Chapter 12

Migration

by Daniel Courgeau

Totally isolated human populations are very rare. In Chapter 16, we shall study several cases of human populations which have remained separated from other groups for some time (for example, the Kel Kummer Tuareg, or the Bedik tribesmen of Senegal). Apart from a few such extreme cases, however, most human and animal populations communicate with other populations of the same species by migration. The amount of migration that will take place, and the exact type of migration that occurs, will depend on geographical, sociological and economic factors.

If a population receives a number of migrants from other communities, its genetic structure may become altered. This will happen if the frequencies of the various alleles at a locus are different in the group of immigrants from those in the host population. Thus migration is a cause of changes in the genetic constitution of populations. In this chapter, we shall study such genetic changes due to migration.

First of all, we shall assume that all other conditions for Hardy-Weinberg equilibrium are satisfied, and we will consider the case of a number of groups, each of infinite size, which exchange members by migration. Such a model is clearly unrealistic; we shall therefore next study finite populations and shall establish "stochastic models", which enable us to study the variance in allele frequencies between groups when there is migration. Thirdly, we shall try to show which of the assumptions used in these models of migration need to be altered in order to agree more closely with reality; this will be done by considering some studies of human migration.

1. Deterministic Models with Migration

Consider an infinitely large population divided into m groups, each of "infinite" size. Suppose that in each generation migration between groups occurs; the proportion of members of the k-th group who are

migrants from the r-th group, in generation g, will be written as $l_{kr}^{(g)}$. We are assuming that all migrations take place in the period between birth and the start of the reproductive period; during the reproductive period itself, we assume that each group is panmictic. The proportion of individuals in the r-th group who are native to that group will be written as $l_{rr}^{(g)}$; hence we have:

$$\sum_r l_{kr}^{(g)} = 1.$$

It will be useful to write the set of values of $l_{kr}^{(g)}$ as a matrix:

$$L_g = (l_{kr}^{(g)}),$$

L_g is a square stochastic matrix of order m; the first subscript k represents the row number, and the second subscript r is the column number.

The matrix L_g represents the exchanges of migrants between all the m groups in the g-th generation. Clearly, L_g will change in each generation, since the numbers of migrants exchanged between particular populations will vary; however, in order to simplify matters, we shall assume from now onwards that the terms l_{kr} remain constant during the period in question, so that the exchanges between groups can be characterised by a single matrix L.

1.1. Changes in Genic Structure

Consider a locus with n alleles $A_1, \ldots, A_i, \ldots, A_n$. We shall write $s_k^{(g)}$ for the genic structure of the k-th group in generation g, i.e.:

$$s_k^{(g)} = (p_{k1}^{(g)}, \ldots, p_{ki}^{(g)}, \ldots, p_{kn}^{(g)})$$

where the p_{ki} are the frequencies of the various alleles in the group.

The genic structures of the m groups are therefore represented by m vectors, each with n elements. To simplify the notation, we shall write Ω_g for the matrix of order $m \times n$ whose rows are the vectors $s_k^{(g)}$, i.e.:

$$\Omega_g = (p_{ki}^{(g)}), \quad \text{with} \quad \sum_i p_{ki}^{(g)} = 1.$$

Assuming that the allele frequencies are the same in the migrants from a group as in the whole of the group they came from, we obtain:

$$p_{ki}^{(g)} = \sum_r l_{kr} p_{ri}^{(g-1)}$$

which can be expressed simply, in matrix notation, as:

$$\Omega_g = L\Omega_{g-1}.$$

Going back to the initial generation, we therefore have:

$$\Omega_g = L^g \Omega_0 . \tag{1}$$

Thus, if the matrix L^g tends towards a limit, as $g \to \infty$, the genic structure will also tend towards a limit. In order to study this limit, it is convenient to express the matrix L^g as a function of the eigenvalues and eigenvectors of L. It is shown in Appendix B that, since L is a stochastic matrix, all its eigenvalues are less than or equal to 1 in absolute magnitude, and that at least one of them is equal to 1. Also, if all the elements of the principal diagonal are non-zero, the eigenvalue $\lambda = 1$ is the only eigenvalue whose absolute value is equal to 1. This condition is, in general, satisfied for human populations: it is equivalent to saying that there are no communities which, in every generation, send all their members to other groups, so that none remain in the original group.

The matrix L can then be written in the form:

$$L = USU^{-1}.$$

If we assume that the eigenvalues of L are distinct [1], then the matrix S is a diagonal matrix whose non-zero elements are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ of L (see Appendix B):

$$S \equiv \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & \lambda_m \end{bmatrix}.$$

Using the results for stochastic matrices developed in Appendix B, we see that the product $U S^g U^{-1}$ $(= L^g)$ tends towards a matrix in which all the rows are identical, and equal to the column eigenvector associated with the eigenvalue unity.

The product of this matrix with the matrix Ω_0 will therefore give a matrix Ω in which all the rows are identical. This shows that in the limit, the genic structures of all the groups will become the same. This limiting genic structure, s, depends on the initial structure Ω_0 of the set of populations, and on the migration matrix L.

The speed with which the difference $\Omega_g - \Omega$ tends to zero depends only on the eigenvalue with the largest modulus, other than $\lambda = 1$.

1.1.1. Some particular cases. The simplest case is that of two groups, of which only one receives migrants from the other; the migration matrix

[1] The result derived here does not depend on this assumption, but the proof for the general case is too long to give here.

in this case is:

$$L=\begin{bmatrix}1-m & m\\ 0 & 1\end{bmatrix}$$

where m is the proportion of individuals in group 1 who came from group 2. It is easy to see that:

$$L^g=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}(1-m)^g & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}$$
$$=\begin{bmatrix}(1-m)^g & 1-(1-m)^g\\ 0 & 1\end{bmatrix}.$$

As $g\to\infty$, $L^g\to\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$, and the genic structure of the first group, which is the only one of the two structures which will change, tends towards that of the second group. The genic structure of the first group will be essentially equal to that of the second group after the passage of g generations, where g is of the order of $1/m$.

In the slightly more complex case of reciprocal exchanges, the migration matrix is of the form:

$$L=\begin{bmatrix}1-m_1 & m_1\\ m_2 & 1-m_2\end{bmatrix}.$$

This gives the equilibrium:

$$\underset{g\to\infty}{L^g}=\begin{bmatrix}\dfrac{m_2}{m_1+m_2} & \dfrac{m_1}{m_1+m_2}\\ \dfrac{m_2}{m_1+m_2} & \dfrac{m_1}{m_1+m_2}\end{bmatrix}. \tag{2}$$

If the initial genic structure of the set of two populations was:

$$\Omega_0=\begin{bmatrix}p_1 & 1-p_1\\ p_2 & 1-p_2\end{bmatrix}$$

the limit of the genic structure of the population would be:

$$\Omega=\begin{bmatrix}\dfrac{p_1m_2+p_2m_1}{m_1+m_2} & 1-\dfrac{p_1m_2+p_2m_1}{m_1+m_2}\\ \dfrac{p_1m_2+p_2m_1}{m_1+m_2} & 1-\dfrac{p_1m_2+p_2m_1}{m_1+m_2}\end{bmatrix}.$$

It is obvious that this structure depends both on the initial structure and on the migration matrix.

1.2. Changes in Genotypic Structure

The genotypic structures of the different groups will tend, like the genic structures, towards a common limit, as $g \to \infty$.

However, until the homogenisation of the populations is complete, the populations will have different genotypic structures. This state will last for a period of time which is greater the lower the migration rates. The heterogeneity which remains at any time can be measured by the deviation of the actual genotypic structure S of the whole population, from the theoretical structure of the population in Hardy-Weinberg equilibrium, with the same genic structure as the population in question, i.e. the corresponding Hardy-Weinberg structure.

Suppose that the size of the k-th group (which is assumed to be very large) is N_k; we shall write $N = \sum_k N_k$. Then the genotypic structure of the set of groups can be written as:

$$S = \frac{\sum_k N_k S_k}{N} = \frac{\sum_k N_k \vec{s}_k^2}{N}.$$

Hence, the overall genic structure is:

$$s = \frac{\sum_k N_k s_k}{N}$$

and so the corresponding Hardy-Weinberg structure is:

$$S_P = \vec{s}^2.$$

The deviation of the actual genotypic structure from the corresponding Hardy-Weinberg structure is therefore:

$$S - S_P = \frac{\sum_k N_k \vec{s}_k^2}{N} - \vec{s}^2 = \frac{\sum_k N_k (\overrightarrow{s_k - s})^2}{N}. \tag{3}$$

The i-th diagonal element of this trimat is:

$$\sum_k \frac{N_k (p_{ik} - p_i)^2}{N} = V_{ii}$$

where V_{ii} is the variance in gene frequency of the i-th allele between groups. The element in the i-th row and the j-th column is:

$$\frac{2 \sum N_k (p_{ik} - p_i)(p_{jk} - p_j)}{N} = 2 V_{ij}$$

where V_{ij} is the covariance in frequency of the i-th and j-th alleles. The elements of the trimat $S-S_P$ will thus all be equal to zero at the equilibrium state, when all the variances and covariances are zero.

1.2.1. The two allele case. In the particular case when there are only two alleles, the difference between the actual genotypic structure and the Hardy-Weinberg structure can be characterised by a single coefficient. This can be shown as follows. Write

$$s=(p, q) \quad \text{with} \quad p+q=1.$$

Then we have:

$$\frac{\sum_k N_k(p_k-p)^2}{N}=\frac{\sum_k N_k(q_k-q)^2}{N}=V$$

$$\frac{2\sum_k N_k p_k q_k}{N}-2pq=-2V.$$

We can therefore write:

$$S-S_P=\begin{vmatrix} V & \\ -2V & V \end{vmatrix}=V\begin{vmatrix} 1 & \\ -2 & 1 \end{vmatrix}.$$

If we define f by the relation:

$$V=fpq$$

we can write:

$$S=S_P+f\begin{vmatrix} pq & \\ -2pq & pq \end{vmatrix}=S_P+f\begin{vmatrix} p & \\ 0 & q \end{vmatrix}-f\begin{vmatrix} p^2 & \\ 2pq & q^2 \end{vmatrix}$$

or

$$S=(1-f)S_P+fS_H \tag{4}$$

where S_H is the corresponding homozygous structure of the population with the same gene frequencies as the population in question.

This corresponds to the classical formulae of Wahlund (1928) and Wright (1943). If we compare Eq. (4) above with Eq. (24b) of Chapter 8, it is clear that they are formally identical, and that f here plays the same role as the coefficient of kinship α in the earlier equation. However, in the general case, with any number of alleles, we do not have a single coefficient which measures the inbreeding effect of subdivision of a population; when there are more than two alleles, the trimat $S-S_P$ depends on more than one coefficient. It is also important to notice the difference between α and f. The coefficient α is a probability, while f is the ratio of the variance in frequency between the groups to the product pq of the mean frequency over all the groups.

1.3. Applications to Actual Populations

This model can be applied in a number of ways, according to the kind of data that is available. Very few genic structures of different human populations are known for more than short periods of time. However, a number of cases are known where several sub-populations have been formed by migration from a large population and have colonised new areas. Assuming that the genic structure of the population of origin has remained unchanged, and that the groups of migrants were representative of the population they came from, we know the initial structures of the new sub-populations; Ω_0 is thus assumed to be the same as the present-day structure of the population from which the new populations originated. If we know the number of generations g which have elapsed since the colonisation occurred, and the present genic structure Ω_g of a colony, we can deduce the migration matrix L, using the relation:

$$\Omega_g = L^g \Omega_0 .$$

1.3.1. The black and white populations of the United States. This method has been used by Glass (1955), Glass and Li (1953) and Roberts and Hiorns (1962) to estimate the rates of migration between the black and white populations of the U.S.A.

Glass and Li (1953) give data for 4 loci (14 alleles). They assume that the gene frequencies in the white population have remained constant, and that 10 generations have passed since the two populations first came into contact. This corresponds to the case we gave above where:

$$L = \begin{bmatrix} 1-m & m \\ 0 & 1 \end{bmatrix}.$$

These authors obtained estimates of m which were all of the same order of magnitude:

$$0.028 \leqq m \leqq 0.056.$$

If the present migration rate is maintained (mean $m=0.036$) the equilibrium frequency of the R^0 gene (rhesus blood group cDe), for example, would not be attained until 60.7 generations had passed; this corresponds to 1 670 years, if we take 27.5 years as the generation time in man. Also a migration rate of 0.03 would mean that about 30% of the genes of the present "black" population must have originated from white ancestors.

Roberts and Hiorns (1962) basing their calculations on more up-to-date views about the original populations from which the black population was derived, obtained estimates of m between 0.02 and 0.025.

1.3.2. The Sudanese Nilotes. In this case, data are available on intermarriages between different tribes. Knowing the genic structure at any time, we can use these to find the structure at a given time in the past or future, on the assumption that the migration rates are constant in time.

The data show that marriages between members of different groups are in general rare[2]. For example, Roberts and Hiorns (1962) give the following matrix of frequencies of marriage between three Sudanese tribes, the Nuer, Dinka and Shilluk:

$$L=\begin{bmatrix}0.9850 & 0.0125 & 0.0025\\ 0.0138 & 0.9775 & 0.0087\\ 0.0000 & 0.0098 & 0.9902\end{bmatrix}.$$

With these migration rates, if the genic structures of these populations are different, it would take a large number of generations for these populations to reach an equilibrium. It is because of low migration rates between groups that we can talk of human "races". The results of this section show that, if the population sizes of the different races are effectively infinite, genetic differences between the races must steadily decrease, and eventually will disappear altogether. In Section 2, however, we shall see that in the case when the sizes of the groups are finite, heterogeneity between the groups may be maintained by random fluctuations in allele frequencies.

1.4. Deterministic Models of Migration when Other Forces for Change are Acting

In the following sections, we shall continue to assume that the subgroups are all of infinite size, and that the matrix of migration rates does not change with time.

1.4.1. The continuous-time model of migration[3]. If we assume that migration, reproduction and death occur continuously in time, as is reasonable for a human population, we can write the genetic structure of a set of m populations as the matrix $\Omega(t)$, which is a function of the continuous variable t. Let us define a matrix M such that in a small interval of time dt we have:

$$\Omega(t+dt)=\Omega(t)+M\Omega(t)\,dt+\text{terms of order } (dt^2)$$

[2] These rates are, however, much higher than mutation rates, which, as we saw in Chapter 11, are of the order of 10^{-5}.

[3] This model has been studied by Roberts and Hiorns (1962); they made the implicit assumption that individuals are capable of reproducing as soon as they are born. Courgeau (1971) has studied the case when there is a lag between birth and the beginning of reproductive life.

so that:

$$\frac{d\Omega(t)}{dt} = M\,\Omega(t). \tag{5}$$

Note that M is not a stochastic matrix, but that it must satisfy the relation $\sum_j m_{ij} = 0$. The matrix $M+I$, where I is the unit matrix, is therefore a stochastic matrix.

If we assume that all the eigenvalues μ_j of M are distinct, as we have done previously, it follows from the theory of differential equations that Eq. (5) has the solution:

$$\Omega(t) = U\,S^t\,U^{-1}\,\Omega(0)$$

where S is a diagonal matrix whose j-th term is equal to e^{μ_j}.

As before, if S^t tends to a limit as $t \to \infty$, then the genic structure of the population tends towards a limiting structure. We have seen that the matrix $M+I$ is a stochastic matrix; it follows that all its eigenvalues are of absolute magnitude less than or equal to 1. Now, the eigenvalues of this matrix are equal to $1+\mu_j$, where μ_j are the eigenvalues of M. It therefore follows that:

$$|1+\mu_j| \leqq 1 \qquad \text{for all values of } j.$$

If μ_j is real, we therefore have $\mu_j \leqq 0$, hence $e^{\mu_j t}$ will tend towards 0 (if $\mu_j < 0$) or 1 (if $\mu_j = 0$), as $t \to \infty$.

If μ_j is a complex number, we can write $\mu_j = r_j(\cos\Theta + i\sin\Theta)$. The preceding condition gives $r_j \cos\Theta \leqq 0$. Hence:

$$e^{\mu_j t} = e^{r_j t\cos\Theta}\, e^{r_j i t \sin\Theta}.$$

If $\cos\Theta < 0$, the first term on the right hand side of this equation tends to 0 as $t \to \infty$; the second term can be written

$$\cos(r_j t \sin\Theta) + i\sin(r_j t \sin\Theta)$$

and this is always finite. In this case, therefore, $e^{\mu_j t} \to 0$ as $t \to \infty$. On the other hand, if $\cos\Theta = 0$, we have $r_j = 0$, hence $e^{\mu_j t} = 1$.

Thus, as $t \to \infty$, S^t tends towards a particular matrix, and $\Omega(t)$ has a limit. Therefore, as $t \to \infty$, the genic structures of the different groups become identical, as in the discrete-time case.

If there are just two groups, and the migration matrix is:

$$M = \begin{bmatrix} -m_1 & m_1 \\ m_2 & -m_2 \end{bmatrix}$$

we obtain:

$$S = \begin{bmatrix} 1 & 0 \\ 0 & e^{-(m_1+m_2)} \end{bmatrix}.$$

Hence:

$$\Omega(t)=\begin{bmatrix}1 & -m_1\\ 1 & m_2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & e^{-t(m_1+m_2)}\end{bmatrix}\begin{bmatrix}\dfrac{m_2}{m_1+m_2} & \dfrac{m_1}{m_1+m_2}\\ -\dfrac{1}{m_1+m_2} & \dfrac{1}{m_1+m_2}\end{bmatrix}\Omega(0)$$

$$=\begin{bmatrix}\dfrac{m_2+m_1e^{-t(m_1+m_2)}}{m_1+m_2} & \dfrac{m_1(1-e^{-t(m_1+m_2)})}{m_1+m_2}\\ \dfrac{m_2(1-e^{-t(m_1+m_2)})}{m_1+m_2} & \dfrac{m_1+m_2e^{-t(m_1+m_2)}}{m_1+m_2}\end{bmatrix}\Omega(0).$$

As $t\to\infty$, therefore:

$$\Omega(t)\to\begin{bmatrix}\dfrac{m_2}{m_1+m_2} & \dfrac{m_1}{m_1+m_2}\\ \dfrac{m_2}{m_1+m_2} & \dfrac{m_1}{m_1+m_2}\end{bmatrix}\Omega(0).$$

This continuous-time model of migration is of interest when the population is studied for only a small number of years. However, the study of the limits shows that these are identical in both the continuous and discontinuous cases, as can be seen by referring back to Eq. (2); the only difference is in the rate at which the population approaches the equilibrium state.

In the discontinuous case, the rate of convergence towards the equilibrium can be obtained from:

$$\Omega_g=\begin{bmatrix}1 & -m_1\\ 1 & m_2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \{1-(m_1-m_2)\}^g\end{bmatrix}\begin{bmatrix}\dfrac{m_2}{m_1+m_2} & \dfrac{m_1}{m_1+m_2}\\ \dfrac{-1}{m_1+m_2} & \dfrac{1}{m_1+m_2}\end{bmatrix}\Omega_0$$

$$=\begin{bmatrix}\dfrac{m_2+m_1\{1-(m_1+m_2)\}^g}{m_1+m_2} & \dfrac{m_2(1-\{1-(m_1+m_2)\}^g)}{m_1+m_2}\\ \dfrac{m_2(1-\{1-(m_1+m_2)\}^g)}{m_1+m_2} & \dfrac{m_1+m_2\{1-(m_1+m_2)\}^g}{m_1+m_2}\end{bmatrix}\Omega_0.$$

Thus when the migration is discontinuous in time, $(\Omega_g-\Omega_\infty)$ tends to zero as $(1-a)^g$, where $a=m_1+m_2$. In the continuous-time case, however, this difference tends towards zero as e^{-ag}, when the unit of time is the generation time. The discontinuous model therefore predicts a more rapid approach to the limit than the continuous one. The difference is greater the larger a is, i.e. the greater the amount of migration which takes place.

1.4.2. Migration and mutation. We shall now examine the case when the genes are subject to mutation. The probability (or rate) of a mutation from one allele to another is assumed to be the same for all the groups of populations. Let v_{xy} be the probability of mutation from allele A_x to A_y (so that $v_{xx}=1-\sum_{y\neq x} v_{xy}$).

The frequency $p_{ki}^{(g)*}$ of the i-th allele in the k-th group before mutation, but after migration, is given by:

$$p_{ki}^{(g)*}=\sum_r l_{kr}\, p_{ri}^{(g-1)}.$$

After mutation, this frequency will have changed to:

$$\begin{aligned} p_{ki}^{(g)} &= p_{ki}^{(g)*} - \sum_{y\neq i} v_{iy}\, p_{ki}^{(g)*} + \sum_{x\neq i} v_{xi}\, p_{kx}^{(g)*} \\ &= p_{ki}^{(g)*}\left[1-\sum_{y\neq i} v_{iy}\right] + \sum_{x\neq i} v_{xi}\, p_{kx}^{(g)*} \\ &= \sum_x v_{xi}\, p_{kx}^{(g)*}. \end{aligned}$$

Writing V for the stochastic matrix of order $n\times n$, whose terms are the v_{xy}, this relation can be written:

$$\Omega_g = L\Omega_{g-1}V = L^g\Omega_0 V^g.$$

We have seen that when the eigenvalues of the matrix L are all distinct, L^g tends towards a matrix Ω whose rows are all identical. With the same conditions, V^g tends towards a matrix Y whose rows are all identical. Therefore, under these conditions, the product $L^g\Omega_0 V^g$ also tends towards a matrix whose rows are all identical, and it is easy to show that this matrix is Y. In this case, therefore, the limiting structure is independent of the initial structure of the population and also of the migration matrix L. It depends only on the matrix of mutation frequencies.

Malécot (1948) has treated the particular case of two alleles. In this case, writing u for the mutation rate v_{12} and v for v_{21}, the matrix Y is equal to:

$$Y=\begin{bmatrix} \dfrac{v}{u+v} & \dfrac{u}{u+v} \\[2ex] \dfrac{v}{u+v} & \dfrac{u}{u+v} \end{bmatrix}.$$

Thus, for example, the frequency

$$p_{k1}^{(g)} \to \frac{v}{u+v} \quad \text{as } g\to\infty.$$

The speed with which the gene frequencies $p_{k1}^{(g)}$ tend towards their limits depends only on $(1-u-v)^g$. The limiting value is therefore essentially reached when g is greater than approximately $\frac{1}{u+v}$. Now, we have seen that mutation rates are generally of the order of 10^{-5}. Therefore the time necessary for the gene frequencies to be equalised is very high, of the order of 10^5 generations.

However, migration rates are usually much higher than mutation rates. Hence the set of populations will at first behave as if mutation did not occur, and will tend towards an equilibrium structure Ω, which is a function of the initial structure and of the migration matrix L, as we showed above. Later, this structure will slowly move towards the equilibrium structure Y under mutation, which depends only on the matrix of mutation rates. In the first stage, we would find inbreeding effects due to the splitting-up of the population, but in the second stage Hardy-Weinberg frequencies would be satisfied. It is, however, important to realise that these conclusions are based on a quite unrealistic model; in particular, we have had to assume that the mutation and migration rates remain constant throughout all these stages.

2. Stochastic Models with Migration

We shall now consider a finite population divided into m groups of sizes N_k between which migration can occur. First of all, we shall assume that all the other conditions for Hardy-Weinberg equilibrium are satisfied. We shall study the limit to which the distribution of gene frequencies in the different groups tends, and the values of the first two moments of this distribution; these provide a measure of the differences between the groups, at equilibrium[4]. We shall then study some cases in which some of the other conditions for Hardy-Weinberg equilibrium are dropped.

In order to define the conditions under which $2N$ "successful gametes" are drawn in generation g, we have to make several assumptions. First, we assume that the number of gametes produced by members of generation $g-1$ is very large. From this assumption, it follows that the gametes which will go to form the new individuals of generation g are drawn from an essentially infinite pool of gametes, among which the genes have the proportions $p_{ki}^{(g-1)}$. We also have to assume that the sample of these gametes which will go to form males has an identical composition with the sample of gametes which will form the females of

[4] See footnote to p. 364.

generation g. We are thus implicitly assuming that the sampling is with replacement.

With these assumptions, the number of successful gametes $2N_k p_{ki}^{(g)}$ which carry the i-th allele is a binomial variate; the probability that it takes the value n_{ki} (which is an integer such that $0 \leqq n_{ki} \leqq 2N_k$) is:

$$P(n_{ki}) = \binom{2N_k}{n_{ki}} \{p_{ki}^{(g-1)}\}^{n_{ki}} \{1 - p_{ki}^{(g-1)}\}^{2N_k - n_{ki}}.$$

As we saw in Chapter 8 (Eq. (35)) the mean and variance of $p_{ki}^{(g)}$, given the value of $p_{ki}^{(g-1)}$, are:

$$E_{g-1}\{p_{ki}^{(g)}\} = p_{ki}^{(g-1)},$$

$$V_{g-1}\{p_{ki}^{(g)}\} = \frac{p_{ki}^{(g-1)}\{1 - p_{ki}^{(g-1)}\}}{2N_k}.$$

We shall now examine how the introduction of migration modifies these equations.

2.1. Migration

We shall use the same notation as in Section 1 for the migration matrix L, which is assumed to be constant from generation to generation.

Also, consideration of Eq. (1) of Section 1.1 shows that we can consider the changes in the frequency of one allele in isolation from the others. The frequencies of the i-th allele in the m groups constitute a column vector:

$$p_i^{(g)} \equiv \begin{bmatrix} p_{1i}^{(g)} \\ p_{2i}^{(g)} \\ \vdots \\ p_{mi}^{(g)} \end{bmatrix}$$

We shall leave out the index i in what follows.

Finally, we assume that migration takes place before the start of the reproductive period, and that individuals remain in the same group throughout their reproductive life. This is in particular the case when we consider the case of matings between individuals who belong to different groups, e.g. different races or tribes, in man.

2.1.1. The expectations of gene frequencies. Since migration is assumed to occur before reproduction, the genic structure from which the gametes which will go to form generation g are drawn is:

$$p^{(g)*} = Lp^{(g-1)}.$$

After the random draw, the genic structure obtained can be written:

$$p^{(g)} = Lp^{(g-1)} + E^{(g)}$$

where $E^{(g)}$ is a column vector whose k-th element $e_k^{(g)}$ represents the chance variation in the frequency of the allele in question, in the k-th group. The expectation and variance of $e_k^{(g)}$, given the genic structure in generation $g-1$ are:

$$E_{g-1}\{e_k^{(g)}\}=0,$$
$$V_{g-1}\{e_k^{(g)}\}=\frac{p_k^{(g)*}\{1-p_k^{(g)*}\}}{2N_k}.$$

The expectation of $p^{(g)}$ is therefore:

$$E\{p^{(g)}\}=E[E_{g-1}\{p^{(g)}\}]=LE\{p^{(g-1)}\}.$$

If the genic structure of the initial generation is assumed to be known, the solution of this equation gives:

$$E\{p^{(g)}\}=L^g p^{(0)}.$$

Thus the expectation is the same as for the deterministic case treated in Section 1. As $g\to\infty$, $E\{p^{(g)}\}\to p$, which is a vector whose elements are all identical. This vector depends in general on the migration matrix and on the genic structures of the initial sub-populations.

In the particular case when one group sends migrants to all the other groups, but does not itself receive migrants from other groups, this result shows that the expected frequency of the allele in question will become the same as the initial frequency of the "colonising" group.

2.1.2. The second moments[5]. We shall write the covariances and variance between the different groups as:

$$u_{jk}^{(g)}=\mathrm{Cov}\{p_i^{(g)}, p_k^{(g)}\}=u_{kj}^{(g)},$$
$$u_{jj}^{(g)}=V\{p_j^{(g)}\}.$$

We shall also need the conditional variance and covariances, given the genic structure of generation $g-1$, which we shall denote by $u_{jk}^{(g|g-1)}$.

We want to find a difference equation for the $u_{jk}^{(g)}$ in terms of the $u_{jk}^{(g-1)}$, and to see whether this leads to a limit for the set of $u_{jk}^{(g)}$, as $g\to\infty$.

First let us examine the conditional covariances (and variances). If the gene frequencies in generation $g-1$ are assumed to be known, we

[5] The variances and covariances are found as follows. We suppose that we can make repeated "trials", each time starting the ensemble of populations in the same initial state, and that we observe the outcomes in a particular population (for the variances) or pair of populations (for the covariances). The variances and covariances are calculated over the different "trials". When the set of populations reaches a steady state (in cases where this can occur), these variances and covariances are equivalent to the variances and covariances between the different populations in the set.

obtain:

$$u_{jk}^{(g|g-1)} = \{\sum_z l_{jz}[p_z^{(g-1)} - E\{p_z^{(g-1)}\}]\} \times \{\sum_w l_{kw}[p_w^{(g-1)} - E\{p_w^{(g-1)}\}]\} + E_{g-1}(e_j e_k).$$

Assuming that the gametes in the j-th and k-th "pools" are sampled independently, we have:

$$\text{if } j \neq k\text{: } E_{g-1}(e_j e_k) = 0,$$

$$\text{if } j = k\text{: } E_{g-1}(e_j e_k) = \frac{p_j^{(g)*}\{1 - p_j^{(g)*}\}}{2N_j}.$$

The expectation of the conditional covariance gives us the *a priori* covariance:

$$u_{jk}^{(g)} = \sum_z \sum_w l_{jz} l_{kw} u_{zw}^{(g-1)} + \delta_{jk} \frac{E\{p_j^{(g)*}\} - E[\{p_j^{(g)*}\}^2]}{2N_j}$$

where $\delta_{jk} = 0$ if $j \neq k$ and $\delta_{jj} = 1$. Now we know that:

$$E\{p_j^{(g)*}\} = \sum_k l_{jk} E\{p_k^{(g-1)}\}$$

$$E[\{p_j^{(g)*}\}^2] = [E\{p_j^{(g)*}\}]^2 + E[(p_j^{(g)*} - E\{p_j^{(g)*}\})^2]$$

$$= [\sum_k l_{jk} E\{p_k^{(g-1)}\}]^2 + \sum_z \sum_w l_{jz} l_{kw} u_{zw}^{(g-1)}.$$

Thus the *a priori* covariances in generation g can be expressed in terms of the covariances and expectations in generation $g-1$.

We saw in the last section that $E\{p_j^{(g)}\}$ tends towards a limit p: this limit will be effectively reached when g exceeds a certain number which is a function solely of the eigenvalue of L with the largest modulus other than 1. We shall assume in what follows that this limit has been reached. Then the equation for $u_{jk}^{(g)}$ given above takes the simplified form:

$$u_{jk}^{(g)} = \sum_z \sum_w l_{jz} l_{kw} \left(1 - \frac{\delta_{jk}}{2N_j}\right) u_{zw}^{(g-1)} + \delta_{jk} \frac{p - p^2}{2N_j}.$$

We can write the covariances and expectations as column vectors with m^2 elements[6]:

[6] We could write them as vectors with $\frac{m(m-1)}{2}$ elements, since $u_{jk} = u_{kj}$, but the form we have used is easier to handle in the formation of the matrix A, which we use below; for doing calculations, however, the vectors with the lower number of elements would be preferable.

$$\mathscr{V}_g = \begin{bmatrix} u_{11}^{(g)} \\ u_{12}^{(g)} \\ \vdots \\ u_{1m}^{(g)} \\ u_{21}^{(g)} \\ \vdots \\ u_{2m}^{(g)} \\ u_{31}^{(g)} \\ \vdots \\ u_{mm}^{(g)} \end{bmatrix} \qquad \mathscr{R} \equiv \begin{bmatrix} \frac{p-p^2}{2N_1} \\ 0 \\ \vdots \\ 0 \\ \frac{p-p^2}{2N_2} \\ \vdots \\ 0 \\ \frac{p-p^2}{2N_3} \\ \vdots \\ 0 \end{bmatrix}$$

We shall also introduce the matrix A of order m^2 whose elements are:

$$a_{rs} = l_{jz}\, l_{kw} \left(1 - \frac{\delta_{jk}}{2N_j}\right)$$

(where the two subscripts j and k determine the row number r, and the pair of subscripts z and w determine the column number s).

In this notation, the relation between the variances in generations g and $g-1$ becomes:

$$\mathscr{V}_g = A\mathscr{V}_{g-1} + \mathscr{R}. \tag{6}$$

If we again assume that $E\{p_j^{(g)}\}$ is effectively equal to its limiting value from generation g onwards, we have:

$$\mathscr{V}_{g+n} = A^{n+1}\mathscr{V}_{g-1} + \sum_{j=0}^{n} A^j \mathscr{R}. \tag{7}$$

For rows for which $j \neq k$, the elements of A sum to 1; when $j=k$, the sum of these elements is $1 - \frac{1}{2N_j}$. Now it is known that for a non-singular matrix $A=(a_{ij})$ whose elements are all greater than zero, the following property holds:

$$\text{minimum } s_i \leqq \lambda \leqq \text{maximum } s_i$$

$$\text{where } s_i = \sum_{j=1}^{n} a_{ij}$$

and λ is the eigenvalue with the largest modulus. Furthermore, the two relations can only be equalities when all the "row-sums" $s_1, \ldots, s_n$ are equal [7], and we have seen above that this is not the case for the matrix A which we are considering. It follows that all the eigenvalues of A are less than 1 in value.

If we assume, as before, that the eigenvalues of the matrix A are all distinct [8], A can be written as:

$$A = US^{-1}U^{-1}$$

where S is a diagonal matrix whose non-zero elements are the eigenvalues of A.

Since we have just shown that all the eigenvalues of A are less than 1 in absolute value, it follows that, as $g \to \infty$, A^g (i.e. US^gU^{-1}) tends towards a matrix all of whose elements are zero. Eq. (7) thus tends towards the expression:

$$\mathscr{V}_{g+n} = \sum_{j=0}^{n} A^j \mathscr{R}.$$

To determine $\mathscr{V}_{g+n}$, it thus remains to determine the limit of

$$\sum_{j=0}^{g} A^j = U\left(\sum_{j=0}^{g} S^j\right) U^{-1}$$

as $g \to \infty$.

The matrix $\sum_{j=0}^{g} S^j$ consists of terms of the form:

$$1 + \lambda + \lambda^2 + \cdots + \lambda^g$$

where λ is an eigenvalue of A. This is a geometric series. Since $|\lambda| < 1$, the series tends, as $g \to \infty$, towards the limit $\frac{1}{1-\lambda}$. The matrix $\sum_{j=0}^{g} A^j$ therefore tends to a finite limit:

$$A^* \equiv \begin{bmatrix} \frac{1}{1-\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{1-\lambda_2} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots\cdots\cdots & & \frac{1}{1-\lambda_{m^2}} \end{bmatrix}$$

[7] For a proof of these results, see Gantmacher (1959), Vol. II, p. 63, remark 2.

[8] The result remains valid when this assumption is dropped, but the proof is long and tedious.

and $\mathscr{V}_{g+n}$ tends towards the limit:

$$\mathscr{V} = A^* \mathscr{R}.$$

$\mathscr{V}$ therefore depends on the set of eigenvalues of A, and on the vector $\mathscr{R}$. Another method of finding $\mathscr{V}$ is to write the equilibrium condition $\mathscr{V}_g = \mathscr{V}_{g-1} = \mathscr{V}$, which, from Eq. (6), gives $\mathscr{V} = A\mathscr{V} + \mathscr{R}$. This matrix equation can be solved to give $\mathscr{V}$.

It is also possible to show that the speed with which the variances approach their asymptotic values is a function both of the speed with which the mean approaches its limit (which is a function of the largest eigenvalue of L that is less than 1), and the speed with which $\sum_{j=0}^{g} A^j \mathscr{R}$ tends towards its limit. This is a function of the largest eigenvalue of A.

The following case is of interest. Suppose that the k-th group is infinite in size and that migrants from this group go to all the other groups, but that group k itself receives no migrants. We have seen that the expected genic structures of the other groups tend towards that of the k-th group. Let us now examine whether the variance also tends towards a limit. Since the "colonising" group k is of infinite size, $\frac{p-p^2}{2N_k} = 0$; also, since the k-th group does not receive any migrants, we have $V_{kk}^{(g)} = 0$ and $V_{kj}^{(g)} = 0$. The relation between the second moments in generations g and $g-1$ can therefore be written, excluding this k-th group:

$$\mathscr{V}_g' = A' \mathscr{V}_{g-1}' + \mathscr{R}'$$

where the vectors $\mathscr{V}'$ and $\mathscr{R}'$ are of order $m^2 - m = m(m-1)$ and the square matrix A' is of order $m(m-1)$. Furthermore, since the k-th group sends migrants to every other group, we have:

$$\sum_{s=m+1}^{m^2} a_{rs} < 1.$$

Therefore the vector $\mathscr{V}_{g+n}'$ tends towards some constant vector, as $n \to \infty$.

Consider the case when there are just two groups, one of which is infinite; the other group is of size N, of which a proportion m are migrants from the first case. This is a particular case of Wright's (1943) "island model". It is easy to prove that the variance u_g of the group of size N is not zero, since it must satisfy the relation:

$$u_g = (1-m)^2 \left(1 - \frac{1}{2N}\right) u_{g-1} + \frac{p(1-p)}{2N}$$

where p is the frequency of the allele in question in the infinite population. The limit u of u_g as g tends to infinity satisfies the relation:

$$u\left[1-(1-m)^2\left(1-\frac{1}{2N}\right)\right]=\frac{p(1-p)}{2N}.$$

Hence:

$$u=\frac{p(1-p)}{1+4Nm-2m-2Nm^2+m^2}.$$

When m is small, we can neglect terms in m, Nm^2 and m^2, and then:

$$u=\frac{p(1-p)}{1+4Nm}. \tag{8}$$

Now let us find the rate at which u_g will approach u. We have seen that this depends on the value of the eigenvalue of L which has the largest modulus less than 1; in this case, this is m. The quantity $E\{p^{(g)}\}-p$ can be considered as essentially equal to zero when g is greater than approximately $1/m$. Assuming, therefore, that this quantity is equal to zero, we can write:

$$\begin{aligned}u_{g+n}-u&=(1-m)^2\left(1-\frac{1}{2N}\right)(u_{g+n-1}-u)\\&=(1-m)^{2n}\left(1-\frac{1}{2N}\right)^n(u_g-u)\\&\approx\left(1-2m-\frac{1}{2N}\right)^n(u_g-u).\end{aligned}$$

This shows that when n exceeds the order of magnitude of $2m+\frac{1}{2N}$, the equilibrium state will have been reached. Thus this state is reached after approximately $\frac{1}{m}+\frac{2N}{1+4Nm}$ generations. Notice that this is the maximum number of generations that must elapse before equilibrium is attained.

Bodmer and Cavalli-Sforza (1968) studied the two-allele case when the gene frequencies remain close to $\frac{1}{2}$. Using the transformation: $p_k^{(g)}=\sin^2\Theta_k^{(g)}$, where Θ is in radians, they obtained the following expression for the conditional variance:

$$V_{g-1}\{\Theta_k^{(g)}\}=\frac{1}{8N_k}+O\left(\frac{1}{N_k}\right)^2 \tag{9}$$

which is independent of the gene frequencies in generation g.

In this case, it is not necessary to assume that the limit of $E\{p_k^{(g)}\}$ has been attained, and we can study the change in the variance simply from the first generation onwards. If we write:

$$\mathscr{R}' \equiv \begin{bmatrix} \frac{1}{8N_1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{8N_2} \\ \vdots \\ 0 \end{bmatrix}$$

we have:

$$\mathscr{V}_g = A^g \mathscr{V}_0 + \sum_{j=0}^{g} A^j \mathscr{R}' = \sum_{j=0}^{g} A^j \mathscr{R}'.$$

The results obtained by Bodmer and Cavalli-Sforza for the case when there are groups which send out migrants to other groups, but which do not themselves receive migrants, show that the approximation expressed in Eq. (9) remains satisfactory so long as the gene frequencies are not close to 0 or 1. The variance increases rapidly with time, and reaches its equilibrium value after a number of generations which is of the order of $1/\alpha$, where α is the smallest proportion of migrants from another group that any group receives.

2.1.3. Alternative methods of investigating the effects of migration. Another way of studying the second moments of the distribution of gene frequencies (which is due to Malécot, 1948) consists of considering the coefficient of kinship between individuals in two sub-populations, k and j; this coefficient of kinship Φ_{kj} is defined as the probability that a gene taken at random from the gamete pool in group k is identical with a gene taken at random from the gamete pool in group j. It can be shown that the *a priori* variance and covariances between groups can be expressed in terms of the coefficients of kinship, as follows.

We introduce the random variables X_{ky}; these take the value 0 or 1 according to whether the y-th gamete in the k-th colony (out of the $2N_k$ gametes in this colony) is the allele in question or not. We can then write:

$$p_k^{(g)} = \frac{\sum_y X_{ky}^{(g)}}{2N_k}, \qquad p_j^{(g)} = \frac{\sum_z X_{jz}^{(g)}}{2N_j}.$$

We know that once equilibrium is reached $E\{p_k^{(g)}\}=E\{p_j^{(g)}\}=p$. The covariance in gene frequencies between colonies is found by the formula:

$$\text{Cov}\{p_k^{(g)}, p_j^{(g)}\}=\frac{1}{4N_kN_j}\sum_{yz}E[(X_{ky}^{(g)}-p)(X_{jz}^{(g)}-p)].$$

We have to consider two possible cases; either two gametes drawn one from the k-th colony and one from the j-th colony are identical by descent – the probability of this is $\Phi_{kj}^{(g)}$, hence we can write, remembering that $X_{ky}^2=X_{ky}$:

$$E\{(X_{ky}^{(g)}-p)^2\}=E\{(X_{ky}^{(g)})^2\}-p^2=p-p^2.$$

The other possibility is that the two genes drawn are not identical; this has the probability $1-\Phi_{kj}^{(g)}$, and in this case we have the following relation:

$$E\{(X_{ky}^{(g)}-p)(X_{jz}^{(g)}-p)\}=0$$

because the two draws are independent. We therefore have:

$$\text{Cov}\{p_k^{(g)}, p_j^{(g)}\}=\frac{1}{4N_kN_j}\times 4N_kN_j(p-p^2)\Phi_{kj}^{(g)}=(p-p^2)\Phi_{kj}^{(g)}.$$

In a similar way, we find that:

$$V\{p_k^{(g)}\}=(p-p^2)\Phi_{kk}^{(g)}+\frac{p-p^2}{2N_k}(1-\Phi_{kk}^{(g)}).$$

If $\frac{1}{2N_k}$ can be neglected in comparison with $\Phi_{kk}^{(g)}$, we therefore obtain:

$$V\{p_k^{(g)}\}=(p-p^2)\Phi_{kk}^{(g)}.$$

These expressions enable us to study the moments of order 2 in terms of the coefficients of kinship.

2.2. Stochastic Models of Migration with Other Evolutionary Forces also Acting

In this section we shall study the effects of migration together with mutation and linearised selection pressure. We shall consider only the stationary state which the population will reach, and we shall assume that the dispersion of gene frequencies around the point of equilibrium which they would reach under selection alone is small, so that selection can be linearised.

2.2.1. Migration and mutation. We now introduce the possibility of mutations, whose frequencies are given by values v_{xy}, as in Section 1.4.2.

We assume that migration occurs first, and that then mutation occurs at the time of production of an infinite number of gametes by the members of the group; finally the "successful gametes", which will go to form the members of the next generation, are assumed to be drawn at random. We will have to consider all the alleles at the locus in question, and cannot consider just one of them, in isolation; we shall therefore use the matrix Ω of gene-frequencies again.

We have seen that, before the drawing of the successful gametes:

$$p_{ki}^{(g)*} = \sum_{xr} v_{xi}\, l_{kr}\, p_{rx}^{(g-1)}$$

where p_{ki} is the frequency of the i-th allele in the k-th group. Consequently:

$$E(\Omega_g) = LE(\Omega_{g-1})\, V = L^g\, \Omega_0\, V^g.$$

Thus the change in the expected gene frequency is the same as for the deterministic model: as $g \to \infty$, $E(\Omega_g)$ tends towards a matrix Y whose rows are all identical. Let the expected frequency of allele A_x be p_x. The covariances and variances will be written:

$$\mathrm{Cov}\,(p_{ki}, p_{lj}) = u_{klij} \quad \text{and} \quad V(p_{ki}) = u_{kkii}.$$

The corresponding conditional covariances and variances will be denoted by u^*_{klij} and u^*_{kkii}.

We can write the following expression for the conditional covariances:

$$u_{klij}^{*(g-1)} = \sum_{xr} v_{xi}\, l_{kr} (p_{rx}^{(g-1)} - p_x) \sum_{ys} v_{yj}\, l_{ls} (p_{sy}^{(g-1)} - p_y) + E_{g-1}(e_{ki}\, e_{lj}).$$

The *a priori* covariance is therefore:

$$u_{klij}^{(g)} = \sum_{xrys} v_{xi}\, v_{yj}\, l_{kr}\, l_{ls}\, u_{rsxy}^{(g-1)} - \frac{\delta_{klij}\, E\{p^*_{ki}\, p^*_{kj}\}}{2N_k} + \frac{\delta^*_{klij}\, E\{p^*_{ki}(1-p^*_{ki})\}}{2N_k}$$

$$= \sum_{xrys} v_{xi}\, v_{yj}\, l_{kr}\, l_{ls} \left(1 - \frac{\delta_{kl}}{2N_k}\right) u_{rsxy}^{(g-1)} - \delta_{klij} \frac{p_i\, p_j}{2N_k} + \delta^*_{klij} \frac{(p_i - p_i^2)}{2N_k}$$

where:

$$\begin{aligned}
\delta_{klij} &= 0, && \text{if } k \neq l \text{ or if } k = l \text{ and } i = j \\
\delta_{kkij} &= 1, && \text{if } i \neq j \\
\delta^*_{klij} &= 0, && \text{if } k \neq l \text{ or if } i \neq j \\
\delta^*_{kkii} &= 1 \\
\delta_{kl} &= 0, && \text{if } k \neq l \\
\delta_{kk} &= 1.
\end{aligned}$$

This can be written more simply in matrix notation if we assume that the means have reached their stationary state by generation g. The pair of indices k and l which refer to the sub-populations will be used to define the row number, and the pair of indices i and j, which denote the alleles in question, will be used to define the column number. We then define the following matrices:

A: a square matrix of order m^2 whose terms are of the form

$$l_{kr}\, l_{ls}\left(1-\frac{\delta_{kl}}{2N_k}\right).$$

W: a square matrix of order n^2, whose general term is $v_{xi}v_{yj}$.

U_g: a matrix of order $m^2 \times n^2$, whose general term is u_{rsxy}.

R: a matrix of order $m^2 \times n^2$ in which the only non-zero terms are r_{kkij}, and these are equal to:

$$-\frac{p_i p_j}{2N_k} \quad \text{if } i \neq j \qquad \text{and} \qquad \frac{p_i - p_i^2}{2N_k} \quad \text{if } i=j.$$

In this notation, the expression for the covariance above becomes:

$$U_g = AU_{g-1}\, W + R.$$

This gives:

$$U_{g+n} = A^n\, U_{g-1}\, W^n + \sum_{j=0}^{n} A^j R W^j.$$

Now we know that A has no eigenvalues equal to 1, so that $A^n \to 0$, and the limit of U_{g+n} as $n \to \infty$ is equal to the limit of the sum:

$$U = \sum_{j=0}^{\infty} A^j R W^j.$$

This limit can be found by solving the matrix equation:

$$U = AUW + R.$$

2.2.2. Migration, mutation and linearised selection (the two allele case). We make similar assumptions to those in the preceding section: migration is assumed to take place first, then mutation and linearised zygotic selection, and random sampling of the gametes is assumed to occur last of all. It is not difficult to obtain useful expressions for this type of model, for the case of a locus with just two alleles. We shall therefore drop the subscript for the allele in what follows, and will confine our attention to one allele.

If the frequency of the allele in question is $p_k^{(g)}$ in the k-th group, it will be changed, after migration has occurred, to:

$$p_{k1}^{(g)} = \sum_r l_{kr} p_r^{(g)}.$$

Mutation and linearised selection can be taken into account by a coefficient h of approach towards the equilibrium position, at which the value of p is $\hat{p}$. The gene frequency after this second stage will then be:

$$p_{k2}^{(g)} = p_{k1}^{(g)} - h(p_{k1}^{(g)} - \hat{p}).$$

Finally, the effect of chance during the random draw of the $2N_k$ gametes can be expressed as follows:

$$p_k^{(g+1)} = p_{k2}^{(g)} + e_k^{(g)}.$$

As before, we have

$$E_g\{e_k^{(g)}\} = 0, \qquad V_g\{e_k^{(g)}\} = \frac{p_{k2}^{(g)}\{1 - p_{k2}^{(g)}\}}{2N_k}.$$

The passage from one generation, g, to the next, $g+1$, can be expressed by the matrix equation:

$$p^{(g+1)} - p = (1-h)\, L\{p^{(g)} - p\} + e^{(g)}$$

where p is the column vector whose elements are all equal to $\hat{p}$.

It is simple to derive the following expression for the mean gene frequencies:

$$E\{p^{(g+1)} - p\} = (1-h)\, LE\{p^{(g)} - p\}.$$

This shows that $E\{p^{(g)} - p\}$ tends to zero, as $g \to \infty$; hence, in the limit, $E\{p^{(g)}\}$ tends towards p. The expectation of the gene frequency in a group therefore tends to a limit which is independent of the initial gene frequency in the group.

We shall now study the special case of a simple migration matrix, such that the groups are arranged in a linear order, and each is of constant size N and exchanges a fixed proportion m of migrants with the two neighbouring groups which lie on either side of it (Fig. 12.1). We assume that no other type of migration occurs. This model is called the "stepping-stone" model of migration.

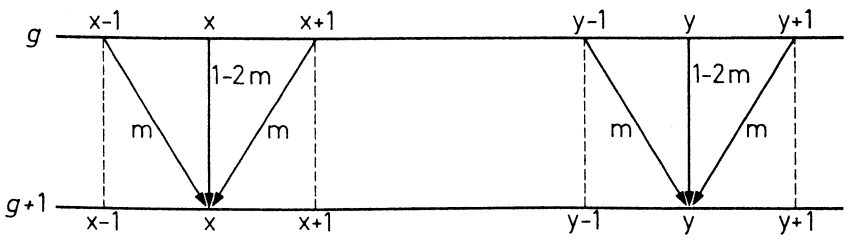

Fig. 12.1. Exchange of migrants between neighbouring groups, with constant migration coefficient

The asymptotic stationary state is defined by relations of the following form:

$$u_{kl}=(1-h)^2\{m^2(u_{k-1,l-1}+u_{k-1,l+1}+u_{k+1,l-1}+u_{k+1,l+1})+m(1-2m)(u_{k-1,l}+u_{k+1,l}+u_{k,l-1}+u_{k,l+1})+(1-2m)^2 u_{kl}\}\times\left(1-\frac{\delta_{kl}}{2N}\right)+\delta_{kl}\frac{(\hat{p}-\hat{p}^2)}{2N}.$$

It turns out that the covariance between colonies depends only on the distance between them. If we write d for the distance between two groups, k and l, and neglect terms in m^2, we have the following expression for the covariance between populations situated at a distance d apart:

$$\begin{aligned}\mathrm{Cov}(d)=(1-h)^2\{2m[\mathrm{Cov}(d+1)+\mathrm{Cov}(d-1)]+(1-4m)\,\mathrm{Cov}(d)\}\\+\delta_{kl}\frac{(\hat{p}-\hat{p}^2)}{2N}-\delta_{kl}\frac{(1-h)^2}{2N}[(1-4m)\,\mathrm{Cov}(0)+4m\,\mathrm{Cov}(1)].\end{aligned}\tag{10}$$

In the case when $d\neq 0$, $\mathrm{Cov}(d)$ is therefore defined by a second order difference equation; the solutions of this type of equation are known (see Appendix A) to be of the form μ^d, where μ^d satisfies the relation:

$$\mu^d=(1-h)^2\{2m\mu^{d+1}+2m\mu^{d-1}+(1-4m)\mu^d\}.$$

Neglecting terms in h^2 and in $h\,m$, this equation simplifies to:

$$\mu=(1-4m-2h)\mu+2m+2m\mu^2$$

which gives:

$$\mu^2-\left(2+\frac{h}{m}\right)\mu+1=0,$$

$$\mu=1+\frac{h}{2m}\pm\sqrt{\frac{h^2}{4m^2}+\frac{h}{m}}.$$

As $d\to\infty$, the solution $\mathrm{Cov}(d)$ remains bounded; it is therefore only necessary to consider the root which is less than 1. We then have:

$$\mathrm{Cov}(d)=\lambda\left\{1+\frac{h}{2m}-\sqrt{\frac{h^2}{4m^2}+\frac{h}{m}}\right\}^d\tag{11}$$

where λ is a constant which can be determined as a function of $\mathrm{Cov}(0)$, as follows.

From Eq. (10), we have the following relation for $d=1$:

$$\mathrm{Cov}(1)=(1-4m+2h)\,\mathrm{Cov}(1)+2m\,\mathrm{Cov}(0)+2m\,\mathrm{Cov}(2).$$

Replacing $\mathrm{Cov}(1)$ and $\mathrm{Cov}(2)$ by the values given by Eq. (11), we obtain:

$$2m\lambda\mu^2-(4m+2h)\lambda\mu+2m\,\mathrm{Cov}(0)=0.$$

Hence:

$$\lambda = \mathrm{Cov}(0).$$

Now let us calculate the variance in gene frequency between colonies, $V(=\mathrm{Cov}(0))$; this comes from Eq. (10):

$$V=(1-4m-2h)\,V+4m\,\mathrm{Cov}(1)+\frac{\hat{p}-\hat{p}^2-(1-4m-2h)\,V-4m\,\mathrm{Cov}(1)}{2N}.$$

Hence:

$$(4m+2h)\,V-4m\,\mathrm{Cov}(1)=\frac{\hat{p}-\hat{p}^2-V}{2N-1}$$

$$2(\sqrt{4mh+h^2})\,V=\frac{\hat{p}-\hat{p}^2-V}{2N-1}.$$

If N is large, $2N-1\approx 2N$, so we obtain, finally:

$$V=\frac{\hat{p}-\hat{p}^2}{1+4N\sqrt{4mh+h^2}}. \tag{12}$$

This enables us to find the general expression:

$$\mathrm{Cov}(d)=\frac{\hat{p}-\hat{p}^2}{1+4N\sqrt{4mh+h^2}}\left\{1+\frac{h}{2m}-\sqrt{\frac{h^2}{4m^2}+\frac{h}{m}}\right\}^d. \tag{13}$$

When the changes in gene frequency due to migration are high, compared with those due to mutation and selection, h can be neglected, in comparison to m, and Eq. (13) simplifies to:

$$\mathrm{Cov}(d)=\frac{\hat{p}-\hat{p}^2}{1+8N\sqrt{mh}}\left\{1-\sqrt{\frac{h}{m}}\right\}^d.$$

This can be approximated by the exponential formula:

$$\mathrm{Cov}(d)=\frac{\hat{p}-\hat{p}^2}{1+8N\sqrt{mh}}\,e^{-d\sqrt{h/m}}. \tag{14}$$

This shows that the correlation coefficient $\frac{\mathrm{Cov}(d)}{V}$ decreases exponentially with distance, and also that the correlation is independent of N, the size of the sub-populations.

These results are due to Malécot (1966).

2.2.3. Migration and mutation in a spatially continuous population. A rather different type of model of the interaction of stochastic factors and migration has been studied by Wright (1943, 1946) and Malécot

(1948, 1969). In this model, the population is assumed to be continuously distributed in space, over an infinite line or plane. Although the total population size is infinite in this model, as in the discontinuous "stepping-stone" model discussed above, differences in gene frequency between different points on the line or plane can arise, because the parents of an individual born at a given point are themselves likely to therefore have been born nearby, and they therefore have a finite chance of being related. Wright has reviewed his approach in his book (Wright, 1969), so we shall here confine ourselves to a consideration of Malécot's method, for the simple case of a linear continuum.

It is convenient to treat this problem in terms of the coefficient of kinship, rather than in terms of the variances and covariances of gene frequencies. As we saw in Section 2.1.3, results obtained in terms of the coefficient of kinship can be translated into variance terms. We shall consider how to determine the coefficient of kinship, $\Phi(x)$, between two individuals who were born at α and β, which are a distance x apart (see Fig. 12.2). The limit, $\Phi(0)$, of $\Phi(x)$ as x tends to zero is the inbreeding coefficient of an individual in such a population, since two individuals born a distance 0 apart are, in fact, the same individual.

In the case of migration along a continuous line, we have to characterise the migration process by a continuous function, the "migration distribution", which specifies the probability density $g(y)$ that a parent of a given individual was born at a distance y from the place where the individual himself was born (we shall arbitrarily assign a negative sign to displacements in the left-hand direction in Fig. 12.2, and a positive sign to displacements to the right, with respect to the birth-place of the individual in question). We shall assume that the probability density $g(y)$ is independent of time and also of the position of the individual in question, i.e. that the migration is homogeneous in both space and time. Clearly, we have:

$$\int_{-\infty}^{\infty} g(y)\,dy = 1.$$

We will also make the simplifying assumption that the number of individuals per unit length, ρ, is independent of position. The number of individuals in an element of length dx is thus $\rho\,dx$.

Finally, we shall assume that the probability that a given allele mutates to some other allele is the same for all alleles, i.e. that $\sum_{j \neq i} v_{ij}$ (see Section 2.1 of Chapter 11) is the same for all i. We shall denote this probability by v.[9]

[9] Note that it follows from this assumption that the mean frequencies of all the alleles, at equilibrium, are the same.

Referring to Fig. 12.2, we want to calculate the coefficient of kinship $\Phi(x)$ between individuals I and J, who were born a distance x apart. We shall assume that the population has reached a steady state with respect to the distribution of allele frequencies, so that $\Phi(x)$ remains constant from generation to generation. Let us consider the relations between two genes, one drawn at random from I and one from J. Suppose that the first comes from the parent I′ of I, and the second from the parent J′ of J. The probability that I′ was born in a neighbourhood of length dy around a point γ, a distance y from α is $g(y)\,dy$; the probability that J′ was born in a neighbourhood of length dz around δ, a distance z from β is $g(z)\,dz$. The probability that I′ and J′ both come from a neighbourhood of length dz around the point δ, a distance z from β (and

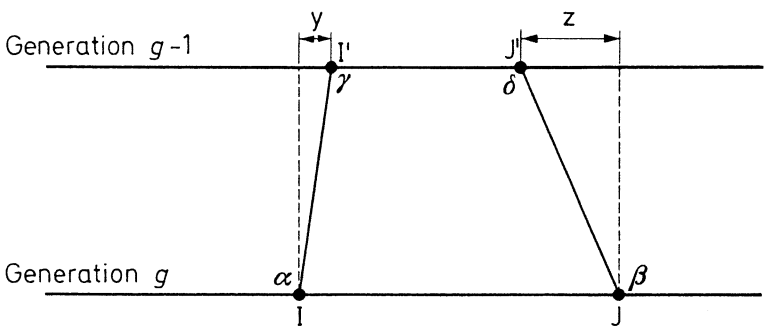

Fig. 12.2. Distances between parents and offspring in a spatially continuous population

therefore, from Fig. 12.2, a distance $x+z$ from α) is $g(z)\,g(x+z)\,dz$. In this case, there is a probability of $\dfrac{1}{\rho\,dz}$ that I′ and J′ are the same individual. If this is the case, the probability that the two genes are identical by descent is $\dfrac{1+\Phi(0)}{2}$, given that neither of them has mutated.

If the two genes are descended from different individuals in generation $g-1$, who were born a distance $(x-y+z)$ apart, their probability of identity by descent is $\Phi(x-y+z)$.

Taking all the possibilities into account, and noting that the chance that neither gene has mutated is $(1-v)^2$, we obtain:

$$\Phi(x)=(1-v)^2\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi(x-y+z)\,g(y)\,g(z)\,dy\,dz + \frac{1}{\rho}\int_{-\infty}^{\infty}\left[\frac{1+\Phi(0)}{2}-\Phi(0)\right]g(x+z)\,g(z)\,dz\right\}. \tag{15}$$

$\Phi(0)$ is subtracted from $\frac{1+\Phi(0)}{2}$ in the right-hand integral, to correct for the fact that the left-hand integral should include no contribution from the cases when the two genes come from the same individual.

This integral equation can be transformed into a linear differential equation with constant coefficients by replacing $\Phi(x-y+z)$ in the double integral above by its Taylor series:

$$\Phi(x-y+z)=\Phi(x)+(z-y)\Phi'(x)+\frac{(z-y)^2}{2!}\Phi''(x)+\cdots.$$

If we neglect terms in v^2, and those containing the differential coefficients $\Phi'(x)$, $\Phi''(x)$ etc. multiplied by v, we get:

$$\begin{aligned}2v\Phi(x)-\Phi'(x)&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(z-y)\,g(y)\,g(z)\,dz\\&-\frac{\Phi''(x)}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(z-y)^2\,g(y)\,g(z)\,dy\,dz+\cdots\\&=\frac{1-\Phi(0)}{2\rho}\int_{-\infty}^{\infty}g(x+z)\,g(z)\,dz.\end{aligned}$$

The left-hand side of this equation includes the moments of the dispersal distribution. If we assume that this distribution is symmetrical, so that odd moments are equal to zero, then, neglecting moments of higher order than 2, we obtain the expression:

$$2v\Phi(x)-\sigma^2\Phi''(x)=\frac{1-\Phi(0)}{2\rho}\int_{-\infty}^{\infty}g(x+z)\,g(z)\,dz. \tag{16}$$

Since $g(x+z)$ tends towards zero with increasing x, it must be negligible for large x. For large x, therefore, $\Phi(x)$ is given by the solution of the differential equation:

$$\Phi''(x)=\frac{2v}{\sigma^2}\Phi(x)$$

which gives the exponential form for $\Phi(x)$:

$$\Phi(x)\propto e^{-\sqrt{2v}|x|/\sigma}. \tag{17}$$

To determine $\Phi(0)$, the inbreeding coefficient, we require the solution of the fundamental Eq. (16). It can be shown that this equation has the solution:

$$\Phi(x)=\frac{1-\Phi(0)}{4\pi\rho}\int_{-\infty}^{\infty}\frac{G^2(t)}{2v+\sigma^2 t}e^{-itx}\,dt$$

where $G(t)$ is the Fourier transform of $g(x)$:

$$G(t)=\int_{-\infty}^{\infty} e^{itx} g(x)\,dx$$

(with t a real number), and $i=\sqrt{-1}$.

For the case of a normal dispersal distribution such that:

$$g(x)=\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

we have:

$$G(t)=e^{-\frac{\sigma^2 t^2}{2}}$$

so that:

$$\Phi(0)=\frac{1}{1+4\sigma\rho\sqrt{2v}}. \tag{18}$$

The even moments of order 2 and higher of a normal distribution are all powers of σ. It is therefore reasonable to neglect moments of higher order than 2, provided that the unit of distance with which we are working (i.e. one over which there is a significant decline in genetic relationship) is large compared with the standard deviation of the migration distribution.

These results can be compared with those obtained with the very similar discontinuous model of migration which we studied in Section 2.2.2. If we assume that only two alleles are present at the locus in question, it follows from the discussion of Section 2.1.3, together with the assumption of equal mutation rates for all the alleles (so that $p=1-p=\frac{1}{2}$), that $\frac{\Phi(0)}{4}$ measures the variance in gene frequency between regions, and $\frac{\Phi(x)}{4}$ measures the correlation coefficient of gene frequency between populations separated by a distance x. Eq. (18) is clearly analogous to the corresponding expression of Section 2.2.2, Eq. (12), if we equate $2m$ with σ^2 (which we can do if we take the distance between adjacent groups in the model of Section 2.2.2 as unity). Eq. (17) is analogous to Eq. (13).

A similar model can be set up for the case of dispersion over a two-dimensional plane. If migration is assumed to follow a normal distribution, with the same standard deviation, σ, in all directions, then the following expressions for $\Phi(0)$ and $\Phi(x)$ are obtained:

$$\Phi(0)=\frac{1}{1-8\pi\rho\sigma^2(1/\log 2v)}$$

and when the distance x is large:

$$\Phi(x) \propto \frac{1}{\sqrt{x}} e^{-\sqrt{2v}|x|/\sigma}.$$

The effect of migration is clearly very sensitive to dimension: $\Phi(0)$ is smaller in the two-dimensional case than the one-dimensional case, and $\Phi(x)$ falls off more quickly.

3. Data on Migration in Human Populations

In all the models which we have discussed so far we assumed that the matrix of all the migration rates between different populations are known. These were assumed to remain constant during the whole period of approach to the stationary state. We also made further assumptions during the development of several of the models; in particular we had to assume that all migration occurred before the reproductive period.

We shall now compare these hypotheses with some observational data on human migration, from various countries.

We shall first discuss some of the many models that have been proposed which would afford a precise description of migration and its evolutionary consequences.

3.1. Models of the Migration Process

First, we need to define the term "migration" more clearly. As population geneticists, we are obviously not interested in short-term movements of individuals, but only in permanent migrations from one population to another. In particular, we are concerned with "matrimonial migration", which can be measured in one of two ways.

1. By comparing the places of birth of individuals with the places where their offspring are born (we would therefore have to take the migrations of both parents into account).
2. By comparing the birthplaces of men and their wives.

3.1.1. The principal types of model. Many studies have shown that the number Y_{ab} of migrants from one community a to another b depends on the distance between the two communities, and on the size of both of them. The most commonly used model gives Y in terms of these parameters, according to the equation:

$$Y_{ab} = k \frac{\delta(a)\,\delta(b)}{r^{\alpha}} ds(a)\, ds(b)$$

where $\delta(x)$ is the density of the population at x, r is the distance between the two places, k and α are constants and $ds(a)$ and $ds(b)$ are the areas of the two places in question. This is called the Pareto model.

Stouffer (1940) has tried to substitute a measure of "social distance" for the measure of geographic distance in this equation; social distance is defined in terms of the total number of migrants from community a who are found in all the intervening communities between a and b. It is obviously very highly correlated with geographic distance, and only empirical data can tell us which model best fits a given population.

Hägerstrand (1957) considered a different model of migration. He assumed that the migration rate at any time is closely correlated with the rates at earlier periods. One of the parameters in this model is therefore the number of migrants from a to b, n years before the time we are considering; another parameter, which implicitly introduces the distance between the two communities, depends on the number (assumed to be small) of individuals who move to b but are not attracted there by earlier migrants.

Other, more complex, models introduce a variety of socio-economic variables. Olsson (1965a, b), for example, showed that the geographic distance that a migrant moves depended on eight parameters which could be used to characterise the migrants. The distance moved thus sums up information of many different sorts about the migrant.

Models of matrimonial migration are usually analogous to the first two types of model.

3.1.2. Some applications to real populations. The Pareto model has been applied to population data from several countries (including France, America, Sweden and Japan), and generally gives satisfactory agreement with observed migration data. The coefficient α, which was intitially thought to be a constant, has, however, proved to vary from one region to another, and also in time.

For Swedish populations, for example, Hägerstrand (1957) found values of α from 0.4 to 3.3; the large values mostly corresponded to migration in rural communities, and the small values to towns. α was generally found to decrease with time.

Courgeau (1969) studied migration in French populations, between 1896 and 1962. He found that α could be taken as equal to 2 for the whole of this period, by introducing a correction term l, which is a function of time. The expression for Y then becomes:

$$Y_{ab} = \delta(a)\,\delta(b)\left(\frac{k}{r^2}+1\right) ds(a)\,ds(b)$$

which gives good agreement with observation.

The finding that migration rates change with time means that it is not strictly valid to apply the genetic theory we have developed above to human populations, since this theory was based on the assumption of constant migration rates.

A difficulty of the Pareto model is that it is not applicable over the whole of the interval $(0, \infty)$, since the number of non-migrants ($r=0$) is undefined. This can be overcome, by introducing a constant β, and writing the equation for Y as:

$$Y_{ab}=k\frac{\delta(a)\,\delta(b)}{(\beta+r)^{\alpha}}\,ds(a)\,ds(b).$$

In order to find the frequency distribution of migration starting at a point a, we must know the form of the territory in which the migration occurs, and the population density at all points. Two particular cases give simple results. First, we can assume that the territory is unidimensional (e.g. a valley), and that the density is constant. The probability distribution is then of the form:

$$f(r)=\frac{(\alpha-1)\,\beta^{\alpha-1}}{(\beta+r)^{\alpha}} \quad \text{for } r\geqq 0.$$

Alternatively, we could consider the territory to be an infinite surface, and assume constant density δ. We then have the following probability distribution:

$$f(r)=\frac{(\alpha-1)(\alpha-2)\,\beta^{\alpha-2}\,r}{(\beta+r)^{\alpha}}.$$

Cavalli Sforza (1962) used a model similar to this one, in a study of matrimonial migration in a valley in Parma:

$$f(r)=\frac{k^{\alpha}}{\Gamma(\alpha)}\,e^{-kr}\,r^{\alpha-1}$$

where

$$\Gamma(\alpha)=\int_0^{\infty} e^{-x}\,x^{\alpha-1}\,dx.$$

This model using the Γ function of Pearson is easier to fit to empirical data than the one given above.

The Pareto model has been compared with Stouffer's model, for American and Swedish population data, and both proved almost equally satisfactory. However, Hägerstrand's model was shown, in a Swedish population, to be greatly superior to the other two. It is also capable, in a modified form, of fitting data on migration in French populations, and their changes between 1896 and 1962. Unfortunately, this model has not yet been applied to population genetics.

3.2. Comparison of the Genetic Models with the Models of Migration

The models of migration in human populations and of the genetic consequences of migration were developed independently. We would therefore like to know whether the assumptions on which the population genetic theory is based agree with the observations of anthropologists and demographers.

3.2.1. Migration independent of time. In the theoretical sections of this chapter, we generally assumed that migration rates were constant in time. This assumption is not necessary for the study of deterministic models, since it is possible to write L_g for the migration matrix in generation g, and we then have:

$$\Omega_g = L_g \Omega_{g-1} = L_g L_{g-1} \ldots L_1 \Omega_0 .$$

If migration rates are constant, or only change gradually, this equation allows us to study the changes in the genetic composition of a population, and even to predict future changes.

However, when we dealt with the stochastic model of migration we assumed that the population had reached its stationary state. If migration rates are changing in time, this assumption is clearly invalid.

Studies in Sweden have shown that migration rates were effectively constant from 1785 to 1870, but have since changed considerably. This also appears to be the case in France, since 1896. Thus it is probable that European populations were in a stationary state before 1870, and could have been studied using our stochastic model, but that these populations are at present changing, and are far from the new stationary state which they may one day attain.

There may, of course, be populations which have not undergone this process of change, so that we could use the stochastic model. It is, however, important to establish over a long period, that the population is really stationary. Cavalli Sforza (1962), for example, states that migration rates in the province of Parma have remained the same for three centuries.

These considerations show that it would be desirable to have a treatment, not only of the stationary state, but also of the changes in the variance before this state is reached. Bodmer and Cavalli Sforza (1968) have studied the changes in variance from generation to generation, with various migration matrices. However, they did not consider the possibility of the migration matrices changing with time.

3.2.2. The nature of the migration matrices. Most genetical models of migration involve matrices of migration rates. Non-genetic studies of migration, however, consider migration as a function of distance. We

have described one genetic model which has this property, and others have also been developed by Wright and Malécot. However, the probability distributions for migration to various distances which these models use (usually the normal distribution) do not correspond to reality at all. It would be desirable to study models like this, but incorporating migration distributions which agree better with observations on migration, and to see what changes this introduces.

Another important point is that most genetic models of migration assume constant population density over the whole of the area in question. This is of course quite unlike the situation for human populations, which vary greatly in density, in particular if we compare country areas and towns.

Finally, the probability distribution for migration is usually assumed to be the same for all points of origin, whereas studies on real populations show that there can be large differences between different populations.

For these reasons, it would seem that there is as yet no satisfactory treatment of the genetic consequences of migration which is continuous in space.

However, as a first attempt to study this problem, we can use the matrix model. We can divide the population into a larger or smaller number of sub-populations and calculate the rates of migration between the sub-populations, whose sizes and gene frequencies are known. Then we can predict the future genetic state of the population, provided that the migration matrix can be assumed to be changing slowly.

3.2.3. Stability of marriages. In a discrete-generation treatment of migration, we assume that mates remain together, once a couple has formed. This is not the case in human populations. In order to allow more than one migration during a lifetime, we would have to make a model with overlapping generations. Many studies have been done of the probability that an individual who has migrated once will do so again during the next n years. It is often assumed that second migrations follow the same frequency distribution of migration distances as the first ones; it would be desirable to see whether this is a reasonable approximation to the real situation.

Although this factor could thus be introduced into the genetic theory, it seems probable *a priori* that the effect would be slight.

4. Conclusions

Observations on real populations show that several of the assumptions behind our theories of the genetic effects of migration are of doubtful validity, when applied to human populations. The stochastic

models developed by Malécot (1948, 1969) are possibly applicable to populations of the snail *Cepaea nemoralis* (Lamotte, 1951), but cannot validly be used for man, since migration rates in man change considerably with time. Since we do not know the limit towards which human migration rates are moving, we can only predict future changes in the short or medium term. For further understanding, we need a theory of the changes during the approach to equilibrium, as well as of the stationary state. This will require precise knowledge of the rates of matrimonial migration between groups. Sutter (1958) and Sutter and Thanh (1962) give data which show that these are very similar to the rates of migration between groups, so that we can probably use this type of data in genetic studies.

It should perhaps also be mentioned that our assumption, in Section 2.2 that mutation rates and selection coefficients are the same in all groups, may not be valid. Malécot (1966) has developed a model in which these values depend on the group in question. Under this model, we can no longer estimate the migration rates from the variance between groups, unless we also know the mutation rates and selection parameters for all the groups; conversely if the migration rates are known, we could estimate the parameter which characterises the mutation rate and selection coefficient. However, this model has only been developed for the stationary state.

Finally, it is important to remember that all the models we have discussed assumed that migration occurred between the gamete pools (assumed to be of infinite size) of the different groups, and not at the level of individuals. In this way, we could consider the frequency of a gamete in a group after migration as deterministic, and the only random stage was the stage of drawing the useful gametes.

In fact, of course, migration takes place because individuals move. This introduces another random process into the system, so it is only the expected frequency of a gene in the group who migrate which is equal to the frequency in the group they come from.

If we also consider the number of migrants exchanged by two groups to be random variable, instead of deterministic, we have a third random stage. It is very difficult to study such a model, with several random stages, except by "Monte Carlo" methods. These will be described in the next chapter.

Further Reading

Alström, C. H.: First-cousin marriages in Sweden 1750–1844 and a study of the population movement in some Swedish populations from the genetic-statistical point of view. Acta Genet. (Basel) **8**, 295–369 (1958).

Azevedo, E., Morton, N.E., Miki, C., Yee, S.: Distance and kinship in Northeastern Brazil. Amer. J. Hum. Genet. **21**, 1-22 (1969).

Cavalli-Sforza, L.L.: Genetic drift in an Italian population. Sci. Amer. **221**, 30-37 (1969).

Cavalli-Sforza, L.L., Zonta, L.A., Nuzzo, F., Bernini, L., Jong, W.W.W. de, Meera Khan, P., Ray, A.K., Went, L.N., Siniscalco, M., Nijenhuis, L.E., Loghem, E. van, Modiano, G.: Studies on African pygmies. I. A pilot investigation of Babinga pygmies in the Central African Republic (with an analysis of genetic distances). Amer. J. Hum. Genet. **21**, 252-274 (1969).

Courgeau, D.: Mutations, migrations et structures géniques. Population **24**, 935-940 (1969).

Haldane, J.B.S.: The theory of a cline. J. Genet. **48**, 277-284 (1948).

Hanson, W.D.: Effects of partial isolation (distance), migration, and different fitness requirements among environmental pockets upon steady-state gene frequencies. Biometrics **22**, 453-468 (1966).

Hiorns, R.W., Harrison, G.A., Boyce, A.J., Kuchemann, C.F.: A mathematical analysis of the effects of movement on the relatedness between populations. Ann. Hum. Genet. **32**, 237-250 (1969).

Imaizumi, Y.: Variation of inbreeding coefficient in Japan. Hum. Heredity **21**, 216-230 (1971).

Imaizumi, Y., Morton, N.E., Harris, D.E.: Isolation by distance in an artificial population. Genetics **66**, 569-582 (1970).

Katz, A., Hill, R.: Residential propinquity and marital selection: a review of theory, method and fact. In: Les deplacements humains, Entretiens de Monaco en sciences humaines. Paris: Hachette 1962.

Kimura, M., Weiss, G.H.: The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49**, 461-576 (1964).

Morton, N.E., Harris, D.E., Yee, S., Lew, R.: Pingelap and Mokil atolls: migration. Amer. J. Hum. Genet. **23**, 339-349 (1971).

Roberts, D.F.: Genetic fitness in a colonising human population. Hum. Biol. **40**, 494-507 (1968).

Saldanha, P.H.: Gene flow from white into negro populations in Brazil. Amer. J. Hum. Genet. **9**, 299-309 (1957).

Smith, C.A.B.: Local fluctuations in gene frequencies. Ann. Hum. Genet. **32**, 251-260 (1969).

Ward, R.H., Neel, J.V.: Gene frequencies and microdifferentiation among the Makiritare Indians. IV. A comparison of a genetic network with ethnohistory and migration; a new index of genetic isolation. Amer. J. Hum. Genet. **22**, 538-561 (1970).

Workman, P.L., Niswander, J.D.: Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. Amer. J. Hum. Genet. **22**, 24-49 (1970).

Wright, S.: The genetical structure of populations. Ann. Eugen. (Lond.) **15**, 325-354 (1951).

Yanase, T.: A note on the patterns of migration in isolated populations. Jap. J. Hum. Genet. **9**, 136-152 (1964).

Zipf, G.K.: The P_1P_2/D hypothesis: on the intercity movement of persons. Amer. Sociol. Rev. **11**, 677-686 (1946).