics*, *91*, 609-626.