

# GENE FLOW AND GENETIC DIFFERENTIATION

P. T. SPIETH

*Department of Genetics, University of California, Berkeley, California 94720*

Manuscript received January 10, 1974

## ABSTRACT

A brief analysis is presented for the effects of gene flow upon genetic differentiation within and between populations generated by mutation and drift. Previous results obtained with the "island" model are developed into a form that lends itself to biological interpretation. Attention is focused upon the effective local population size and the ratio of the genetic identity of two genes in different populations to that of two genes in the same population. The biological significance of this ratio, which is independent of population size, is discussed. Similarities between the results of this model and those of the "stepping-stone" model are noted.

THE basic unit of evolutionary genetics is the rather poorly defined grouping of actively interbreeding individuals known as a population. The gene pool of a local population is the arena in which all micro-evolutionary changes must transpire. However, the potential for gene flow among all populations of the same species blurs the distinctness of local populations and makes for a genetic interdependence such that micro-evolutionary changes in any one may be transmitted to the others. It is not obvious as to what extent the gene pool of an entire species will evolve as a unit and to what extent the gene pools of local populations will evolve in separate directions. Undoubtedly, local selective pressures are the principal determinant of the extent to which a population will be influenced by changes in other populations; however, the different possibilities for selection generate such a broad spectrum of possibilities for differentiation between two gene pools that the population biologist is faced with difficulty in knowing what to expect or how to interpret observed situations. A reference situation, with respect to which expectations and observations can be compared, may be obtained by analyzing the amount of differentiation to be expected when selection is not taken into consideration. That is, consideration of traits (whether they exist or not) for which selection is neither promoting nor inhibiting genetic differentiation provides a theoretical framework upon which the effects of any particular selection regime may be superimposed in either a rigorous or an intuitive manner as the case permits.

In practice, this is accomplished through the analysis of the differentiation which can be generated by the processes of mutation, drift and migration. Such analyses date back to the classic work of WRIGHT (1931 and 1943); however, in recent years they have been the object of a considerable amount of work. MARUYAMA has published a series of papers which analyze a variety of "step-

ping-stone" models of migration (for an apropos summary see KIMURA and MARUYAMA 1971); MAYNARD SMITH (1970) investigated the topic using the "island" model of WRIGHT; CROW and MARUYAMA (1971) and LATTER (1973) have obtained results of remarkable theoretical generality. This note develops the results of MAYNARD SMITH (1970) into a form that is particularly amenable both to biological interpretation and for noting rather striking similarities with results obtained from other models.

The model utilized by MAYNARD SMITH (1970) presupposes that a sexual, diploid species is divided into  $r + 1$  populations, each of effective size  $N_i$ . For a given locus,  $u$  is rate of selectively neutral mutation per generation, with each mutant assumed to be unique. Finally, each generation an individual has the probability  $m$  of breeding in a population other than the population of his birth; so  $m$  is a measure of the total amount of migration out of and into each population per generation. The term "island model" denotes the assumption that an emigrant from one population is equally likely to immigrate into any one of the  $r$  other populations. Hence,  $m/r$  measures the unidirectional migration between any pair of populations per generation.

The basic results are obtained through the following two quantities: Let  $F$  denote the probability that an individual's genes at a given locus are identical by descent. Let  $S$  denote the probability that two homologous genes randomly drawn from two different populations are identical.  $F$  measures the genetic homogeneity of a local population while  $S$  measures the genetic similarity between two different populations. By obtaining recursion formulae for  $F$  and  $S$  MAYNARD SMITH solved for the equilibrium condition in which neither of these two variables changes from one generation to the next. The assumption was made that  $m$  and  $u$  are both sufficiently small that terms involving  $m^2$ ,  $u^2$  or  $mu$  can be ignored. His basic algebraic results are given by  $F = \{1 + 4N_i m S\} / \{1 + 4N_i (m + u)\}$  and  $S = mF / (m + ur)$ .

He investigated the behavior of  $S$  and  $F$  by considering three cases, depending upon whether  $ur \gg m$ ,  $m \gg ur$ , or  $m \sim ur$ , respectively. He argued that, with one exception,  $S$  will be close to zero in both the first and third cases. However, it can be seen that  $S$  will be near zero if  $F$  is near zero; hence, simply investigating the value of  $S$  does not provide an adequate measure of the resemblance between two populations. A preferable measure is the ratio of  $S$  to  $F$ , which indicates the genetic similarity of two different populations *relative* to the extent of differentiation within a single population. This ratio is conveniently expressed in terms of the quantity  $k$  which is defined by  $ku = m/r$ . (In words,  $k$  is the ratio of the rate of migration between any pair of populations to the mutation rate.) Then,  $S/F = k/(k + 1)$ . This expression, introduced by MAYNARD SMITH, shows the remarkable fact that the ratio of  $S$  to  $F$  depends solely upon the mutation and migration rates (as expressed by  $k$ ) and is *independent of population size*. This fact has also been observed by LATTER (1973), who introduced the quantity  $\gamma$ , which is equal to  $1 - S/F$ . Additionally, the above expression for  $S/F$  indicates that there will be substantial differentiation among populations only if  $k \ll 1$ , which is very similar to the situation with a two-dimensional stepping-stone model for which

equation (8) of KIMURA and MARUYAMA (1971) obtains. However, if the total size of the ensemble,  $N_T = (r + 1)N_i$ , is sufficiently small or mutation so rare that  $N_T u \ll 1$ , their results show that differentiation among populations is possible under the stepping-stone model even though  $k > 1$ .

For a given  $N_i$ ,  $r$  and  $u$  the effect of migration upon genetic differentiation can be illustrated by its effects upon  $F$  and  $k$ . MAYNARD SMITH'S equation for  $F$  can be put into the form:

$$1/F = 1 + 4N_i u \left\{ 1 + r \left( \frac{k}{k+1} \right) \right\}. \quad (1)$$

Since the principal effect of immigration is to increase the size of the local gene pool, the most biologically intuitive approach is in terms of the effective size,  $N_e$ , of a local population. For a given  $F$ , this is defined by  $1/F = 1 + 4N_e u$  (KIMURA and CROW 1964). (A clear distinction should be made between  $N_i$  and  $N_e$ , both of which are "effective" population sizes. The former is the size of the local population adjusted to take into account all such *internal* considerations as unequal sex ratio, variance in offspring number and overlapping generations. It does not reflect any influences on the effective size resulting from other populations.  $N_e$ , in addition to reflecting the internal considerations, adjusts the effective size to account for migration.) We have, then, the simple relation  $N_e = YN_i$ , where  $Y$  is the term in brackets in (1). With increasing migration  $k$  varies from zero (no migration) to being considerably greater than unity. Hence, the effective local population size varies continuously from  $N_i$ , its effective size in total isolation, to  $N_T$ , the size of the total ensemble of populations. A plot of  $Y$  versus  $k$  will show the manner with which  $N_e$  increases with increasing migration. It is clear that for  $k > 1$ ,  $N_e$  is of the order of  $N_T$ .

Additional insight into the processes of genetic differentiation may be gained by conceptually reversing the roles of  $k$  and  $F$ . The above results can be reworked to give the following weak inequality on  $k$ :

$$B \left( \frac{r+1}{r} \right) - 1 < k < B \left( \frac{r+1}{r} \right), \quad (2)$$

where  $B = 4N_i m / \{ (1/F) - 1 \}$ . This expression provides an indication of the expected amount of genetic differentiation between populations (as measured by  $k$ ) for given values of  $F$  and the *number* of migrants,  $N_i m$ . In particular, if each population averages one migrant per generation ( $N_i m = 1$ ) then (2) indicates that  $k$  will be on the order of  $4/(1/F - 1)$  and will be greater than unity unless  $1/F$  is at least three and preferably five or more. That is,  $S$  will be similar to  $F$  except for the case of an extremely high mutation rate (say,  $N_T u \gg 1$ ) for which it is theoretically conceivable that  $S \ll F \ll 1$ . Note, however, that for this exception to obtain  $F$  must be considerably less than unity, which represents a level of intrapopulation genetic differentiation considerably greater than has been observed to date even though the currently utilized electrophoretic techniques provide underestimates. The biological interpretation is that at this rate of migration there will be little differentiation between populations

unless mutation is so large as to produce substantial differentiation within the individual populations. To an extent this point merely confirms previous results from island models (e.g., MORAN 1962, chapter 9). However, the full implications of (2) are remarkably similar to those of a two-dimensional stepping-stone model (KIMURA and MARYUYAMA 1971)—even to the qualification concerning the effect of an extremely high mutation rate—which suggests a greater robustness than is normally credited to island and stepping-stone models.

Two notable observations stem out of the preceding discussion. First is the observation that the ratio of  $S$  to  $F$  depends solely upon  $k$ . Biologically this implies that, while the actual amounts of differentiation,  $F$  and  $S$ , depend upon the mutation rate and the effective local population size,  $N_e$  (which is influenced by the migration rate), the matter of whether or not an ensemble of populations drifts together as a group depends solely upon whether or not migration is sufficiently strong to distribute mutants throughout the group as fast as mutation creates new ones. To be precise, it follows from the definition of  $k$  that  $k = \{(r+1)/r\} \{N_1 m / N_T u\}$ . Hence, the critical factor is the ratio of the number of migrants per local population per generation to the total number of new mutants in the entire ensemble each generation.

This point is of especial theoretical interest when considered in light of CROW and MARUYAMA's (1971) result. Instead of  $S$  they used the variable  $f$ , which is defined as the probability that two homologous genes, randomly drawn from the *entire ensemble* of populations, are identical. They obtain the very general result that  $(1 - F)/f = 4N_T u$  regardless of the rate and pattern of migration! It would seem, therefore, that the use of  $f$ , which is a weighted average of  $F$  and  $S$ , tends to confound and obscure the dynamic relationship between  $F$  and  $S$ . Unfortunately, the expression given here for  $S/F$  is only strictly valid for the island model. Whether or not its independence from  $N_e$  is applicable to other migration patterns is open to conjecture. In any event, a clear distinction among genetic differentiation within a population, genetic differentiation between populations and the average genetic differentiation of a species (or ensemble of populations) is essential for maintaining proper perspective.

The second point is the observation that migration at the rate of one migrant individual per local population per generation is, generally, sufficient to obscure any disruptive effects of drift. (It should be emphasized that this is merely a "rule of thumb" and, as was brought out in the discussion following (2), is not without qualification.) The essential feature is the extremely low level of migration, from which several significant biological implications follow. First, in relation to suggestions such as those of EHRlich and RAVEN (1969) that migration between natural populations may be extremely low, it shows that such a rate may be, nonetheless, sufficient to cause the populations to evolve together unless selection is actively operating to drive them apart. Second, the existence of genetic differentiation between populations in the presence of gene flow of even a low presumptive level is suggestive of differing selection pressures. Mere intrapopulation drift is ruled out as the cause of the differentiation. However, population structure—the subdivision of a species into local, interbreeding units—is

susceptible to being organized and reorganized over comparatively short, ecological time lengths. Consequently, genetic differentiation between local groups may possibly be a reflection of a non-equilibrium situation attributable to recent historical shifts in population structure rather than different selective regimes. Perhaps of greater interest is the converse situation in which there is a lack of significant genetic differentiation between populations. The very low critical level of migration means that it is impossible to draw a direct conclusion as to whether migration and selection are working in the same direction or whether migration is operating alone. Third, this point emphasizes the great significance of the *potential* ability of conspecifics for interbreeding and demonstrates the even greater significance of speciation whereby such ability is lost. In terms of gene flow, the distinction between absolutely none and almost none is enormous.

## LITERATURE CITED

- CROW, J. F. and T. MARUYAMA, 1971 The number of neutral alleles maintained in a finite, geographically structured population. *Theoret. Pop. Biol.* **2**: 437-453.
- EHRlich, P. R. and P. H. RAVEN, 1969 Differentiation of populations. *Science* **165**: 1228-1232.
- KIMURA, M. and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725-739.
- KIMURA, M. and T. MARUYAMA, 1971 Pattern of neutral polymorphism in a geographically structured population. *Genet. Res.* **18**: 125-131.
- LATTER, B. D. H., 1973 The island model of population differentiation: a general solution. *Genetics* **73**: 147-157.
- MAYNARD SMITH, J., 1970 Population size, polymorphism, and the rate of non-Darwinian evolution. *Am. Naturalist* **104**: 231-237.
- MORAN, P. A. P., 1962 *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97-159. —, 1943 Isolation by distance. *Genetics* **28**: 114-138.

Corresponding editor: W. EWENS