THE DETERMINISTIC BEHAVIOR OF SELF-INCOMPATIBILITY ALLELES[1]

THOMAS NAGYLAKI

Department of Medical Genetics, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received June 4, 1974

ABSTRACT

For a system of n self-incompatibility alleles, neglecting mutation and random drift, it is shown that the completely symmetric equilibrium is locally stable, and any allelic frequency less than $q = 1 + a - \sqrt{1 + a^2}$, where $a = [2(n-1)]^{-1}$, will increase. For all n, $q > (2n)^{-1}$, but if $n >> 1$, $q \approx (2n)^{-1}$.

I. INTRODUCTION

MANY flowering plants possess a self-incompatibility locus such that pollen will not function on a style carrying the allele in the pollen grain. The prediction of the expected number of alleles present in terms of the population size and the mutation rate is an interesting and difficult problem involving frequency-dependent multiallelic selection and random drift. The system was first treated mathematically by WRIGHT (1939), and then by FISHER (1958), whose approach differed from WRIGHT's only in relatively minor details (WRIGHT 1960). For highly instructive discussions of various conceptual questions and further analysis, the reader may refer to MORAN (1962), EWENS (1964, 1969), EWENS and EWENS (1966), and WRIGHT (1964, 1969). EWENS' (1969) derivation, employing the mean time a new allele remains in the population before loss due to random drift, is particularly lucid.

In order to effect a calculation, EWENS (1969), FISHER (1958), and WRIGHT (1969) focus attention on one of the alleles, and make various approximations which treat the others symmetrically. Thus, they obtain rather similar expressions for the expected value and variance of the change in the frequency of any one of the alleles. The approximations for the expected gene frequency change in one generation indicate that, as the self-incompatibility mechanism leads us to surmise, neglecting mutation and random drift, any allelic frequency less than some number equal or close to the reciprocal of the number of alleles should increase. If the mutation rates of self-incompatibility alleles are low, as they appear to be, then it is the selective advantage of rare alleles which allows a relatively large number of alleles to remain in fairly small populations in spite of random drift. Therefore, even if our main interest is in the maintenance of genetic variability in self-incompatibility populations, the theoretical justification

[1] Paper No. 1755 from the Genetics Laboratory, University of Wisconsin, Madison, Wisc. 53706. Supported by the National Institutes of Health (Grant GM-15422).

of the stochastic attack on this problem requires, as a prerequisite, the analysis of the deterministic behavior of the system. The remainder of this paper is devoted to that task.

Intuition and the approximations discussed above suggest that the only equilibrium of the self-incompatibility system is the totally symmetric one, and that this equilibrium is globally stable. For the special case of three alleles (the smallest number possible), this is easily proved: the deviations of all genotypic and allelic frequencies from their values in this equilibrium are multiplied by $-1/2$ every generation (see, e.g., MORAN 1962). The results presented below will add to the plausibility of the general conjecture. For four alleles, the range of gene frequencies not covered by the analysis was exhaustively studied numerically. The evaluation of the gene frequency change in more than 3000 genotypic configurations supports the hypothesis. In Section II we shall demonstrate that the symmetric equilibrium is locally stable, and display the rather striking dependence of the nature of the approach to equilibrium on the number of alleles. A sufficient condition for the increase of rare alleles will be derived in Section III by finding a lower bound on the strength of selection.

II. LOCAL STABILITY

We denote the frequencies of the n self-incompatibility alleles S_i by p_i, and those of the ordered genotypes S_iS_j by P_{ij} $(=P_{ji})$. Since S_i and S_j pollen will not fertilize an S_iS_j plant, there are no homozygotes: $P_{ii}=0$. The frequency of the unordered genotype S_iS_j is $2P_{ij}$. Employing a prime to signify the next generation, the recursion relations for the genotypic frequencies may be written for $i \neq j$ as (FISHER 1958)

$$2P'_{ij} = \sum_{k\neq i,j} \frac{P_{ik}p_j}{1-p_i-p_k} + \sum_{k\neq i,j} \frac{P_{kj}p_i}{1-p_k-p_j} . \tag{1}$$

Summing (1) over j, we find (FISHER 1958) that the gene frequencies satisfy

$$2p_i' = p_i + \sum_{j\neq i} \sum_{k\neq i,j} \frac{P_{kj}p_i}{1-p_k-p_j} . \tag{2}$$

It is trivial to verify that the symmetry point

$$\hat{P}_{ij} = [n(n-1)]^{-1}, \quad \hat{p}_i = 1/n \tag{3}$$

is a stationary point of (1) (MORAN 1962). To examine the local stability of this equilibrium, we write

$$P_{ij} = \hat{P}_{ij} + \varepsilon_{ij}, \quad p_i = \hat{p}_i + \eta_i , \tag{4}$$

with

$$\eta_i = \sum_{j\neq i} \varepsilon_{ij}, \quad \sum_i \eta_i = 0 . \tag{5}$$

Substitution of (3), (4), and (5) into (1) and linearization in ε_{ij} and η_i yields

$$\varepsilon'_{ij} = -\mu_n \varepsilon_{ij} + \left(\frac{\eta_i + \eta_j}{n-2}\right)\left(1 - \frac{1}{(n-2)(n-1)}\right) , \tag{6}$$

where

$$\mu_n = 1/(n-2) . \tag{7}$$

Summing (6) over j, we obtain

$$\eta_i' = \lambda_n \eta_i , \tag{8}$$

with the solution

$$\eta_i(t) = \lambda_n^t \eta_i(0) , \tag{9}$$

where

$$\lambda_n = 1 - \frac{n}{(n-2)(n-1)} . \tag{10}$$

Inserting (9) into (6) gives an elementary difference equation with the solution

$$\varepsilon_{ij}(t) = \mu_n[\eta_i(0) + \eta_j(0)]\lambda_n^t + \{\varepsilon_{ij}(0) - \mu_n[\eta_i(0) + \eta_j(0)]\}(-\mu_n)^t . \tag{11}$$

Equation (11) is the complete solution for the genotypic frequencies for small initial deviations from equilibrium: $|\varepsilon_{ij}(0)| << \dot{P}_{ij}$. As n increases from 3, λ_n increases from $-1/2$, never quite reaching $+1$. If $n = 3$, (5) and (7) show that the brace in (11) is zero. As n increases from 4, μ_n decreases from 1/2, never quite reaching 0. Therefore, (11) proves that $\varepsilon_{ij}(t) \to 0$ as $t \to \infty$, i.e., the equilibrium is locally stable.

The nature of the approach to equilibrium is rather interesting. If $n = 3$, (1) reduces to a simple linear system. Hence, the linearized solution $\varepsilon_{ij}(t) = \varepsilon_{ij}(0)(-1/2)^t$, derived at once from (11) by recalling (5), (7), and (10), must be exact, as is readily verified (see, e.g., MORAN 1962). Thus, for three alleles, the genotypic and allelic frequencies tend to equilibrium in an oscillatory manner, the deviations being multiplied by $-1/2$ every generation. For four alleles, $\mu_4 = 1/2$ and $\lambda_4 = 1/3$. Therefore, (9) shows that the gene frequencies approach equilibrium without oscillation at the rate $(1/3)^t$, while from (11) we conclude that (since $\mu_4 > \lambda_4$) eventually the genotypic frequencies approach the stationary point, oscillating every generation according to $(-1/2)^t$. For $n \geq 5$, $\lambda_n > \mu_n$, so that all deviations ultimately tend to zero at the rate λ_n^t without oscillation. In the biologically important case of many alleles, $n >> 1$, $\lambda_n \approx 1 - n^{-1}$, and the rate of approach to equilibrium is given by $\lambda_n^t \approx e^{-t/n}$, which is slower the larger the number of alleles.

III. THE INCREASE OF RARE ALLELES

We seek a lower bound on the strength of selection favoring rare alleles. Let us rearrange (2) to read

$$\Delta p_i = p_i X_i , \tag{12}$$

where

$$2X_i = -1 + \sum_{j \neq i} \sum_{k \neq i,j} P_{kj}(1 - p_k - p_j)^{-1} \quad . \tag{13}$$

Now,

$$(1 - p_k - p_j)^{-1} = \sum_{m=0}^{\infty} (p_k + p_j)^m > 1 + \sum_{m=1}^{\infty} (p_k^m + p_j^m) \ , \tag{14}$$

and, since the two terms in the sum in (14) contribute equally to (13), substituting (14) into (13) yields

$$2X_i > 1 + \sum_{k \neq i} \sum_{k \neq i,j} P_{kj}(1 + 2 \sum_{m=1}^{\infty} p_j^m) \ . \tag{15}$$

We sum over k,

$$\sum_{k \neq i,j} P_{kj} = p_j - P_{ij} \ , \tag{16}$$

and use the obvious inequality $P_{ij} \leq p_i$ when inserting (16) into the last term of (15) to obtain

$$2X_i > -1 + \sum_{j \neq i} [p_j - P_{ij} + 2(p_j - p_i)(p_j + \sum_{m=2}^{\infty} p_j^m)] \ . \tag{17}$$

Since the linear terms are easily summed over j, and the sum over m is trivial, after minor simplification we find

$$X_i > -p_i(2 - p_i) + (1 - p_i)Y_i \ , \tag{18}$$

where

$$Y_i = \sum_{j \neq i} p_j^2(1 - p_j)^{-1} \ . \tag{19}$$

We desire to find the minimum value of Y_i for fixed p_i and $0 < p_k \leq 1/2$ for $k = 1, 2, \ldots, n$. (The maximum gene frequency is $1/2$ and not 1 due to the absence of homozygotes.) Employing Lagrange multipliers, one can readily prove that Y_i has one and only one internal minimum, and that this minimum is at the point $p_j = (1 - p_i)(n - 1)^{-1}$, for all $j \neq i$, where

$$Y_i = (1 - p_i)^2(n - 2 + p_i)^{-1} \ . \tag{20}$$

If any one of the gene frequencies approaches zero, we must replace n by $n - 1$ in (20), so that the minimum value of Y_i is increased. If one of them is $1/2$, recalling (19), and replacing n by $n - 1$ and p_i by $p_i + (1/2)$ in (20), we deduce the local minimum

$$Y_i = (1/2) + [(1/2) - p_i]^2[n - (5/2) + p_i]^{-1} \ . \tag{21}$$

Routine algebra shows that (20) is less than (21) for $n \geq 3$, $0 < p_i \leq 1/2$. Consequently, (20) is a global minimum of Y_i, and, dropping the subscript i, we may rewrite (18) as

$$X > f(p) \ , \tag{22}$$

where

$$f(p) = -p(2 - p) + (1 - p)^3(n - 2 + p)^{-1} \ . \tag{23}$$

With moderate effort, it has not been possible to significantly improve the lower bound (23). From (23), we see directly that $f(0) > 0$ and $f(1/2) < 0$. Hence, $f(p) = 0$ will have at least one real root in the interval (0,1/2), and $f(p) > 0$ for $0 \leq p < q$, where q is the smallest root in (0,1/2). Setting (23) equal to zero, we obtain

$$q = 1 + a - \sqrt{1 + a^2} \ , \tag{24}$$

where

$$a = [2(n-1)]^{-1} \ . \tag{25}$$

Recalling (12), we conclude that any allelic frequency less than q will increase. For any n, it follows from (24) and (25) that $q > (2n)^{-1}$, but in the biologically important case $n >> 1$, $q \approx (2n)^{-1}$. For $n = 3,4,5$, the respective values of q are 0.219, 0.153, 0.117. The nature of our derivation makes it clear that the condition $p < q$ is sufficient for the increase of an allele, but by no means necessary. For four alleles, for example, the computer calculations referred to in Section I indicate that gene frequencies increase as long as they are less than 0.23. The naive expectation that allelic frequencies increase if they are less than $1/n$ and decrease otherwise, however, is false. With highly asymmetric genotypic configurations, in the four-allele case, examples were found in which a gene frequency decreased from 0.24 and increased from 0.26. This does not contradict the nonoscillating approach to equilibrium of the gene frequencies for four alleles deduced in Section II by linear local analysis because the latter was based on the assumption that the *genotypic* frequencies were close to equilibrium.

Our results enable us to comment on the approximations of WRIGHT (1964), FISHER (1958), and EWENS (1969). The expected number of alleles maintained in a finite population for a given mutation rate depends most strongly on the strength of selection for very small ($p << 1/n$) gene frequencies. Restricting ourselves to the most common situation in which there are many alleles, we can easily convince ourselves that the above authors' approximations reduce roughly to $\Delta p \approx p/n$ for $p << 1/n$. From (12), (22), and (23), we infer that this approximation is rigorously valid as a lower bound. Therefore, assuming diffusion approximations hold, the number of alleles maintained in a population for fixed mutation rate and population size should be at least as high as estimated by EWENS, by FISHER, and by WRIGHT. This is indeed fortunate, since, to the extent there are discrepancies between theory and observation (EWENS 1969; WRIGHT 1969), the number of alleles predicted appears to be too low.

I am happy to thank MR. JÜRGEN PÄTAU for the care and efficiency with which he carried out the computing reported in this paper.

LITERATURE CITED

EWENS, W. J., 1964 On the problem of self-sterility alleles. Genetics **50**: 1433–1438. ——, 1969 Pp. 71–76. In: *Population Genetics*. Methuen, London.

EWENS, W. J. and P. M. EWENS, 1966 The maintenance of alleles by mutation. Monte Carlo results for normal and self-sterility alleles. Heredity **21**: 371–378.

FISHER, R. A., 1958 Pp. 104–110. In: *The Genetical Theory of Natural Selection.* Dover Publications, New York.

MORAN, P. A. P., 1962 Pp. 159–163. In: *The Statistical Processes of Evolutionary Theory.* Clarendon Press, Oxford.

WRIGHT, S., 1939 The distribution of self-sterility alleles in populations. Genetics **24:** 538–552. ——, 1960 On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in a population of a given size: A re-examination. Biometrics **16:** 61–85. ——, 1964 The distribution of self-incompatibility alleles in populations. Evolution **18:** 609–619. ——, 1969 Pp. 143–148, 402–416. In: *Evolution and the Genetics of Populations.* Vol. 2. The University of Chicago Press, Chicago.

Corresponding editor: W. J. EWENS