THE STEPPING STONE MODEL OF POPULATION STRUCTURE AND THE DECREASE OF GENETIC CORRELATION WITH DISTANCE[1]

MOTOO KIMURA[2] and GEORGE H. WEISS[3]

c/o Department of Genetics, L.I.G.B. Pavia Section, University of Pavia, Italy, and Institute of Fluid Dynamics and Applied Mathematics, University of Maryland, College Park

WHEN a species occupies a very large territory, local differentiation is usually noticeable in the form of geographical races. Each race may in turn consist of numerous colonies which are differentiated to a less noticeable extent. The underlying differentiation in genetic constitution may reflect the local differences of selective pattern or may be the results of chance occurrence of different mutant genes, but these factors cannot act effectively unless some sort of isolation ensures the accumulation of genetic differences.

It is well known that existence of geographical barriers greatly favors the formation of races and new species. However, even if such barriers do not exist, the large size of the whole area as compared with the migration distance of an individual may prevent the species from forming a single panmictic unit, and this will produce a sort of isolation which WRIGHT called "isolation by distance" (WRIGHT 1943). He proposed a model of population structure in which a population is distributed uniformly over a large territory, but the parents of any given individual are drawn from a small surrounding region. He studied, by his method of path coefficients, the pattern of change in the inbreeding coefficient of subgroups relative to a larger population in which they are contained (WRIGHT 1940, 1943, 1946, 1951). The problem of local differentiation may also be studied in terms of change in correlation with distance as considered by MALÉCOT (1948, 1955, 1959); individuals living nearby tend to be more alike than those living far apart. In the mathematical theory of population genetics, the problem of local differentiation of gene frequencies in a structured population is one of the most intricate, and so far the main results are due to these two authors.

In natural populations, individuals often are distributed more or less discontinuously to form numerous colonies, and individuals may be exchanged between adjacent or nearby colonies. To analyze such a situation, one of us proposed a model which he termed "stepping stone model" of population structure (KIMURA

1953). The purpose of the present paper is to present a solution of this model and to discuss its biological implications.

Let us consider an infinite array of colonies with their position represented by integers on a line (Figure 1). The simplest situation for this one dimensional stepping stone model is that in each generation an individual can migrate at most "one step" in either direction between colonies. In other words, exchange of individuals is restricted to be between adjacent colonies.
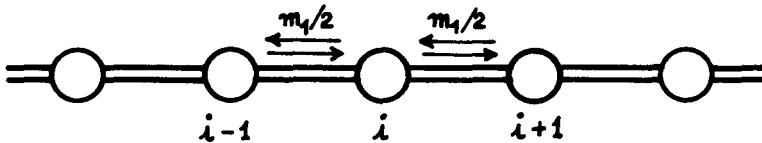

FIGURE 1.—One dimensional stepping stone model.

Consider a single locus with a pair of alleles A and A'. The frequency of A in each colony may change from generation to generation and to simplify the treatment we will assume discrete generation time. Also we assume that the gene frequency changes systematically by linear evolutionary pressures (mutation, migration but not selection, which will be considered later) and fortuitously by random sampling of gametes (small population number). If we denote by p_i the (relative) frequency of A in the ith colony in the present generation, then its value in the next generation may be given by

$$(1.1) \qquad p'_i = (1 - m_1 - m_\infty) p_i + \frac{m_1}{2} (p_{i-1} + p_{i+1}) + m_\infty \bar{p} + \xi_i$$

In the above expression, m_1 stands for the rate of migration per generation to neighboring colonies such that $m_1/2$ is the proportion of individuals exchanged each generation between a pair of adjacent colonies. Also m_∞ stands for the rate of long range dispersal per generation, namely the rate by which a colony exchanges individuals in each generation with a random sample taken from the entire population in which the frequency of A is $\bar{p}$. The effect of the long range dispersal is formally equivalent to mutation; if there is mutation between A and A' in addition to long range dispersal of gametes, then m_∞ should be replaced by $\mu + v + m_\infty$ and $m_\infty \bar{p}$ by $v + m_\infty \bar{p}$, where μ is the mutation rate from A to A' and v is the mutation rate in the reverse direction. Furthermore, ξ_i stands for the change in p_i due to random sampling of gametes in reproduction, namely, due to the relatively small number of gametes being randomly chosen to form the next generation out of the very large number of gametes produced by the parents. Thus, if N_e is the effective size (cf. WRIGHT 1940; KIMURA and CROW 1963) of the colony, ξ_i follows the binomial distribution with mean and variance given by

$$E_\delta (\xi_i) = 0 \text{ and } E_\delta (\xi_i{}^2) = \frac{p_i(1 - p_i)}{2N_e}$$

where E_{δ} stands for an operator of taking expectation with respect to this random change.

Let us denote by $\tilde{p}_i$ the deviation of gene frequency from its mean, i.e.

$$\tilde{p}_i = p_i - \bar{p},$$

then (1.1) becomes

(1.2) $$\tilde{p}'_i = \alpha \tilde{p}_i + \beta(\tilde{p}_{i-1} + \tilde{p}_{i+1}) + \xi_i,$$

where $\alpha = 1 - m_1 - m_{\infty}$ and $\beta = m_1/2$.

We will denote by V_p the variance in the probability distribution of the gene frequency among colonies,

$$V_p = E_{\phi}(\tilde{p}_i{}^2),$$

and by r_k the correlation coefficient of the gene frequencies between two colonies which are k steps apart,

$$r_k = E_{\phi}(\tilde{p}_i \tilde{p}_{i+k})/V_p,$$

where E_{ϕ} stands for an operator for taking expectation with respect to gene frequency distribution among colonies.

In order, to obtain the formula for the variance, we square both sides of (1.2) and take expectations. Noting that the product terms between $\tilde{p}$'s and ξ_i have expectation zero and that

$$E(\xi_i{}^2) = E_{\phi}\{E_{\delta}(\xi_i{}^2)\} = E_{\phi}\left\{\frac{p_i(1-p_i)}{2N_e}\right\}$$

$$= \{\bar{p} - (V_p + \bar{p}^2)\}/(2N_e),$$

we obtain

$$V_p{}' = \alpha^2 V_p + 4\alpha\beta V_p r_1 + 2\beta^2 V_p(1+r_2) - \frac{V_p}{2N_e} + \frac{\bar{p}(1-\bar{p})}{2N_e}$$

where the prime indicates that it is the value in the next generation. At equilibrium in which $V'_p = V_p$, the above reduces to

(1.3) $$V_p\left\{1 - \alpha^2 - 4\alpha\beta r_1 - 2\beta^2(1+r_2) + \frac{1}{2N_e}\right\} = \frac{\bar{p}(1-\bar{p})}{2N_e}$$

In order to obtain the formula for the correlation coefficients, we will consider the expectation of product $\tilde{p}'_i \tilde{p}'_{i+k}$ ($k \neq 0$). Noting that terms like $\tilde{p}_{i+k}\xi_i$ as well as $\xi_i\xi_{i+k}$ have zero expectations, we obtain

$$r'_k = \alpha^2 r_k + 2\alpha\beta(r_{k+1} + r_{k-1}) + \beta^2(r_{k+2} + 2r_k + r_{k-2})$$

At equilibrium in which $r'_k = r_k$, this reduces to

(1.4) $(\alpha^2 + 2\beta^2 - 1)r_k + 2\alpha\beta(r_{k+1} + r_{k-1}) + \beta^2(r_{k+2} + r_{k-2}) = 0.$ $(k \neq 0)$

Equation (1.4) holds for $k \geqq 1$. However, for $k = 1$, r_{-1} should be replaced by r_1 to give

(1.5) $(\alpha^2 + 2\beta^2 - 1)r_1 + 2\alpha\beta(r_2 + 1) + \beta^2(r_3 + r_1) = 0.$

In order to solve (1.4), let $r_k = \lambda^k$ and substitute in (1.4). This leads to a 4th order equation in λ with the following four roots;

$$(1.6) \begin{cases} \lambda_1 = \dfrac{1}{2\beta}\left\{ (1-\alpha) + \sqrt{(1-\alpha)^2 - (2\beta)^2} \right\} \\[2mm] \lambda_2 = \dfrac{1}{2\beta}\left\{ (1-\alpha) - \sqrt{(1-\alpha)^2 - (2\beta)^2} \right\} \\[2mm] \lambda_3 = \dfrac{-1}{2\beta}\left\{ (1+\alpha) + \sqrt{(1+\alpha)^2 - (2\beta)^2} \right\} \\[2mm] \lambda_4 = \dfrac{-1}{2\beta}\left\{ (1+\alpha) - \sqrt{(1+\alpha)^2 - (2\beta)^2} \right\} \end{cases}$$

where $\lambda_1 > 1$, $1 > \lambda_2 > 0$, $\lambda_3 < -1$, and $-1 < \lambda_4 < 0$.

The required solution should then be expressed in the form

$$(1.7) \qquad r_k = \sum_{i=1}^{4} C_i \lambda_i{}^k$$

where the C_i's are constants, which may be determined through the following consideration: First, in order that r_k vanishes at $k = \infty$, we must have $C_1 = C_3 = 0$, since both λ_1 and λ_3 are larger than unity in absolute value. Secondedly, $r_0 = 1$ by definition and also r_k has to satisfy relation (1.5). From these requirements, C_2 and C_4 are determined;

$$(1.8) \qquad C_2 = \frac{R_1}{R_1 + R_2} \text{ and } C_4 = \frac{R_2}{R_1 + R_2},$$

where

$$R_1 = \sqrt{(1+\alpha)^2 - (2\beta)^2}$$

and

$$R_2 = \sqrt{(1-\alpha)^2 - (2\beta)^2},$$

in which $\alpha = 1 - m_1 - m_x$ and $2\beta = m_1$.

Therefore, writing $r(k)$ instead of r_k, we obtain

$$(1.9) \qquad r(k) = C_2 \lambda_2{}^k + C_4 \lambda_4{}^k$$

as the correlation of gene frequencies between two colonies which are k steps apart, with λ_2 and λ_4 given by (1.6) and C_2 and C_4 given by (1.8).

Substituting the values of r_1 and r_2 into (1.3), we get

$$(1.10) \qquad V_p = \frac{\bar{p}(1-\bar{p})}{1 + 2N_e C_0}$$

where $C_0 = 2R_1 R_2 / (R_1 + R_2)$.

In the special case $m_1 = 0$, the above formula reduces to

$$(1.11) \qquad V_p = \frac{\bar{p}(1-\bar{p})}{1 + 2N_e(2m_x - m^2{}_x)}.$$

This case should correspond to WRIGHTS "island model" (WRIGHT 1943) and indeed (1.11) agrees with his formula except for the negligible term $m^2{}_x$ in the denominator.

For our purpose, however, a really interesting case is one in which $m_1 \gg m_\infty$. In this case, we have approximately

$$R_1 = 2\sqrt{1 - m_1}, \qquad R_2 = \sqrt{2m_1 \, m_\infty}$$

and (1.10) reduces to

(1.12)
$$V_p = \frac{\bar{p}(1 - \bar{p})}{1 + 4N_e \sqrt{2m_1 m_\infty}}$$

Furthermore, (1.9) may be approximated by

(1.13)
$$r(k) = e^{-\sqrt{\frac{2m_\infty}{m_1}}\, k}$$

with sufficient accuracy.

GENERAL TREATMENT INCLUDING TWO AND THREE DIMENSIONS

Since the detailed account of the general treatment will be published elsewhere (WEISS and KIMURA 1964), only the main results will be presented here.

In the one dimensional case, the correlation of gene frequencies between colonies which are k steps apart may be given by

(2.1)
$$r(k) = \frac{C_0}{2\pi} \int_0^{2\pi} \frac{\cos k\theta \, d\theta}{1 - H^2(\cos\theta)}$$

where

(2.2)
$$C_0^{-1} = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{1 - H^2(\cos\theta)}$$

and

(2.3)
$$H(\cos\theta) = 1 - m_\infty - m_1(1 - \cos\theta).$$

Then the variance of gene frequencies between colonies is expressed in the form:

(2.4)
$$V_p = \frac{\bar{p}(1 - \bar{p})}{1 + 2N_e C_0}.$$

It is possible to show that the above results agree with those given in the previous section; formula (2.1) and (2.4) respectively reduce to (1.9) and (1.10).

For numerical calculations, the following expressions turn out to be useful:

(2.5)
$$r(k) = \frac{A_1(k) + A_2(k)}{A_1(0) + A_2(0)}$$

(2.6)
$$C_0^{-1} = A_1(0) + A_2(0),$$

where

(2.7) $\quad A_1(k) = \dfrac{1}{2} \cdot \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} \dfrac{\cos k\theta \, d\theta}{1 - H(\cos\theta)} = \dfrac{1}{4\pi} \displaystyle\int_0^{2\pi} \dfrac{\cos k\theta \, d\theta}{m_\infty + m_1(1 - \cos\theta)}$

and

(2.8) $\quad A_2(k) = \dfrac{1}{2} \cdot \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} \dfrac{\cos k\theta \, d\theta}{1 + H(\cos\theta)} = \dfrac{1}{4\pi} \displaystyle\int_0^{2\pi} \dfrac{\cos k\theta \, d\theta}{2 - m_\infty - m_1(1 - \cos\theta)}$

For small values of m_1 and m_∞ which we are interested in, the contribution of $A_2(k)$ is negligible in comparison with that of $A_1(k)$.
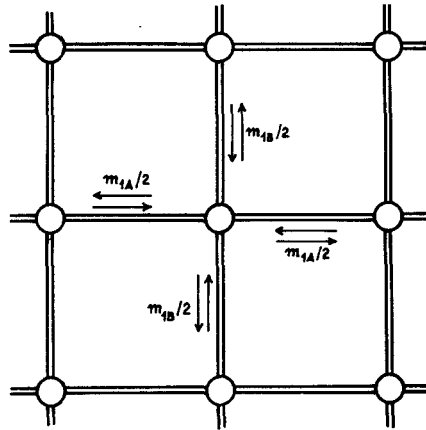
FIGURE 2.—Two dimensional stepping stone model.

In the two dimensional case (Figure 2), we assume that the entire population consists of a rectangular array of colonies, each of which occupies a point denoted by a pair of integers (k_1, k_2). We will also assume that in each generation, an individual colony exchanges migrants with four surrounding colonies, but the effective population number in each colony remains the same (N_e). The rate of migration may be different in X and Y directions: let m_{1A} be the rate (taking one generation as a unit) of migration along the X-axis or horizontal direction, such that $m_{1A}/2$ is the proportion of individuals exchanged between a pair of adjacent colonies in this direction. Similarly let m_{1B} be the rate of migration per generation along the Y-axis. The proportion of individuals which migrates to four neighboring colonies per generation is $m_1 = m_{1A} + m_{1B}$.

We will denote by m_x the rate of long range dispersal as defined in the one dimensional case.

At equilibrium, the correlation of gene frequencies between colonies which are k_1 steps apart in the X direction and k_2 steps apart in the Y direction may be written

$$(2.9) \qquad r(k_1, k_2) = \frac{A_1(k_1, k_2) + A_2(k_1, k_2)}{A_1(0, 0) + A_2(0, 0)}$$

where

$$(2.10) \quad A_1(k_1, k_2) = \frac{1}{2} \cdot \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \frac{\cos k_1 \theta_1 \cos k_2 \theta_2 \, d\theta_1 \, d\theta_2}{m_x + m_{1A}(1 - \cos \theta_1) + m_{1B}(1 - \cos \theta_2)}$$

and

$$(2.11) \quad A_2(k_1, k_2) = \frac{1}{2} \cdot \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \frac{\cos k_1 \theta_1 \cos k_2 \theta_2 \, d\theta_1 \, d\theta_2}{2 - m_x - m_{1A}(1 - \cos \theta_1) - m_{1B}(1 - \cos \theta_2)}$$

The variance of gene frequencies between colonies is given by

$$(2.12) \qquad V_p = \frac{\bar{p}(1 - \bar{p})}{1 + 2N_e C_0}$$

where N_e is the effective population number of a colony and

(2.13)
$$C_0^{-1} = A_1(0, 0) + A_2(0, 0).$$

The last formula can be expressed in terms of a complete elliptic integral:

(2.14)
$$C_0^{-1} = \frac{1}{2\pi\sqrt{m_{1A}\, m_{1B}}} \left\{ \frac{1}{M_1} K\left(\frac{1}{M_1}\right) + \frac{1}{M_2} K\left(\frac{1}{M_2}\right) \right\}$$

where

$$M_1 = \sqrt{\left(1 + \frac{m_\infty}{2m_{1A}}\right)\left(1 + \frac{m_\infty}{2m_{1B}}\right)},$$

$$M_2 = \sqrt{\left(1 - \frac{2 - m_\infty}{2m_{1A}}\right)\left(1 - \frac{2 - m_\infty}{2m_{1B}}\right)}$$

and $K(\cdot)$ stands for the complete elliptic integral of the first kind defined by

(2.15)
$$K(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2\theta}}.$$

The following approximations are useful to evaluate C_0^{-1} when m_∞ is much smaller than m_{1A} and m_{1B}, and also m_{1A} and m_{1B} themselves are small:

(2.16)
$$K(1 - \varepsilon) \approx \log_e \frac{4}{\sqrt{2\varepsilon}} \qquad (0 < \varepsilon \ll 1)$$

(2.17)
$$K(\varepsilon) \approx \frac{\pi}{2}\left(1 + \frac{\varepsilon^2}{4}\right)$$

In order to obtain an accurate figure for the correlation from (2.9) for given values of k_1 and k_2, numerical integration has to be employed to evaluate (2.10) and (2.11). However, if the distance between colonies is large, simpler expressions are available. Namely, if we put

(2.18)
$$\zeta = \left(\frac{k_1^2}{m_{1A}} + \frac{k_2^2}{m_{1B}}\right)^{1/2}$$

then, we obtain as an approximation for large ζ

(2.19)
$$A_1(k_1, k_2) = \frac{1}{2\pi\sqrt{m_{1A}\, m_{1B}}} K_0(\sqrt{2m_\infty}\, \zeta)$$

and

(2.20)
$$A_2(k_1, k_2) = \frac{(-1)^{k_1 + k_2}}{2\pi\sqrt{m_{1A}\, m_{1B}}} K_0(\sqrt{2(2 - m_\infty - 2(m_{1A} + m_{1B}))}\, \zeta)$$

where $K_0(\cdot)$ is the modified Bessel function of zeroth order.

If m_∞ is much smaller than m_{1A} and m_{1B}, which are themselves small as compared with unity, $A_2(k_1, k_2)$ is neglible as compared with $A_1(k_1, k_2)$. Then we get

(2.21)
$$r(k_1, k_2) = \frac{C_0}{2\pi\sqrt{m_{1A}\, m_{1B}}} K_0(\sqrt{2m_\infty}\, \zeta)$$

Furthermore, if the rate of migration is equal in X and Y directions, so that $m_{1A} = m_{1B} = m_1/2$, the above formula reduces to

(2.22)
$$r(\rho) = \frac{C_0}{\pi m_1} K_0\left(\sqrt{\frac{4m_\infty}{m_1}}\, \rho\right),$$

where $\rho = \sqrt{k_1{}^2 + k_2{}^2}$ is the distance between two colonies, and $m_1 = m_{1A} + m_{1B}$. Since asymptotically

$$K_0(z) = \sqrt{\frac{\pi}{2z}} e^{-z}, \qquad (z \to \infty),$$

the correlation at long distance is proportional to

$$e^{-\sqrt{\frac{4m_\infty}{m_1}}\,\rho} \Big/ \sqrt{\rho},$$

which shows that the correlation falls off more quickly than in the one dimensional case.

The variance in gene frequency between colonies is

(2.23)
$$V_p = \frac{\bar{p}(1-\bar{p})}{1 + 2N_eC_0}$$

where C_0 is given by (2.14).

We will now turn to the three dimensional case. Let us assume that the cubic array of colonies extends to infinity in all directions. In the present case, each colony has six adjacent colonies to exchange individuals in each generation. The rates of migration in three perpendicular directions may be different and we will denote by m_{1A}, m_{1B} and m_{1C} the respective rates per generation in the X, Y and Z directions. Thus the amount of individuals exchanged per generation between two adjacent colonies parallel to the X direction is $m_{1A}/2$, between those parallel to the Y direction is $m_{1B}/2$, etc. As before m_x stands for the rate of long range dispersal per generation. The position of each colony may be designated by a triplet of integers (k_1, k_2, k_3), and we assume that each colony has the effective population number of N_e, which is constant in each generation.

The correlation of gene frequencies at equilibrium between two colonies which are respectively k_1, k_2 and k_3 steps apart in the X, Y and Z directions is

(2.24)
$$r(k_1, k_2, k_3) = \frac{A_1(k_1, k_2, k_3) + A_2(k_1, k_2, k_3)}{A_1(0, 0, 0,) + A_2(0, 0, 0)}$$

where

(2.25)
$$A_1(k_1,k_2,k_3) = \frac{1}{2\pi^3} \int_0^\pi \int_0^\pi \int_0^\pi \frac{\cos k_1\theta_1 \cos k_2\theta_2 \cos k_3\theta_3\, d\theta_1 d\theta_2 d\theta_3}{m_x + m_{1A}(1-\cos\theta_1) + m_{1B}(1-\cos\theta_2) - m_{1C}(1-\cos\theta_3)}$$

and

(2.26)
$$A_2(k_1,k_2,k_3) = \frac{1}{2\pi^3} \int_0^\pi \int_0^\pi \int_0^\pi \frac{\cos k_1\theta_1 \cos k_2\theta_2 \cos k_3\theta_3\, d\theta_1\, d\theta_2\, d\theta_3}{2 - m_x - m_{1A}(1-\cos\theta_1) - m_{1B}(1-\cos\theta_2) - m_{1C}(1-\cos\theta_3)}$$

Generally, numerical integration has to be employed to evaluate these integrals. Fortunately, however, tables are available for the important case of $m_{1A} = m_{1B}$. This represents an isotropic migration in a plane but a different amount of migration in the third dimension. In this case, the integrals can be expressed in terms of the Green's functions for monatomic simple cubic lattices which are defined by

(2.27)
$$I(a,b,c;\alpha;\beta) = \frac{1}{\pi^3} \int_0^\pi \int_0^\pi \int_0^\pi \frac{\cos ax \cos bx \cos cz\, dx\, dy\, dz}{(2+\alpha)\beta - \cos x - \cos y - \alpha\cos z}$$

This function is extensively tabulated by MARADUDIN et al. (1960). In terms of this function, A's are expressed as follows:

$$(2.28) \qquad A_1(k_1,k_2,k_3) = \frac{1}{2m_{1A}} I\left(k_1,k_2,k_3; \frac{m_{1C}}{m_{1A}}; 1 + \frac{m_\infty}{2m_{1A}+m_{1C}}\right)$$

$$(2.29) \quad A_2(k_1,k_2,k_3) = \frac{(-1)^{k_1+k_2+k_3}}{2m_{1A}} I\left(k_1,k_2,k_3; \frac{m_{1C}}{m_{1A}}; \frac{2-m_\infty}{2m_{1A}+m_{1C}} -1\right)$$

The variance of gene frequencies between colonies is given by

$$V_p = \frac{\bar{p}(1-\bar{p})}{1 + 2N_e C_0}$$

(2.30)

where

(2.31) $$C_0^{-1} = A_1(0,0,0) + A_2(0,0,0).$$

If

$$R^2 = k_1{}^2 + k_2{}^2 + \frac{m_{1A}}{m_{1C}} k_3{}^2$$

is much larger than unity, a simple approximation formula is available for the correlation:

$$(2.32) \qquad r(k_1,k_2,k_3) = \frac{C_0 e^{-\sqrt{\frac{2m_\infty}{m_{1A}}} R}}{4\pi \sqrt{m_{1A}m_{1C}} R}.$$

In addition, if $m_{1A} = m_{1C}$ i.e. migration is completely isotropic and if m_∞ is much smaller than m_{1A} and also m_{1A} is at most of the order of a few percent, we have approximately

$$C_0 = 4m_{1A}.$$

In this case, the correlation of gene frequencies between two colonies which are distance ρ apart may be given by the following simple approximation formula

$$(2.33) \qquad r(\rho) = \pi^{-1} e^{-\sqrt{\frac{6m_\infty}{m_1}} \rho} \Big/ \rho$$

where

$$\rho = (k_1{}^2 + k_2{}^2 + k_3{}^2)^{1/2}$$

and

$$m_1 = m_{1A} + m_{1B} + m_{1C}$$

Thus, at long distance, the correlation falls off more quickly in the three dimensional case than in the two dimensional one, which in turn falls off more rapidly than the one dimensional case.

In Figure 3, the relation between the correlation coefficient $r(\rho)$ and distance ρ is plotted for the one, two and three dimensional cases, taking $m_\infty = 4 \times 10^{-5}$, $m_1 = 0.1$ and assuming complete isotropic migration, namely, $m_{1A} = m_{1B} = 0.05$ for two dimensions and $m_{1A} = m_{1B} = m_{1C} = 0.0333$ for three dimensions. It can be seen that a distinct difference exists between the three cases.

DISCUSSION

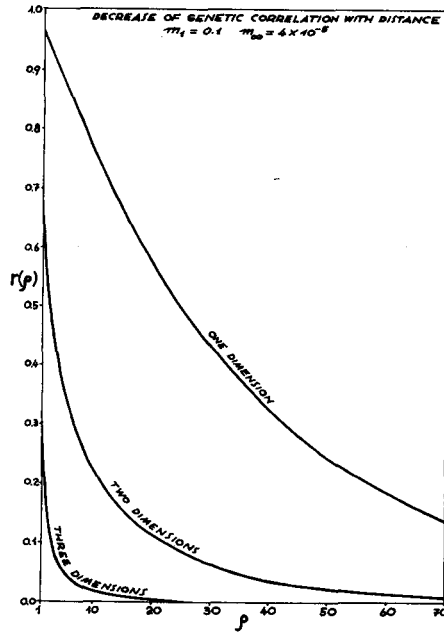(i) *More general forms of migration:* In the foregoing treatments, we have as-

FIGURE 3.—Graphs showing the decrease of genetic correlation with distance for one, two and three dimensions taking $m_\infty = 4 \times 10^{-5}$, $m_1 = 0.1$ and assuming complete isotropic migration. Abscissa: distance between two colonies (ρ). Ordinate: correlation coefficient of gene frequencies between colonies, $r(\rho)$.

sumed that the migration is restricted to one step per generation. However, the results can be extended to the situation where there is more than one step migration per generation. For most purposes this can be done simply by substituting for m_1 the variance of migration distance per generation. For example, in the one dimensional case, if $m_j/2$ is the proportion of individuals exchanged per generation between two colonies which are j steps apart, then

$$\sigma_m^2 = \sum_{j=1}^{\infty} j^2 m_j$$

should be replaced for m_1 in (1.13), to give

(3.1) $$r(k) = e^{-\sqrt{\frac{2m_\infty}{\sigma_m}} k}$$

In the two dimensional case, let $m_{ij}/2$ be the proportion of individuals exchanged in each generation between two colonies, one at $(0,0)$ and the other at (i,j). If we define by

$$\sigma_{mA}^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i^2 m_{ij}$$

and

$$\sigma_{mB}^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} j^2 m_{ij},$$

the two variances, then the variance $\sigma^2{}_{mA}$ should be used instead of m_{1A}, and similarly the variance $\sigma^2{}_{mB}$ should be used instead of m_{1B}, to calculate the correlation coefficient and the variance.

More detailed discussion together with the mathematical justification will be given in WEISS and KIMURA (1964).

(ii) *Relationship between the stepping stone model and the "continuum model":* The latter model may be obtained as a limiting case of the former by letting the actual distance between the adjacent subgroups approach zero. The various results obtained for the stepping-stone model may easily be adapted to the continuum case by substituting σ^2 for m_1 and $(N_e/N)\,\delta$ for N_e, where σ^2 is the variance of the migration distance of individual per generation, N_e/N is the ratio between the effective number and the actual number of individuals per subgroup and δ is the density of individuals or the number of inhabitants per unit size of the habitat (e.g. per unit area in the case of two dimensional habitat).

We may note here that the correlation of gene frequencies does not depend on the number of individuals or density, while the variance is clearly dependent on it. The latter, namely V_p, in the stepping stone model may be translated into the inbreeding coefficient of an individual in the continuum by the relation:

(3.2)
$$f_0 = \frac{V_p}{\overline{p}\,(1 - \overline{p})}$$

This relation is derived easily by considering the expected frequency of heterozygotes, i.e.

$$E\,\{2p(1 - p)\} = 2\overline{p}(1 - \overline{p})\,(1 - f_0)$$

which yields

$$2\,\{\overline{p} - (V_p + \overline{p}^2)\} = 2\overline{p}(1 - \overline{p})\,(1 - f_0)$$

or

$$V_p = \overline{p}(1 - \overline{p})f_0$$

(iii) *Effect of selection:* In our original treatment, only mutation and migration are assumed as factors which cause systematic change in gene frequency. However, we can incorporate selection to the extent that its effect on the change of gene frequency can be expressed linearly. There are two important cases where this can be done as a good approximation.

Firstly, in a polymorphic locus, if deviation of gene frequency p_i from its equilibrium frequency $\hat{p}$ is small, the amount of change in gene frequency per generation may be expressed in the form

(3.3)
$$- K(p_i - \hat{p})$$

where K depends on the selection intensity and $\hat{p}$ may be substituted by $\overline{p}$. For example, if A and A' are heterotic and if s_1, and s_2 are respectively the selection coefficients against both homozygotes AA and $A'A'$ as compared with heterozygote AA', then we may take $K = s_1 s_2/(s_1 + s_2)$ and $\hat{p} = s_2/(s_1 + s_2)$. Therefore expression (1.1) is unchanged, except that m_∞ now include K. In such circumstance, the mutation rates may be negligible as compared with K and if m_∞ is also negligible, then we may take $m_\infty = K$ in all the subsequent formulae.

Secondly, if gene A is unconditionally deleterious, kept in low frequency by

the balance between recurrent mutation and selective elimination, then the selection coefficient against gene A may be included in m_x. With no long range dispersal we may put $m_x = s$ and $m_x\bar{p} = v$ since $\mu + v$ may be neglected as compared with s. Here s is equal to the reciprocal of the average number of generations through which a gene A persists in a population. The incidence of mutant homozygotes among the offspring of marriages within a colony is $\Sigma(\bar{p}^2 + V_p)$, while that of marriages between two neighboring colonies is $\Sigma(\bar{p}^2 + V_p r(1))$ where the summation is over all relevant loci.

(iv) *Probability distribution of gene frequencies and the amount of random local differentiation:* In WRIGHT's island model, immigrants represent a random sample from the entire population and the probability distribution of gene frequency among colonies is given by

$$(3.4) \qquad \phi(p) = \frac{\Gamma(4Nm)}{\Gamma(4Nm\bar{p})\,\Gamma(4Nm(1-\bar{p}))} \; p^{4Nm\bar{p}-1}\,(1-p)^{4Nm(1-\bar{p})-1}$$

where m corresponds to our m_x. WRIGHT (1940) states that "if there is a correlation, r, between immigrants and receiving group, the m of the formula must be replaced by $m(1-r)$ if m is to continue to be the actual amount of replacement by immigration." Thus in our stepping stone model, we may take

$$m = m_x + m_1(1-r(1))$$

to obtain the approximate gene frequency distribution from (3.4). This means that in the typical case of $m_x \ll m_1$, $1-r(1)$ is an important quantity relating to the amount of local differentiation of gene frequencies due to random sampling of gametes. If $4N_e m_1(1-r(1))$ is less than $1/\bar{p}$ and $1/(1-\bar{p})$, the curve giving gene frequency distribution is U-shaped and strong local differentiation will be expected. More generally, the variance in gene frequency distribution may be given approximately by

$$(3.5) \qquad V_p = \frac{\bar{p}(1-\bar{p})}{1 + 4Nm_1(1-r(1))}$$

if $m_x \ll m_1 \ll 1$.

In the one dimensional case, the exact value of $1 - r(1)$ may be obtained from

$$(3.6) \qquad 1 - r(1) = \frac{1}{4\alpha\beta}\left\{ R_1 R_2 - 1 + (1-m_x)^2 \right\}.$$

If m_x is much smaller than m_1, which in turn is much smaller than unity,

$$(3.7) \qquad 1 - r(1) = \sqrt{\frac{2m_x}{m_1}}$$

approximately. In the two dimensional case, assuming isotropy ($m_{1A}=m_{1B}=m_1/2$) the exact expression for $1-r(1)$ is

$$(3.8) \qquad 1-r(1,0) = \frac{1}{m_1}\left\{ \frac{(2-m_x)\,A_2\,(0,0)-m_x\,A_1\,(0,0)}{A_1\,(0,0) + A_2\,(0,0)} \right\}$$

If $m_x \ll m_1 \ll 1$, we have roughly

$$(3.9) \qquad 1 - r(1,0) = \frac{\pi}{2} \left(\log_e \frac{4}{\sqrt{\dfrac{2m_\infty}{m_1}}} \right)^{-1}$$

Again if we assume isotropy, the corresponding expression for the three dimensional case is

$$(3.10) \qquad 1 - r(1,0,0) = \frac{1}{m_1} \left\{ \frac{(2 - m_\infty) \, A_2 \, (0,0,0) - m_\infty \, A_1 \, (0,0,0)}{A_1 \, (0,0,0) + A_2 \, (0,0,0)} \right\}$$

where $m_1 = 3m_{1A} = 3m_{1B} = 3m_{1C}$. If in addition $m_\infty \ll m_1 \ll 1$, this quantity becomes almost independent of m_∞ and m_1, and we have roughly

$$(3.11) \qquad \qquad 1 - r \, (1,0,0) = 2/3.$$

These results clearly show that the tendency toward local differentiation is very much dependent on the number of dimensions.

In the one dimensional case, $4N_e m_1 (1 - r(1))$ is approximately $4N_e \sqrt{2m_1 m_\infty}$ and it can easily be less than unity if m_∞ is very small: with $m_1 = 0.1$ and $m_\infty = 2 \times 10^{-5}$, a considerable local differentiation will be expected if N_e is less than 100. On the other hand, in two dimensions, a tendency toward local differentiation due to random sampling will generally be rather weak. For example, with $m_1 = 0.1$ and $m_\infty = 2 \times 10^{-5}$, $4N_e m_1 (1 - r(1,0))$ becomes less than unity only when N_e is less then about nine. In the three dimensional case, the tendency will be still weaker, since $4N_e m_1 (1 - r(1))$ is approximately $8N_e m_1/3$. It is remarkable that this quantity is now independent of m_∞. This means that however small the value of m_∞ is, random differentiation is impossible in three dimension unless m_1 is very small.

WRIGHT (1940, 1943, 1946) studied the same type of problem for the continuum using an entirely different approach, and arrived at a similar conclusion with respect to one and two dimensions.

(v) *Decrease of genetic correlation with distance:* The fact that tendency toward random differentiation depends very much on the number of dimensions is also reflected in the way in which the correlation coefficient falls off with distance, especially when the distance is large ($\rho \to \infty$): In the one dimensional case, the correlation falls off exponentially with distance

$$(3.12) \qquad \qquad r(\rho) \propto e^{-\sqrt{\frac{2m_\infty}{m_1}} \rho} \qquad ;$$

in two dimensions, it falls off more rapidly, namely

$$(3.13) \qquad \qquad r(\rho) \propto e^{-\sqrt{\frac{4m_\infty}{m_1}} \rho} \Big/ \sqrt{\rho} \qquad ;$$

and in three dimensions, it falls off still more rapidly

(3.14) $$r(\rho) \propto e^{-\sqrt{\frac{6m_\infty}{m_1}}\rho}\Big/\rho \quad .$$

MALÉCOT (1955) obtained a result, based on his elegant method of using an integro-differential equation for a continuum, that the coefficient of relationship as well as the correlation coefficient between gene frequencies decrease approximately in exponential form with distance independent of dimension. It is probable that our present results are more accurate than his result.

(vi) *Concept of dimension:* The one dimensional model can represent a population of organisms living along a river, coastal line or mountain ridge. The two dimensional model can represent a population on a plane and cover the most important cases in nature. CAVALLI-SFORZA and CONTERIO (1960) introduced "coefficient of dimensionality" to measure the pattern of geographical distribution of villages. It varies from 1 to 2. The intermediate case may be represented by the two dimensional stepping stone model with different migration rates in the X and Y directions. The three dimensional model can represent a population in an oceanic habitat and the treatment assuming equal rate of migration in horizontal (X and Y) directions but different rate in a vertical (Z) direction will be useful.

The three dimensional model can also represent a population of organisms living on a plane, but there is a third dimension such as the social rank in which "migration" is restricted to the neighboring classes, if distance is well defined in this third dimension. The results presented in this paper can readily be extended to cover the cases of four and higher dimensions, but they do not appear to have an important application to natural populations.

(vii) *Variance of gene frequency within a restricted region:* In the present paper, V_p represents the theoretical variance of gene frequencies between colonies in an infinitely large distribution range. On the other hand, actual observations cover only a restricted area. Therefore, in applying the present theory to actual data, it is necessary to derive, for each dimension, a formula for the variance of gene frequency within a restricted region. Here we will consider the one and two dimensional cases.

In the one dimensional case, the variance of gene frequency between colonies within a group of n consecutive colonies may be given by

(3.15) $$V_w(n) = V_p \left\{ 1 - \frac{1}{n} - \frac{2}{n^2} \sum_{i=1}^{n} (n-i)\, r(i) \right\} .$$

In the two dimensional case, if we consider a squared region of length n_1 in X directions and length n_2 in Y directions where $n_1 n_2$ colonies are contained, then the variance of gene frequency between colonies within such region is

$$(3.16) \qquad V_w\ (n_1,n_2) = V_p \left[\ 1 - \frac{1}{n_1 n_2} - \frac{2}{(n_1 n_2)^2} \left\{ \sum_{i=1}^{n_1} (n_1 - i) n_2 r(i,0) \right.\right.$$

$$+ \sum_{j=1}^{n_2} n_1 (n_2 - j)\ r(0,j)$$

$$\left.\left. + 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (n_1 - i)(n_2 - j)\ r(i,j) \right\} \right] .$$

SUMMARY

If the distance of individual migration is much smaller as compared with the entire distribution range of the species, the random local differentiation in gene frequency will be expected as shown by Wright in his studies on "isolation by distance". In the present paper, the stepping stone model is used to study this phenomenon. The model assumes that the entire population is subdivided into colonies and the migration of individuals in each generation is restricted to nearby colonies.

The solution of this model is presented for one, two and three dimensional cases, with special reference to the correlation coefficient of gene frequencies between colonies. Also, the variance of gene frequencies between colonies is given for the three cases.

It has been shown that the decrease of genetic correlation with distance depends very much on the number of dimensions: In one dimension, the correlation decreases approximately exponentially with distance:

$$r\ (\rho)\ \propto e^{-A\rho}$$

where ρ is the distance and A is a constant which is equal to $\sqrt{2m_x/m_1}$. While in two dimensions, it falls off more rapidly and if the migration is isotropic in X and Y directions, we obtain, for a large value of ρ, the relation

$$r\ (\rho)\ \propto e^{-B\rho}/\sqrt{\rho}\ ,$$

where B is a constant which is equal to $\sqrt{4m_x/m_1}$. In three dimensions, it falls off still more rapidly and asymptotically, we obtain

$$r\ (\rho)\ \propto\ e^{-C\rho}/\rho$$

where $C = \sqrt{6m_x/m_1}$.

The quantity $1 - r(1)$, where $r(1)$ is the correlation coefficient between two adjacent colonies, is also pertinent in discussing the tendency toward random local differentiation. It has been shown that the relation of this quantity with mutation and migration rates depends very much on the number of dimensions. This, together with the above results on the decrease of correlation at a large distance, clearly indicates that the tendency toward random local differentiation is very much dependent on the number of dimensions; it is strongest in one dimension and becomes weaker as the number of dimension increases.

More general forms of migration and also some effects of selection are discussed.

LITERATURE CITED

CAVALLI-SFORZA, L. L., and F. CONTERIO, 1960   Analisi della fluttuazione di frequenze geniche nella popolazione della Val Parma. Atti Ass. Genet. Ital. Pavia **5**: 333–343.

KIMURA, M., 1953   "Stepping Stone" model of population. Ann. Rept. Nat. Inst. Genetics, Japan **3**: 62–63.

KIMURA, M., and J. F. CROW, 1963   The measurement of effective population number. Evolution **17**: 279–288.

MALÉCOT, G., 1948   *Les mathématiques de l'hérédité*. Masson, Paris. ——— 1955   Remarks on decrease of relationship with distance, following paper by M. KIMURA. Cold Spring Harbor Symp. Quant. Biol. **20**: 52–53. ——— 1959   Le modèles stochastiques en génétique de population. Publ. Inst. Statist. Univ. Paris **8**: 173–210.

MARADUDIN, A. A., E. W. MONTROLL, G. H. WEISS, R. HERMAN, and H. W. MILNES, 1960   Green's functions for monatomic simple cubic lattices. Mémoires, Académie Royal de Belgique, Classe des Sciences **14**, Fascicule 7.

WEISS, G. and M. KIMURA, 1964   A mathematical analysis of the stepping stone model of genetic correlation (in preparation).

WRIGHT, S., 1940   Breeding structure of populations in relation to speciation. Am. Naturalist **74**: 232–248. ——— 1943   Isolation by distance. Genetics **28**: 114–138. ——— 1946   Isolation by distance under diverse systems of mating. Genetics **31**: 39–59. ——— 1951   The genetical structure of populations. Ann. Eugenics **15**: 323–354.