ON THE PROBLEM OF SELF-STERILITY ALLELES

W. J. EWENS

Department of Statistics, Australian National University, Canberra, Australia

Received August 3, 1964

WRIGHT's initial analysis (1939) of the behaviour of a self-sterility population followed the discovery by EMERSON (1938, 1939) of a large number (at least 37, later stated by LEWIS [1948] to be 45) of self-incompatibility alleles in *Oenothera organensis*, a species restricted to a small mountain area in New Mexico. EMERSON found 154 individuals in this population and estimated that the total population size would probably not exceed 500. This is an extremely small population for such a large number of alleles, but in fact illustrates a common pattern; in general it appears to be typical for self-sterility populations to have a large number of different alleles even in a small population. The initial presumption is that the large number of alleles is due to a high rate of mutation to new alleles, and the problem considered by FISHER and WRIGHT is to find how much mutation is necessary to maintain a given number of alleles in such a population when the size of the population is fixed. The mutation rates indicated by their analyses are of the order $10^{-2.8}$ for the *Oenothera* population. Such a high theoretical mutation rate seems to be contradicted, however, by the failure of LEWIS (1948) to find, in a large-scale test, a single mutant in 220×10^6 cell divisions. WRIGHT (1960) then surmised that the large number of alleles is best explained by the hypothesis that the population size was recently several times as large as at present and that a partial isolation of colonies has tended to maintain a larger number of alleles than would be obtained by complete panmixia.

However, FISHER (1961a) has shown, by using a goodness-of-fit test, that differences in presence or absence of alleles in the various collecting grounds of *Oenothera* were not significant. He concluded that it is likely that cross-pollination does occur between the areas, or that if isolation does in fact exist, it has led to no differentiation between the subpopulations.

It seems reasonable that a more general argument is required. Since the existence of large numbers of alleles in small populations is common in self-sterility populations, any explanation of this phenomenon must not be peculiar to the *Oenothera* population, but should hold for all self-sterility populations and thus presumably be associated with the sterility mechanism itself.

FISHER (1961b) attempted to account for the apparent anomaly between LEWIS's result and the high theoretical mutation rate by putting forward the hypothesis that new alleles are synthesized by recombination, and that a new allele formed in this way would not be acceptable in a style secreting antibodies to both parent alleles. The introduction of new alleles would not then be revealed by the techniques of self-pollination. In this way an abundant number of new alleles could be derived without mutation being detected.

is important. (Numerical values suggest that the mean time approaches a value near 2.45^N).

It is well known for ordinary (i.e. non-self-sterility) populations admitting two alleles at any locus, that the mean time until one or other allele is lost by random sampling is at most of order N generations, when the population size is N . If the initial value of one or other allele is very small, this mean time is as low as $2 \log N$. These are completely different orders of magnitude than that obtained in (4) for the three-allele self-sterility population. Why there should be such a difference between the two is easily understood by referring to the deterministic behaviour of the infinite self-sterility population, used here as a description of the "expected" behaviour of the finite population. Whenever x_n approaches zero in the self-sterility population there is a very strong tendency for x_{n+1} to be a large value. For non-self-sterility populations there is no such tendency; for selectively neutral alleles the mean value of x_{n+1} is x_n , so that a gradual drift of x to zero is not unlikely.

More than three alleles. When more than three alleles are present, an exact analysis is much more difficult. However, it is easy to show that the two properties of the three allele process discussed above under *deterministic behaviour* will still hold, although to a lesser extent. There will be a pressure for x_n to approach a quasi-equilibrium value (which is the reciprocal of the number K of alleles currently present), but in a less rapid manner than in the three-allele process.

On the other hand, this tendency should not be allowed to obscure a further property of the process, which is that despite the pressure toward K^{-1} , random sampling will ensure that eventually the frequency of some allele will become zero. This allele is thenceforth lost from the population. In a similar manner, the remaining alleles are lost one by one until only two remain. At this point the population lasts for one generation and is then extinguished, since there will then be no pollen available which is acceptable to the individuals in the population.

Further, this gradual loss of alleles, followed by extinction of the population, will still take place, although at a far slower rate, if stochastic mutation is allowed to new or existing alleles. This is so because in any generation there is always positive probability that the number of alleles present is only two, and that no mutation occurs. Thus eventually this event will occur and the population automatically dies out. It is therefore meaningless, in a mathematical sense, to ask how much mutation will maintain a given number of alleles in equilibrium in these populations. On the other hand, it is likely that an enormously long time will pass before extinction takes place and that in the meantime a quasi-equilibrium distribution of alleles will develop. It would be a problem of some difficulty to derive the form of this quasi-equilibrium distribution. It is not allowable to use WRIGHT's well known stationary distribution formula, as the derivation of this formula makes certain assumptions on boundary conditions which are not met by the quasi-stationary distribution.

It is possible, however, to write down a formula for the mutation rate necessary to maintain in quasi-equilibrium a given number of alleles. That is, one can write down a formula for the mean of the quasi-stationary distribution, as this mean

will be the reciprocal of the mean number of alleles. We do this by balancing the mean number of alleles lost per generation with the mean number created. If the population size is N and the mutation rate (assumed to new alleles only) is ν per gene per generation, then the mean number of new alleles per generation is $2N\nu$. If there are k alleles being maintained in quasi-equilibrium, and if the mean number of generations that a given allele is maintained before being lost by drift and/or mutation is $\bar{T}$, then the mean number of alleles lost per generation is $k/\bar{T}$. Thus the balancing equation is $2N\nu = k/\bar{T}$

$$\text{or} \quad k = 2N\nu\bar{T} \quad (5)$$

In the case of selectively neutral, non-self-sterility populations, to a close approximation this formula become (EWENS 1964)

$$k = 4N\nu \int_{(2N)^{-1}}^1 x^{-1} (1-x)^{4N\nu-1} dx \quad (6)$$

If we simply insert $k = 45$, $N = 500$ in this formula we get approximately $\nu = 10^{-2.5}$, a result agreeing reasonably well with that of FISHER and WRIGHT. However (6) is not an applicable formula for self-sterility populations, since it was derived under assumptions not holding for such populations, so that the solution $\nu = 10^{-2.5}$ is valueless. The analysis of the preceding section makes it plausible that the mean time $\bar{T}$ is much higher for self-sterility populations than it is for ordinary populations, so that the required mutation rate, from equation (5), would be much smaller. If this is so, the mutation rate $\nu = 10^{-2.5}$ for the *Oenothera* population is much higher than the true value.

Unfortunately it appears to be very difficult to find a formula analogous to (6) for self-sterility populations, and numerical methods will probably be necessary for the determination of $\bar{T}$.

The analyses of Fisher and Wright. These are in most respects similar, and it is instructive to consider the methods and assumptions on which they are based for comparison with the previous arguments. It is sufficient for our purposes to consider the analysis of WRIGHT.

WRIGHT supposed that the population is of size N and denotes the sterility alleles $S_1, S_2, \dots, S_k$, with frequencies $q_1, q_2, \dots, q_k$. The method used is to consider the frequency of q of any particular allele (say S_1). He finds an expression for the mean change Δq of q from one generation to the next and also an expression for the variance $\sigma_{\Delta q}^2$ of this change. By substitution in his steady-state formula (1937)

$$\psi(q) = (C/\sigma_{\Delta q}^2) \exp [2 \int (\Delta q / \sigma_{\Delta q}^2) dq] \quad (7)$$

he obtains a supposed steady-state distribution of the frequency q of any allele. This is a continuous approximation to a discrete distribution, and since it breaks down when q equals zero, he restricts attention to "the probability distribution of alleles when present," i.e. to the distribution (7) taken at discrete points $(2N)^{-1}$, $2(2N)^{-1}$, $\dots$, and normalized to give unit total probability. From this WRIGHT derives the mean number n of alleles from the formula $n = \bar{q}^{-1}$, where $\bar{q}$ is the mean of this discrete distribution. Furthermore, bounds for the mutation rate ν

necessary to maintain this number of alleles are found by considering a formula that connects the probabilities that zero or one gene of the allele are present with the mutation rate.

FISHER also uses a formula equivalent to (7), (which is the diffusion approximation formula), and terminal frequencies to derive mutation rates.

MORAN (1962) has discussed WRIGHT's method, and has pointed out that if stationary distributions are to be considered, the probabilistic model used must ensure that the population will not die out; in the above case no such model has been specified. This point was referred to in the previous section. A stationary distribution formula used in a case where stationarity does not exist will give misleading results, even in a case such as that under consideration when a quasi-stationary behaviour is to be expected.

To this may be added the following observations. The "exact" value for $\sigma_{\Delta q}^2$ given by FISHER (1958) and WRIGHT (1960) in equation (7) was $q(1-2q)/2N$. This must be incorrect since it implies that when $q = 1/2$ (all individuals contain one S_1 allele) the number of S_1 alleles in the next generation is determinate. However it is clear that in fact this number is a binomial variate with index N and parameter $1/2$. Derivation of the exact formula for the variance will be complicated by the peculiar reproductive system of the population. For three alleles, it is not hard to show that $\sigma_{\Delta q}^2 = q(1-q)/4N$. Clearly the usual binomial formula for variance is inappropriate.

Secondly, and more importantly, the Fokker-Planck diffusion approximation, from which (7) is essentially derived, cannot be applied for self-sterility populations. This is so because the diffusion approximation holds only under the assumption that $\overline{\Delta q}$ and $\sigma_{\Delta q}^2$ are of the same order of magnitude in N^{-1} . WRIGHT's expression for $\overline{\Delta q}$ was

$$\overline{\Delta q} = -q(q-\alpha)/(1-\alpha)(1-2\alpha) \quad (8)$$

where α is the equilibrium frequency. His expression for $\sigma_{\Delta q}^2$ is given above, and is of order N^{-1} , while (8) is not. Thus any result obtained by inserting these values in (7) may lead to a very misleading result. For example, we consider the use of diffusion methods for the three-allele system. Diffusion methods assume implicitly that terms like $(\overline{\Delta q})^2$ are small compared with $\sigma_{\Delta q}^2$ and may be ignored in comparison with the latter. The value $q = (2N)^{-1}$ is particularly important since this frequency is used by FISHER and WRIGHT when finding mutation rates. For a population of 500 individuals and $q = (2N)^{-1}$, $(\overline{\Delta q})^2$ is about $1/4$, while $\sigma_{\Delta q}^2$ is about 10^{-6} . However it is the former value which is implicitly ignored in comparison with the latter. For the case of more than three alleles the error will not be so extreme but will still be significant. It is because diffusion methods cannot be used that equation (5) cannot be replaced by (6) for self-sterility populations. It is, in fact, believed by the present author that use of diffusion methods has led to very inaccurate results for self-sterility populations.

Conclusion. The reason suggested here for the occurrence of large numbers of alleles in self-sterility populations is not that there exists a high mutation rate to

new alleles, but that either (a) owing to the very long time before an allele is lost by mutation and/or drift, even a very small mutation rate will maintain in quasi-equilibrium a large number of alleles for a long time, or (b) if initially a large number of alleles existed, then even without mutation the alleles will be lost at a very slow rate. Thus any population which is observed may well still have a large number of alleles. For example, even if there were only three alleles initially in an annual population of 500, the mean time (about 2.45^{500} years) until the population dies out is far longer than the supposed age of the earth. For ordinary (i.e. non-self-sterility populations), the time needed to lose one of three alleles is of the order 500 years.

Neither of these alternatives is contradicted by the experimental results of LEWIS (presuming that the hypotheses of WRIGHT and FISHER referred to in the introduction are not valid). However, the alternatives given above should only be regarded as tentative until numerical results and more complete mathematical methods have been obtained.

SUMMARY

The problem of explaining the presence of large numbers of different alleles in small self-sterility populations is considered. Previous treatments have attempted to explain the large numbers by assuming a high mutation rate or by a subdivision of the population into many smaller sub-populations. The present analysis suggests that a different explanation is possible, namely an extremely slow rate of loss of alleles by random sampling for self-sterility populations. This slow rate is due to the tendency for the frequencies of rare alleles to increase towards an equilibrium value. No such tendency occurs in normal populations.

LITERATURE CITED

- EMERSON, S., 1938 The genetics of self-incompatibility in *Oenothera organensis*. *Genetics* **23**: 190-202. — 1939 A preliminary survey of the *Oenothera organensis* population. *Genetics* **24**: 524-537.
- EWENS, W. J., 1964 The maintenance of alleles by mutation. *Genetics* **50**: 891-898.
- FISHER, R. A., 1930, 1958 *The Genetical Theory of Natural Selection*. The Clarendon Press, Oxford. 2nd Revised Edition, Dover, New York. — 1961a Possible differentiation in the wild population of *O. organensis*. *Austral. J. Biol. Sci.* **14**: 76-78. — 1961b A model for the generation of self-sterility alleles. *J. Theoret. Biol.* **1**: 411-414.
- LEWIS, D., 1948 Structure of the incompatibility gene. 1. Spontaneous mutation rate. *Heredity* **2**: 219-236.
- MORAN, P. A. P., 1962 *The Statistical Processes of Evolutionary Theory*. Oxford University Press, London.
- WRIGHT, S., 1937 The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. U.S.* **23**: 307-320. — 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 given size: a re-examination. *Biometrics* **16**: 61-85.