

# THE THEORY OF SOME GENETICAL EFFECTS OF POPULATION SUBDIVISION

By P. A. P. MORAN\*

[Manuscript received September 26, 1958]

## Summary

The genetical effects of the subdivision of a population into partially isolated subgroups are considered in two particular cases. In the first a probability model is studied in which the subpopulations are of finite size with migration between them. In the absence of selection the asymptotic rate of progress to homozygosity is shown to be very little affected by the subdivision. In the second case a deterministic model is studied in which there are two subpopulations in which selective forces are equal and opposite. A stable dimorphism is then shown to exist if there is any small amount of intermigration.

## I. INTRODUCTION

The evolutionary implications of the subdivision of a population into groups which are partly isolated from one another has long been a subject of controversy. Wright and Fisher have shown that in a completely random mating population without selection or mutation the probability of the population remaining heterozygous with respect to two alleles at a single locus is proportional to  $\lambda^t$ , where  $t$  is measured in generations, and  $\lambda = 1 - N(8N_1 N_2)^{-1}$ . Here  $N_1$  and  $N_2$  are the numbers of males and females,  $N = N_1 + N_2$ , the generations are non-overlapping, and the distribution of the number of offspring per parent is approximately Poissonian. The theory of this phenomenon is simpler if we consider a population of  $N$  monoecious individuals. If these are diploid the population is completely defined by the nature of the  $2N$  gametes from which they were formed. The probability of the population remaining heterozygous is then asymptotically proportional to  $\lambda^t$  where  $\lambda = 1 - (2N)^{-1}$ .

It is known that inbreeding resulting from assortative mating or consanguinity can alter this rate considerably. If now we consider a population spread out on a wide geographical area, the conditions of strict random mating can no longer hold and this will be still more evident in a population divided into subgroups between which only a limited amount of migration takes place. It is therefore of some interest to study mathematical models of such subdivided populations.

We could do this for populations of diploid individuals with sexual differentiation but the algebra is then so complicated that we confine ourselves to monoecious populations and thus it is only necessary to consider populations of haploid individuals. This will give a quite adequate idea of the effects of migrations between subpopulations in more complicated models.

\* Department of Statistics, Australian National University, Canberra.

## II. RANDOM APPROACH TO HOMOZYGOSITY IN A SUBDIVIDED POPULATION

Suppose then that we have  $H+1$  subpopulations each of  $N$  haploid individuals. The total population thus consists of  $(H+1)N$  individuals and if there were no subdivisions the probability of heterozygosity would decrease asymptotically as  $\{1-[(H+1)N]^{-1}\}^t$ . We suppose, however, that in each subpopulation the next generation contains  $K$  migrants chosen at random from the offspring of each of the other  $H$  populations, together with  $N-HK$  offspring from the original subpopulation itself. We write  $L = HK$  for the total number of migrants into each subpopulation. Then  $L < N$ . Each of these migrants is a haploid individual.

In this model we have assumed that migrants come equally from all the other subpopulations. A more realistic model would be obtained if we assume that the subpopulations are distributed geographically over a two-dimensional area, for example if they were concentrated at the vertices of a rectangular lattice. Migration would then be permitted only from the nearest neighbouring subpopulations. Difficulties about the special position of subpopulations on the "edge" of the lattice area could be obviated by wrapping the lattice around a torus. However, even in this case the algebra becomes unwieldy. It is therefore necessary to confine ourselves to the case where each subpopulation receives the same number of migrants from each of the other subpopulations.

We consider a single locus with two possible alleles,  $a$  and  $A$ . We suppose that in generation  $t$  the number of  $a$  individuals in the  $i$ th subpopulation ( $i=1, \dots, H+1$ ) is  $k_{ti}$  so that the number of  $A$  individuals is  $N-k_{ti}$ . We can then write

$$k_{ti} = k_{t-1,i} + \sum_j^{i*} k_{t-1,j}, \quad \dots \dots \dots (1)$$

where  $k_{t-1,i}$  is the number of  $a$  individuals which are descended from  $k_{t-1,i}$  and  $k_{t-1,j}$  the number of migrant  $a$  individuals in population  $i$  descended from  $k_{t-1,j}$ .

The symbol  $\sum_j^{i*}$  is used to signify summation over all values of  $j \neq i$ .

The generations are taken as non-overlapping and the offspring from the  $i$ th population are taken to be  $a$  or  $A$  independently with probabilities  $k_{ti}N^{-1}$  and  $(N-k_{ti})N^{-1}$ . To calculate the asymptotic rate of progress to homozygosity we set up recurrence relations between the moments of the variates  $k$  at times  $t$  and  $t+1$ . For simplicity of notation we write  $t = 0$ . We then have

$$E(k_{1i}) = (N-L)N^{-1}k_{0i},$$

$$E(k_{1ij}) = KN^{-1}k_{0j}, \quad (i \neq j)$$

$$\text{var}(k_{1i}) = (N-L)N^{-2}k_{0i}(N-k_{0i}),$$

$$\text{var}(k_{1ij}) = (N-L)N^{-2}k_{0j}(N-k_{0j}).$$

These are expectations at time  $t = 1$  conditional on fixed values of the  $k$ 's at time  $t = 0$ . It follows that the  $k_{1ii}$  and the  $k_{1ij}$  are independent and therefore

$$\begin{aligned}\text{var}(k_{1i}) &= (N-L)N^{-2}k_{0i}(N-k_{0i}) + \sum_j^{i*} KN^{-2}k_{0j}(N-k_{0j}), \\ \text{var}(\sum_i k_{1i}) &= N^{-1} \sum_i k_{0i}(N-k_{0i}).\end{aligned}$$

We now write  $k_t = \sum_i k_{ti}$ . Then

$$\begin{aligned}E(k_1) &= \sum_i E k_{1i} = \sum_i E \left\{ k_{1ii} + \sum_j^{i*} k_{1ij} \right\} \\ &= N^{-1} \sum_i \left\{ (N-L)k_{0i} + \sum_j^{*i} K k_{0j} \right\} \\ &= \sum_i k_{0i} \\ &= k_0.\end{aligned}$$

Put

$$\begin{aligned}A_t &= \sum_i E(k_{ti}^2), \\ B_t &= \sum_i \sum_j^{i*} E(k_{ti} k_{tj}), \\ C_t &= \sum_i E(k_{ti}).\end{aligned}$$

Then from above,  $C_t = C_0$  and is constant. We also have

$$\begin{aligned}E(k_{1i}^2) &= (E k_{1i})^2 + \text{var}(k_{1i}) \\ &= N^{-2} \left\{ (N-L)k_{0i} + K \sum_j^{*i} k_{0j} \right\}^2 \\ &\quad + N^{-2} \left\{ (N-L)k_{0i}(N-k_{0i}) + \sum_j^{*i} K k_{0j}(N-k_{0j}) \right\}.\end{aligned}$$

Multiplying out and summing over all  $i$  we find

$$A_1 = N^{-2} \{ (N-L)^2 + KL - N \} A_0 + N^{-2} \{ 2NK - KL - K^2 \} B_0 + C_0. \quad \dots (2)$$

Next we have, for  $i \neq j$ ,

$$E(k_{1i} k_{1j}) = N^{-2} \left\{ (N-L)k_{0i} + K \sum_l^{*i} k_{0l} \right\} \left\{ (N-L)k_{0j} + K \sum_m^{*j} k_{0m} \right\}.$$

Multiplying out and summing we get

$$\begin{aligned}B_1 &= \sum_i \sum_j^{*i} E(k_{1i} k_{1j}) \\ &= N^{-2} \{ (N-L)^2 B_0 + 2(N-L)KHA_0 + 2(N-L)K(H-1)B_0 + K^2 B_0 \\ &\quad + 2K^2(H-1)B_0 + K^2 H(H-1)A_0 + K^2(H-1)(H-2)B_0 \} \\ &= N^{-2} \{ N^2 - 2NK + LK + K^2 \} B_0 + N^{-2} \{ 2NL - L^2 - KL \} A_0.\end{aligned}$$

The process may be regarded as a Markov chain in which the state of the system is determined by the  $(H+1)$  variables  $k_{ti} (i=1, \dots, H+1) = 0, \dots, N$ . There are exactly two absorbing states for which all the  $k_{ti}$  are zero, or all equal to  $N$ , and from any other state all states are accessible. The matrix of transition

probabilities is far too complicated to be written down explicitly but we may determine the moduli of its three largest roots which we write as  $\lambda_0$ ,  $\lambda_1$ , and  $\lambda_2$ . Clearly  $\lambda_0 = \lambda_1 = 1$  and there are no other roots of unit modulus. Starting from any non-absorbing state the probability of any other non-absorbing state at generation  $t$  will be asymptotically equal to a constant multiple of  $|\lambda_2|^t$  and the probabilities of the two absorbing states will be approximately of the form  $\alpha - \beta |\lambda_2|^t$ . Thus the expression

$$T = E \Sigma (k_{tt} - \frac{1}{2}N)^2, \dots\dots\dots(3)$$

will be asymptotically increasing and of the form  $\gamma - \delta |\lambda_2|^t$  where  $\delta > 0$ . Thus since  $T$  can be expressed in terms of  $A_t$  and  $C_t$  we can find  $|\lambda_2|$  by picking out the largest

TABLE 1  
VALUES FOR  $\nu$

K	H			
	1	2	4	8
1	0.8769	0.8953	0.9246	0.9527
2	0.9378	0.9460	0.9612	0.9752
4	0.9688	0.9726	0.9803	0.9878
8	0.9844	0.9862	0.9901	0.9939
16	0.9922	0.9930	0.9950	0.9970
100	0.9988	0.9989	0.9992	0.9994

root of the matrix of coefficients in the set of difference equations expressing  $A_{t+1}$ ,  $B_{t+1}$ , and  $C_{t+1}$  in terms of  $A_t$ ,  $B_t$ , and  $C_t$ .

Writing  $\lambda = 1 - \mu N^{-1}$ , we have to find the smallest root in  $\mu$  of the equation

$$\begin{vmatrix} \mu N - (2NL - L^2 - KL + N) & 2NK - KL - K^2 \\ 2NL - KL - L^2 & \mu N - (2NK - KL - K^2) \end{vmatrix} = 0$$

and this becomes, on division by  $N$ ,

$$-\mu^2 N - \mu \{ (K+L)(2N-K-L) + N \} + K(2N-K-L) = 0. \dots(4)$$

The total population size is  $(H+1)N$  haploid individuals and if there were no subdivision of the population the largest non unit root would be  $1 - \{(H+1)N\}^{-1}$  which would govern the asymptotic rate of approach to homozygosity. To estimate the effect of subdivision on this root we write

$$\lambda = 1 - \mu N^{-1} = 1 - \nu \{ (H+1)N \}^{-1},$$

and calculate  $\nu$  as a function of  $H$  and  $K$ . As  $K$  gets large we expect  $\nu$  to approach unity. Table 1 gives  $\nu$  for selected values of  $H$  and  $K$ , and shows that the effects

of subdivision on the asymptotic rate of progress to homozygosity is small and rapidly becomes negligible when  $K$ , the number of migrants exchanged between each subpopulation, increases beyond one or two. In fact, it is easily verified that for

$$\lambda = 1 - \nu\{(H+1)N\}^{-1},$$

$\nu$  rapidly tends to unity as  $H$  or  $K$  increases, for the asymptotic form of equation (4) is

$$\mu^2 - \mu\{2(H+1)K+1\} + 2K = 0.$$

and the relevant root of this is nearly  $(H+1)^{-1}$  when  $H$  or  $K$  is large.

These results show that geographical subdivision will have only a very small affect on the asymptotic rate of drift to homozygosity in small populations between which there is some small amount of intermigration. This conclusion, however, is dependent on there being no selective effects. The problem of the effect of selection on the rate of progress to homozygosity is a very difficult one and has been discussed for a single population by Wright and Kerr (1954) and by Kimura (1955). These papers estimate the result of a small selective effect, without dominance, on the principal root and show that the asymptotic rate of progress to homozygosity is increased.

### III. STABLE POLYMORPHISM IN A SUBDIVIDED POPULATION

The situation is entirely changed if we suppose that the population is divided into two parts in which selection operates in different directions. To discuss the theory of a stochastic model of this kind would be a very difficult undertaking. We shall confine ourselves to considering a deterministic model and show that there exists a stable state in which both alleles exist, i.e. a stable polymorphism similar to the well-known case in which the heterozygote has a selective advantage.

Consider two populations within each of which mating is at random so that we can represent the frequencies of the zygotes  $aa$ ,  $Aa$ , and  $AA$  by  $p_1^2$ ,  $2p_1q_1$ ,  $q_1^2$ , and  $p_2^2$ ,  $2p_2q_2$ ,  $q_2^2$  respectively. Next suppose that in the first population the relative reproductive powers of the zygotes are in the ratios  $(1+m):1:(1-m)$ . Similarly in the second population we suppose they are in the ratios  $(1-m):1:(1+m)$ , i.e. we assume that selection operates equally but in opposite directions. Then the frequency of gene  $a$  in the offspring from the first population will be

$$\frac{(1+m)p_1^2 + 2p_1q_1}{(1+m)p_1^2 + 2p_1q_1 + (1-m)q_1^2},$$

and if  $m$  is small this will be equal to  $p_1 + mp_1q_1$ . Similarly the  $a$  gene frequency in the second population will be  $p_2 - mp_2q_2$ . We now suppose that the next generation in each subpopulation consists of a fraction  $(1-k)$  of descendants of this population and a fraction  $k$  which are descendants of the other population. In general  $k$  will be small but for the following results to hold it is necessary only

that  $k < \frac{1}{2}$ . If the values of the frequencies of gene  $a$  in the new generation are  $p'_1$  and  $p'_2$  we then have, supposing  $m$  small,

$$p'_1 = p_1 + k(p_2 - p_1) + (1-k)mp_1q_1 - mkp_2q_2,$$

$$p'_2 = p_2 + k(p_1 - p_2) - (1-k)mp_2q_2 + mkp_1q_1.$$

We first consider what values  $P_1, P_2$  of  $p_1$  and  $p_2$  will result in a stationary state, i.e. one for which  $p'_1 = p_1$  and  $p'_2 = p_2$ . Writing  $l = k^{-1}$  we obtain from the above equations

$$P_2 - P_1 + (l-1)mP_1Q_1 - mP_2Q_2 = 0,$$

$$P_1 - P_2 - (l-1)mP_2Q_2 + mP_1Q_1 = 0,$$

where  $Q_1 = 1 - P_1$ ,  $Q_2 = 1 - P_2$ .

Adding we see that  $P_1Q_1 = P_2Q_2$ . Thus we must have  $P_1 = P_2$  or  $P_1 = Q_2$ . If  $P_1 = P_2$  we must have  $(l-2)P_1Q_1 = 0$  and so  $P_1Q_1 = P_2Q_2 = 0$  so that either  $P_1 = P_2 = 0$  or  $P_1 = P_2 = 1$ . Both of these are possible solutions. If  $P_1 = Q_2$  we substitute for  $P_2$  and  $Q_2$  in the first equation and find

$$P_1^2m(2-l) - P_1\{m(2-l)+2\} + 1 = 0,$$

and since  $l > 2$  we see that the root

$$\frac{1}{2} + \{m(2-l)\}^{-1} - \sqrt{\{\frac{1}{4} + m^{-2}(2-l)^{-2}\}}$$

is less than zero and so is not relevant. The other root

$$\frac{1}{2} + \{m(2-l)\}^{-1} + \sqrt{\{\frac{1}{4} + m^{-2}(2-l)^{-2}\}}$$

lies between  $\frac{1}{2}$  and 1 and results in a stationary solution.

Consider the stability of solution  $P_1 = P_2 = 0$ . For  $p_1$  and  $p_2$  small we have the approximate equations

$$p'_1 = p_1(1-k)(1+m) + p_2k(1-m),$$

$$p'_2 = p_1k(1+m) + p_2(1-k)(1-m).$$

These are a pair of linear difference equations and the population will be stable in the neighbourhood of  $p_1 = p_2 = 0$  if, and only if, the roots of the determinantal equation

$$\begin{vmatrix} \lambda - (1-k)(1+m) & -k(1-m) \\ -k(1+m) & \lambda - (1-k)(1-m) \end{vmatrix} = 0,$$

are less than unity in absolute value. The roots are

$$\lambda = (1-k) \pm \sqrt{\{k^2 + m^2(1-2k)\}}.$$

Since we have assumed  $k < \frac{1}{2}$  the term inside the radical is certainly greater than  $k^2$  and thus one root is greater than  $1 - k + k = 1$ . The solution  $P_1 = P_2 = 0$  is therefore unstable. Similarly the solution  $P_1 = P_2 = 1$  is unstable by symmetry.

For the third solution we have  $\frac{1}{2} < P_1 = Q_2 < 1$ . Write  $P_1 = Q_2 = P = 1 - Q$  say, and

$$\begin{aligned} p_1 &= P + \delta_1, & q_1 &= Q - \delta_1, \\ p_2 &= Q + \delta_2, & q_2 &= P - \delta_2, \end{aligned}$$

where  $\delta_1, \delta_2$  are small. Expanding the recurrence relation about the point  $(P, Q)$  we get

$$\begin{aligned} \delta_1' &= \delta_1(1-k)(1+m(Q-P)) + \delta_2 k(1+m(Q-P)), \\ \delta_2' &= \delta_1 k(1+m(Q-P)) + \delta_2(1-k)(1+m(Q-P)). \end{aligned}$$

The state  $(P, Q)$  will be stable if, and only if, the roots of the determinantal equation

$$\begin{vmatrix} \lambda - (1-k)(1+m(Q-P)) & k(1+m(Q-P)) \\ k(1+m(Q-P)) & \lambda - (1-k)(1+m(Q-P)) \end{vmatrix} = \lambda^2 - 2(1-k)(1+m(Q-P))\lambda + (1-2k)(1+m(Q-P))^2 = 0,$$

are less than unity in absolute value.

These roots are

$$\lambda = 1 + m(Q-P), \quad (1-2k)(1+m(Q-P)),$$

and since  $P > Q$  and  $k < \frac{1}{2}$  both these roots are less than unity so that a stable polymorphism exists. This model could be generalized without much difficulty to a situation in which selection in the two subpopulations is unequal so long as they are of opposite signs, and similarly the migration intensities could be made unequal.

Levene (1953) has considered a similar model with several subpopulations but supposes the individuals to leave their niches immediately before breeding thus forming one large panmictic population whose offspring are later distributed amongst the different niches.

Thus a deterministic model results in the existence of a stable polymorphism. If we take into account the fact that the populations are finite, the ultimate fate of the population is quite different since there are two absorbing states and one of these must ultimately be attained. The stable state with  $0 < P_1 < 1$  now has only a quasi-stable character in that there will be a strong tendency for the population to return to it when disturbed. Thus in the absence of mutation the rate of approach to homozygosity will be very much slower than that which would occur if there were no selection. Unfortunately the problem of calculating this rate appears to be very difficult.

The evolutionary significance of these results is that "drift" to homozygosity in small populations may be very much slowed down if the population contains two niches in which selection operates in opposite directions. This will be true even for two populations between which there is only a small amount of migration. However, in practice, we are unlikely to have niches which differ in their selective effects for only a single character and it may well happen that the accumulation of genetic differences between the two subpopulations finally results in barriers which reduce and ultimately eliminate cross-breeding.

## IV. REFERENCES

- KIMURA, M. (1955).—Stochastic processes and gene frequencies. *Cold Spring Harb. Symp. Quant. Biol.* **20**: 33–53.
- LEVENE, H. (1953).—Genetic equilibrium when more than one ecological niche is available. *Amer. Nat.* **27**: 331–3.
- WRIGHT, S., and KERR, W. E. (1954).—Experimental studies in the distribution of gene frequencies in very small populations of *Drosophila melanogaster*. II. *Evolution* **8**: 225–40.