The Dynamics of Racial Intermixture

D. F. ROBERTS AND R. W. HIORNS

Department of Human Anatomy, University of Oxford,
and
Electronics Computing Laboratory, University of Leeds, England

IN THE VERY CONSIDERABLE BODY of mathematical theory relating to population genetics, developed initially by Fisher (e.g., 1930), Wright (e.g., 1931) and Haldane (e.g., 1924-32), and latterly extended by Malecot (1948), Crow (1958), Kimura (1956), Owen (1959) and others, the more recent studies have tended to be concentrated on the effects of selection and random processes on gene frequencies. Intermixture, that evolutionary process which is bringing about some of the most strikingly different gene combinations in human populations today, has received rather less attention, although by 1931 Wright had established the basic formulas relating to the rate of change of frequency of a gene in a population under immigration, and the distribution of gene frequencies in an array of intermixing populations, both where intermixture is the sole process modifying frequencies and where it occurs concomitantly with one or more other such processes. In anthropological problems theory and application have been elementary. Several attempts have been made to examine from demographic data the effect of observed admixture in restricting random differentiation of gene frequencies among local populations within larger ones, applying Wright's models of "island" population structure (e.g., Lasker, 1952; Roberts, 1956) and of isolation by distance (Roberts, 1956; Alström, 1958). A second type of problem attempted refers to specific cases of hybridization between two populations where observed gene frequency data from parental and hybrid groups have been used to calculate from Bernstein's simple formula the amount of intermixture that had occurred (e.g., Boyd, 1939; Ottensooser, 1944; da Silva, 1948, 1949), while Stevens (1952) examined some of the attendant statistical problems. Glass and Li (1953) pointed out that from the same data, the average rate of gene flow per generation may be calculated, if the number of generations of intermixture is known; they gave a formula for one-way gene flow as exemplified by the American Negro, in which, whereas the white parental population had contributed to the gene pool of the hybrid, the latter had not contributed to the white; further studies of the Negro in America followed (Roberts, 1955; Glass, 1955; Steinberg et al., 1960; Saldanha, 1957). Probably more important in human evolution is the situation in which two or more populations have exchanged genes through occasional intermarriage, each thereby affecting the other's gene pool and being itself affected; the equations and procedure relating to this situation are set out here since they are likely to be of relevance in other studies, e.g., in those aiming to measure the extent of selection or drift in new populations arising from intermixture in a new environment.

Received December 21, 1961.

MODEL I: TWO POPULATIONS, DISCRETE GENERATIONS

This model assumes that generations do not overlap, that migration occurs at the end of each generation, that populations are panmictic, that migrants are drawn at random from them and that intermixture is the sole process modifying gene frequencies.

In two populations, let q_o^I be the original allele frequency at a single locus in population I, q_o^{II} that in population II; let the subscript refer to the generation so that q_n^I is the gene frequency in population I after n generations. Let $m^I(n)$ be the rate of gene flow into population II from population I, and $m^{II}(n)$ that into population I from population II, after the nth generation. Thus the gene flow rate is the proportion of the *recipient* population's genes received by immigration in the given generation; what proportion of the donating population emigrates is irrelevant for present purposes. The gene flow rates are hereinafter referred to as admixture rates. They need not be constant from generation to generation.

In the gametes of population I after one generation of such admixture, there will be a component, m^{II}, deriving from population II, in which the allele frequency will be q_o^{II}; in the remainder, $1 - m^{II}$, deriving from population I, the gene frequencies will be unaltered by loss of the emigrants; the frequency will be in the total population $q_1^I = m^{II}q_o^{II} + (1 - m^{II})\, q_o^I$.

After n generations,

$$q_n^I = [1 - m^{II}(n)]\, q_{n-1}^I + m^{II}(n)\, q_{n-1}^{II} \tag{1}$$

$$q_n^{II} = m^I(n)\, q_{n-1}^I + [1 - m^I(n)]\, q_{n-1}^{II} \tag{2}$$

Suppose that

$$\Delta_n = q_n^I - q_n^{II} \text{ and } \Sigma_n = q_n^I + q_n^{II}$$

then subtracting (1) − (2):

$$\Delta_n = [1 - m^I(n) - m^{II}(n)]\, \Delta_{n-1} \tag{3}$$

which is a recurrence relation implying

$$\Delta_{n-1} = [1 - m^I(n-1) - m^{II}(n-1)]\, \Delta_{n-2},$$

$$\Delta_{n-2} = [1 - m^I(n-2) - m^{II}(n-2)]\, \Delta_{n-3},$$

$$\vdots$$

$$\Delta_1 = [1 - m^I(1) - m^{II}(1)]\, \Delta_o$$

and by back substitution of each equation in the one immediately above it,

$$\Delta_n = [1 - m^I(1) - m^{II}(1)]\, [1 - m^I(2) - m^{II}(2)] \cdots [1 - m^I(n) - m^{II}(n)]\, \Delta_o$$

or

$$\Delta_n = \pi[1 - m^I(k) - m^{II}(k)]\, \Delta_o \tag{4}$$

where π denotes the product of the expressions[ ] for $k = 1, 2, \ldots n$.

Adding (1) and (2) gives

$$\Sigma_n = \Sigma_{n-1} + [m^I(n) - m^{II}(n)]\, \Delta_{n-1} \tag{5}$$

which is a second recurrence relation, so that

$$\Sigma_{n-1} = \Sigma_{n-2} + [m^I(n-1) - m^{II}(n-1)]\, \Delta_{n2}$$

$$\Sigma_1 = \Sigma_0 + [m^I(1) - m^{II}(1)]\ \Delta_0$$

Summing these equations and canceling equal terms on both sides leaves

$$\Sigma_n = \Sigma_0 + [m^I(1) - m^{II}(1)]\ \Delta_0 + [m^I(2) - m^{II}(2)]\ \Delta_1 + \ldots + [m^I(n) - m^{II}(n)]\ \Delta_{n-1} \quad (6)$$

and this may be written, using (4), as

$$\Sigma_n = \Sigma_0 + \Sigma[m^I(k) - m^{II}(k)]\pi\ [1 - m^I(l) - m^{II}(l)]\ \Delta_0 \quad (7)$$

where Σ is used to denote summation over $k = 1, 2, \ldots\ldots n$
and π now indicates the product of the bracketed expressions following it for $l = 1, 2, \ldots k-1$.

The general solution for q_n^I and q_n^{II} now follows from (4) and (7),

$$q_n^I = \tfrac{1}{2}\ [\Sigma_0 + \Sigma[m^I(k) - m^{II}(k) + f]\pi[1 - m^I(l) - m^{II}(l)]\Delta_0] \quad (8)$$

$$q_n^{II} = \tfrac{1}{2}\ [\Sigma_0 + \Sigma[m^I(k) - m^{II}(k) - f]\pi[1 - m^I(l) - m^{II}(l)]\Delta_0] \quad (9)$$

where $f = 0$ for $k \neq n$, $f = 1 - m^I(k) - m^{II}(k)$ for $k = n$.

For constant admixture rates

$$\pi[1 - m^I(l) - m^{II}(l)] = \pi(1 - m^I - m^{II}) = (1 - m^I - m^{II})^{k-1}$$

so that

$$q_n^I = \tfrac{1}{2}\ [\Sigma_0 + (m^I - m^{II})\Sigma(1 - m^I - m^{II})^{k-1}\ \Delta_0 + (1 - m^I - m^{II})^n\ \Delta_0] \quad (10)$$

$$q_n^{II} = \tfrac{1}{2}\ [\Sigma_0 + (m^I - m^{II})\ \Sigma(1 - m^I - m^{II})^{k-1}\ \Delta_0 - (1 - m^I - m^{II})^n\ \Delta_0] \quad (11)$$

These two expressions[1] simplify further to those given by Glass and Li (1953) for the discrete generation case in which there is one-way gene flow, $m^{II} = 0$ and $m = m^I$, so that

$$q_n^I = \tfrac{1}{2}(\Sigma_0 + \Delta_0) = q_0^I \quad (12)$$

$$q_n^{II} = \tfrac{1}{2}[\Sigma_0 + \Delta_0 + 2(1 - m)^n\Delta_0] = q_0^{II}(1 - m)^n + [1 - (1 - m)^n]q_0^I \quad (13)$$

MODEL II: TWO POPULATIONS, CONTINUOUS MIGRATION

All of the above deals with migrations occurring only at the end of generations and the treatment is wholly discrete. It is perhaps more realistic to consider continuous migration during each generation. It is reasonable to expect that for large values of n the difference between the effects of discrete and continuous migration will be small. It will be shown that the analogous expres-

[1] Expressions (10) (11) may be simplified by a slight change in notation: let $m = m^I - m^{II}$, and let $M = 1 - m^I - m^{II}$ then

$$q_n^I = \tfrac{1}{2}\ \Sigma_0 + \Delta_0\ \frac{[m + (-m + 1 - M)\ M^n]}{[1 - M]} \quad (10a)$$

$$q_n^{II} = \tfrac{1}{2}\ \Sigma_0 + \Delta_0\ \frac{[m + (-m - 1 + M)\ M^n]}{[1 - M]} \quad (11a)$$

sions to (12) and (13) derived by Glass and Li (1953) will be a simple case of the fuller treatment which follows.

Let $m^I(t)$ and $m^{II}(t)$ be the rates of admixture from populations I and II to each other, respectively, at time (t) so that in the small interval (t, t + δt) the amounts of migration may be written $m^I(t)\,\delta t$ and $m^{II}(t)\,\delta t$.

If further $q^I(t)$ and $q^{II}(t)$ represent the gene frequencies in the two populations at time t,

$$q^I(t + \delta t) = [1 - m^{II}(t)\delta t]\, q^I(t) + m^{II}(t)\delta t\, q^{II}(t) \qquad (14)$$

$$q^{II}(t + \delta t) = m^I t(t)\delta t\, q^I(t) + [1 - m^I(t)\delta t] q^{II}(t) \qquad (15)$$

Subtracting,

$$\Delta(t + \delta t) = [1 - \delta t\, m^I(t) - \delta t\, m^{II}(t)]\, \Delta(t)$$

or

$$\Delta(t + \delta t) - \Delta(t) = -[m^I(t) + m^{II}(t)]\, \Delta(t)\, \delta t.$$

Dividing by δt and letting $\delta t \longrightarrow 0$,

$$\frac{d\Delta(t)}{dt} = -[m^I(t) + m^{II}(t)]\, \Delta(t)$$

$$\text{so that } \left[\log \Delta(t')\right]_{t'=0}^{t'=t} = -\int_o^t [m^I(t') + m^{II}(t')]\, dt'$$

$$\text{or } \Delta(t) = \Delta_o \exp\left[-\int_o^t [m^I(t') + m^{II}(t')]\, dt'\right] \qquad (16)$$

On the other hand, adding equations (14) and (15)

$$\Sigma(t + \delta t) = \Sigma(t) + \delta t\, [m^I(t) - m^{II}(t)]\, \Delta(t)$$

$$\frac{d\Sigma(t)}{dt} = [m^I(t) - m^{II}(t)]\, \Delta_o \exp\left[-\int_o^t [m^I(t') + m^{II}(t')]\, dt'\right]$$

so that

$$\Sigma(t) = \Sigma_o + \Delta_o \int_o^t [m^I(t'') - m^{II}(t'')] \left[\exp - \int_o^{t''} [m^I(t') + m^{II}(t')]\, dt'\right] dt'' \qquad (17)$$

where t' and t'' are carrying variables for the integration.

Adding and subtracting (16) and (17) will now lead to general expressions for $q^I(t)$ and $q^{II}(t)$ which are

$$q^I(t) = \tfrac{1}{2}\left(\Sigma_o + \Delta_o \int_o^t [m^I(t'') - m^{II}(t'')] \exp\left[-\int_o^{t''} [m^I(t') + m^{II}(t')]\, dt'\right] dt'' + \Delta_o \exp\left[-\int_o^t [m^I(t') + m^{II}(t')]\, dt'\right]\right) \qquad (18)$$

$$q^{II}(t) = \tfrac{1}{2}\left(\Sigma_o + \Delta_o \int_o^t [m^I(t'') - m^{II}(t'')] \exp\left[-\int_o^{t''} [m^I(t') + m^{II}(t')]\, dt'\right] dt'' - \Delta_o \exp\left[-\int_o^t [m^I(t') + m^{II}(t')]\, dt'\right]\right) \qquad (19)$$

These expressions cannot be simplified further and represent the most general solution to this problem for continuous migration at varying rates between two populations.

It is possible to write down general simpler forms for (18) and (19) if one of the rates is zero. Say $m^{II} = 0$. Then

$$q^I(t) = \tfrac{1}{2}\left[\Sigma_o + \Delta_o\, e^{-2\int m^I(t)\, dt}\right] \qquad (18a)$$

$$q^{II}(t) = \tfrac{1}{2}\left[\Sigma_o - 2\,\Delta_o\, e^{-\int m^I(t)\, dt} + \Delta_o \left[e^{-2\int m^I(t)\, dt}\right]_{t=o}\right] \qquad (19a)$$

where the integrals in these two expressions are all indefinite.

Another simple solution occurs if both rates are equal, say

$m^I = m^{II} = m$. Then

$$q^I(t) = \tfrac{1}{2}(\Sigma_o + \Delta_o\, e^{-2\int_o^t m(t)\,dt}) \tag{18b}$$

$$q^{II}(t) = \tfrac{1}{2}(\Sigma_o - \Delta_o\, e^{-2\int_o^t m(t)\,dt}) \tag{19b}$$

Again, if $m^I(t)$ and $m^{II}(t)$ are assumed to be the constants m^I and m^{II}, the simplified solutions for this case are obtained after integrating in (18) and (19) as

$$q^I(t) = \tfrac{1}{2}\left\{\Sigma_o + \frac{\Delta_o\,(m^I - m^{II})}{(m^I + m^{II})} + \frac{2m^{II}\,\Delta_o\, e}{(m^I + m^{II})}^{-(m^I + m^{II})t}\right\} \tag{20}$$

$$q^{II}(t) = \tfrac{1}{2}\left\{\Sigma_o + \frac{\Delta_o\,(m^I - m^{II})}{(m^I + m^{II})} - \frac{2m^I\,\Delta_o\, e}{(m^I + m^{II})}^{-(m^I + m^{II})t}\right\} \tag{21}$$

Simplifying further for the case of unidirectional gene flow at a constant rate ($m^{II} = O$, $m^I = m$) leads to the equation of Glass and Li

$$q^I(t) = \tfrac{1}{2}\,[\Sigma_o + \Delta_o] = q_o{}^I \tag{22}$$

$$q^{II}(t) = \tfrac{1}{2}\,[\Sigma_o + \Delta_o - 2\Delta_o\, e^{-mt}]$$

$$= q_o{}^I - (q_o{}^I - q_o{}^{II})\, e^{-mt} \tag{23}$$

In equations (20) and (21) it is interesting to observe that the asymptotic values (as $t \rightarrow \infty$)

$$q^I(\infty) = q^{II}(\infty) = \tfrac{1}{2}\,[\Sigma_o + \Delta_o\,(m^I - m^{II})] \tag{24}$$

are identical with the corresponding values for the discrete case, q_∞^I, q^{II}, which may be obtained from equations (10) and (11).

A further simple case is that in which gene flow is unidirectional but at a rate increasing hyperbolically with time, e.g., a situation in which initial strong resistance to mating with immigrants broke down more and more rapidly. Substituting in formulas (22), (23)

$$m^{II} = 0,\ m^I = \frac{c}{a - t}$$

where a and c are suitable constants for this model, not necessarily integers, gives solutions

$$q^I(t) = \tfrac{1}{2}(\Sigma_o + \Delta_o) = q_o{}^I \tag{25}$$

$$q^{II}(t) = \tfrac{1}{2}\left\{\Sigma_o + \Delta_o - \Delta_o \frac{2(a - t)^c}{a^c}\right\}$$

$$= q_o{}^I - (q_o{}^I - q_o{}^{II})\,\frac{(a - t)^c}{a^c} \tag{26}$$

A similarly simple case is that where rates of gene flow in both directions are equal at any given time, but are both increasing as above, i.e.,

$$m^I = m^{II} = \frac{c}{a - t}$$

$$q^I(t) = \tfrac{1}{2}\left\{\Sigma_o + \Delta_o \frac{(a-t)^{2c}}{a^{2c}}\right\} \tag{27}$$

$$q^{II}(t) = \tfrac{1}{2}\left\{\Sigma_o - \Delta_o \frac{(a-t)^{2c}}{a^{2c}}\right\} \tag{28}$$

If a constantly increasing rate is needed, such as when $m^I = at^2 + bt + c$,

so that $\frac{d^2m^1}{dt^2}$ = a (a constant) and $\frac{dm^1}{dt}$ = at + b (linearly increasing). Then, if $m^{11} = 0$, (18a) and (19a) give

$q^1(t) = q^1_0$

$q^{11}(t) = q_0^1 - 2\Delta_0 e^{-(\frac{1}{3}at^3 + \frac{1}{2}bt^2 + ct)}$

If $m^{11} = m^1$, (18b) and (19b) give

$q^1(t) = \frac{1}{2}[\Sigma_0 + \Delta_0 e^{-2(\frac{1}{3}at^3 + \frac{1}{2}bt^2 + ct)}]$

$q^{11}(t) = \frac{1}{2}[\Sigma_0 - \Delta_0 e^{-2(\frac{1}{3}at^3 + \frac{1}{2}bt^2 + ct)}]$

MODEL III: MORE THAN TWO POPULATIONS, CONSTANT MIGRATION RATE

The above theory may be extended to cover N = 3 or more populations by writing

$$\underset{\sim}{q}_n = \underset{\sim}{M}\,\underset{\sim}{q}_{n-1} \qquad (29)$$

where $\underset{\sim}{M}$ represents the matrix of coefficients of the elements q^i_{n-1} of $i = 1, 2, \ldots N$ of the vector $\underset{\sim}{q}_{n-1}$. The matrix equation (29) is the N-dimensional analogy of the equations (10) and (11) for the N = 2 case already considered.

As before, it is desired to express $\underset{\sim}{q}_n$ in terms of $\underset{\sim}{q}_0$, the vector of initial frequencies, and n, the number of generations where the generations are discrete. Such an expression is, for constant gene flow rates,

$$\underset{\sim}{q}_n = \underset{\sim}{M}^n\,\underset{\sim}{q}_0 \qquad (30)$$

where $\underset{\sim}{M}^n$ means the matrix multiplied by itself n times; this is one possible way of calculating its value. When however its latent roots are distinct, $\underset{\sim}{M}$ may be written as a linear combination of N particular matrices, $\underset{\sim}{M}_i$, with the property

$$\underset{\sim}{M}^n = \Sigma\lambda^n_i\,\underset{\sim}{M}_i \qquad (31)$$

where the summation extends over $i = 1, \ldots N$ and the λ_i are the latent roots of $\underset{\sim}{M}$. Such an expression of the matrix $\underset{\sim}{M}$ is termed its *spectral resolution* and this property can be exploited in the calculation of $\underset{\sim}{q}_n$ using (30) in the form:

$$\underset{\sim}{q}_n = \Sigma\lambda_i^n\,\underset{\sim}{M}_i\,\underset{\sim}{q}_0 \qquad (32)$$

To construct the spectral set of matrices M_i it is first necessary to obtain the latent roots of $\underset{\sim}{M}$ in the usual way (see for instance Aitken, 1956) and use them to find N pairs of vectors $\underset{\sim}{s}_i$ and $\underset{\sim}{t}_i$ such that

$\underset{\sim}{M}\,\underset{\sim}{s}'_i = \lambda_i\,\underset{\sim}{s}_i$ and $\underset{\sim}{t}'_i\,\underset{\sim}{M} = \lambda_i\,t_i$, where $\underset{\sim}{s}'$ is the transpose of $\underset{\sim}{s}$, and $\underset{\sim}{t}'$ of $\underset{\sim}{t}$.

The spectral set is the set of matrices

$$\underset{\sim}{M}_i = [\underset{\sim}{s}_i\,\underset{\sim}{t}'_i]\,/\underset{\sim}{s}'_i\underset{\sim}{t}_i \qquad (i = 1, 2, \ldots\ldots N) \qquad (33)$$

For N = 3 or more populations the solutions cannot be written down in the manner of the earlier cases but the brief description given above together with the following examples should be sufficient to indicate how such solutions may be obtained.

EXAMPLES

I. *From the northern Nilotes*

To illustrate the problems that can be dealt with by this procedure, consider, among the northern Nilotic populations of the southern Sudan, the group com-

prising the Nuer, Dinka and Shilluk. Observations were made on the incidence of intermarriages in these peoples, and with these data the trends in gene frequencies can be examined, if the necessary assumptions are made. One assumption is that the observed gene frequencies are the real gene frequencies of the populations. The second is that for the purpose of the present model the intermixture rates are taken to be constant, an assumption that does not seem too unrealistic from what is known of the cultural inertia of primitive folk, and which appears to be supported by comparison of the two generations embraced in the present data. The populations are considered as panmictic, though the area over which they are settled makes this unlikely, and intermixture would be expected to be more frequent in those communities of Nuer and Dinka nearest geographically to the territorial confines. However, the observed intermixture rates can be regarded as those of the populations as a whole since the samples on which they are based covered villages throughout the whole of the southern half of Shillukland, the northern block of Dinka tribes from the Ageir to the Ruweng and Ngok, and the whole of the Nuer tribes from the Eastern Jikany to the Bul tribes.

In 288 marriages in Dinka villages, whose progeny would come to be regarded as Dinka, 8 spouses were Nuer and 5 were Shilluk, giving admixture rates of .01389 and .00868; in Shilluk villages, out of 255 marriages, 5 spouses were Dinka and none were Nuer (two further spouses were "Arab," omitted from the present analysis), giving admixture rates of .00980 and 0; and in 200 Nuer matings, one was with a Shilluk woman and 5 were with Dinka, giving admixture rates of .0025 and .0125. These admixture rates may be set out in the form of a table, the principal diagonal comprising the proportion of spouses each recipient population receives from itself.

RECIPIENT POPULATION	DONATING POPULATION		
	Nuer	Dinka	Shilluk
Nuer	.9850	.0125	.0025
Dinka	.0138	.9775	.0087
Shilluk	0	.0098	.9902

These data can be used in the calculation of gene frequencies at any given generation in the past, or in the future. For simplicity, the latter situation is illustrated first.

(a) *Given the constant admixture rates and the present gene frequencies, to find the frequencies in some future generation.*

The present frequencies of the blood group M gene are .5750, .5670 and .5047, in Nuer, Dinka and Shilluk, respectively (Roberts, Ikin and Mourant, 1955).

In equation (30), the vector of initial gene frequencies $q_o =$

$$\begin{bmatrix} .5750 \\ .5670 \\ .5047 \end{bmatrix}$$ and the matrix of admixture rates $\underset{\sim}{M} = \begin{bmatrix} .9850 & .0125 & .0025 \\ .0138 & .9775 & .0087 \\ 0 & .0098 & .9902 \end{bmatrix}$

The latent roots of $\underset{\sim}{M}$ are 1, .986937, .965763, so the spectral resolution of $\underset{\sim}{M}$ (equation 31)

$$\underset{\sim}{M}^n = \left\{ 1 \begin{bmatrix} .302388 & .328683 & .768929 \\ .302388 & .328683 & .768929 \\ .302388 & .328683 & .768929 \end{bmatrix} + \begin{bmatrix} .419578 & .058890 & -.478469 \\ .162804 & .022850 & -.185654 \\ -.488945 & -.068626 & .557571 \end{bmatrix} (.986937)^n + \begin{bmatrix} .278033 & -.387572 & .109539 \\ -.465191 & .648467 & -.183276 \\ .186557 & -.260057 & .073500 \end{bmatrix} (.965763)^n \right\} \quad (34)$$

Multiplying, as indicated in equation (32), the right hand side of equation (34) by the vector $\underset{\sim}{q}_0$ gives

$$\underset{\sim}{q}_n = \begin{bmatrix} .546435 \\ .546435 \\ .546435 \end{bmatrix} 1 + \begin{bmatrix} .033165 \\ .012869 \\ -.038648 \end{bmatrix} (.986937)^n + \begin{bmatrix} -.004600 \\ .007697 \\ -.003087 \end{bmatrix} (.965763)^n \quad (34a)$$

In this form the vector of frequencies for any generation may be readily obtained as in table 1 and in Fig. 1.

TABLE 1.

n =	5	10	15	20	30	40	60
Nuer	.5736	.5723	.5709	.5696	.5672	.5649	.5610
Dinka	.5650	.5632	.5616	.5602	.5578	.5559	.5532
Shilluk	.5076	.5104	.5129	.5152	.5193	.5228	.5285

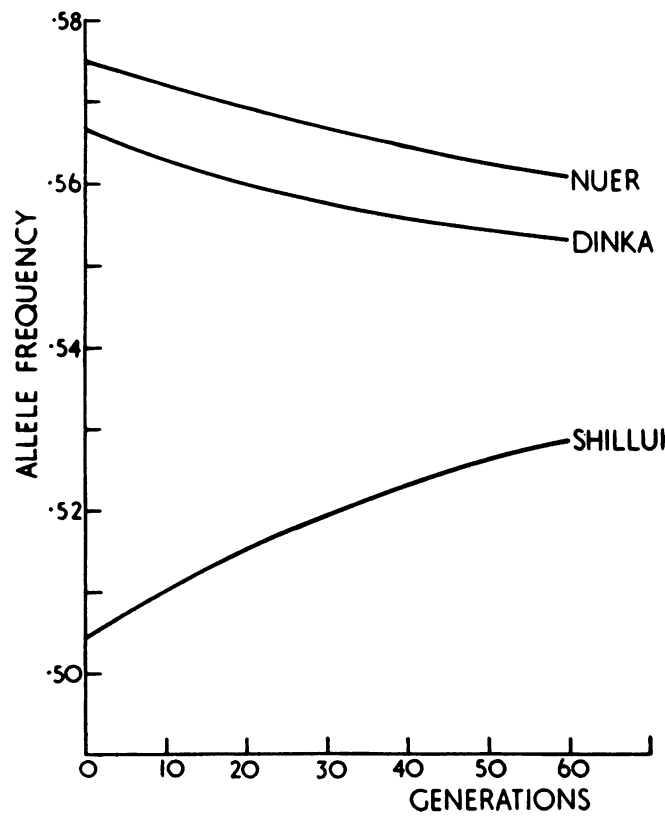

FIG. 1. Predicted approach through intermixture of frequencies of the gene for blood group M in the Northern Nilotes.

The asymptotic value is the first vector on the right hand side of equation (34a), so that the ultimate allele frequency that all three populations will attain is .5464.

(b) *Given the constant admixture rates and the present gene frequencies, to find the frequencies in a previous generation.*

The similarity in blood group gene frequencies of these peoples has been interpreted as indicating a derivation from a common original stock (Roberts *et al.*, 1955). A possible alternative explanation is that the gene frequencies in three populations formerly diverse came to resemble each other more and more closely as a result of continued intermixture so that some at least of the slight frequency differences now observed, instead of being attributable to sampling error, represent the end stages of this process. With the available data this hypothesis can be examined.

The value of n is taken as 20 for the present analysis, for the Shilluk according to folklore seem to have entered their present territory about 1500 A.D. The matrix $\underset{\sim}{M}$ is as before, but the present gene frequencies are taken as the vector q_n. From the equation $q_o = q_n \Sigma \lambda_i^n \underset{\sim}{M}_i$, where the $\underset{\sim}{M}_i$ are now the spectral matrices of the reciprocal matrix $\underset{\sim}{M}^{-1}$, the frequencies in the three populations 20 generations ago of the four alleles (out of those available) at which the Shilluk are today most divergent, viz., those of genes M and S of the MNS blood group system, of gene d of the Rh system, and of the gene for blood group B, may be calculated to be as shown in table 2.

TABLE 2.

	q_o (ca. 1500 A.D.)				q_n			
	M	S	d	B	M	S	d	B
Nuer	.5803	.2559	.1766	.1324	.5750	.2254	.1782	.1270
Dinka	.5786	.1197	.1566	.1158	.5670	.1505	.1759	.1155
Shilluk	.4900	.1501	.2709	.0894	.5047	.1477	.2523	.0941

The over-all effect of intermixture, on the assumptions previously stated, has been quite slight. It seems that 20 generations ago the three peoples were still very similar in their gene frequencies; in Nuer and Dinka the M allele frequencies were even closer than today, while the Shilluk were a little more divergent from the other two populations in frequencies of three of the alleles as expected, though not in the fourth. The importance of initial distance between frequencies for the results of intermixture is well shown by comparing B and M frequencies. In these, Dinka initially occupied an intermediate position between the other two populations, in the former being almost equidistant from both, but in the latter being very close to the Nuer. The same intermixture has brought the Dinka away from the Nuer in M frequency, but has brought the Nuer closer to the Dinka in B frequency. The course followed by the three populations in gene S (Fig. 2) shows that the positions of populations relative to each other with regard to the frequency of a gene are not necessarily constant during the approach to the asymptotic value.

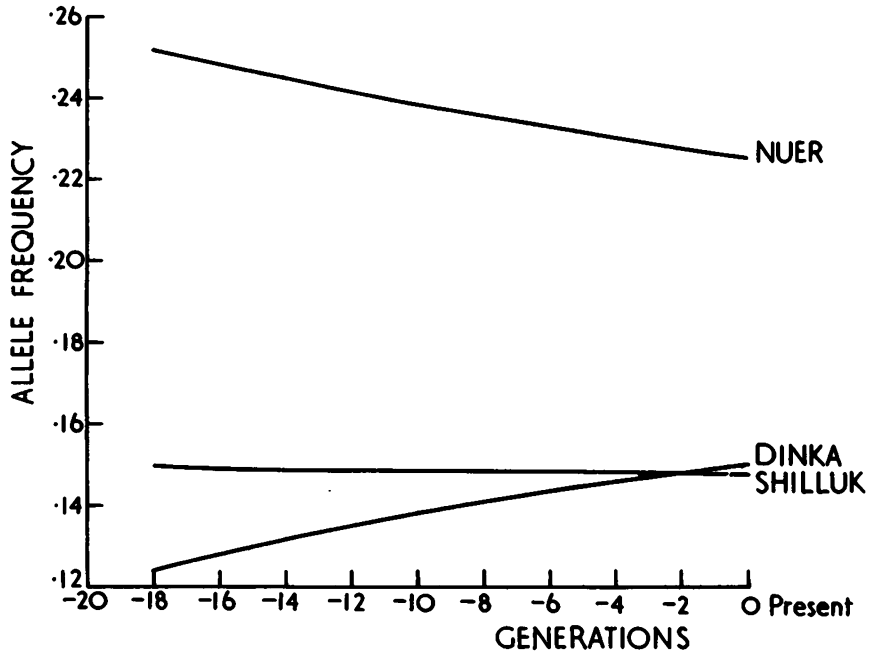

FIG. 2. Calculated past approach of frequencies of the gene for blood group S.

(c) *Given gene frequencies of different generations, to find the constant intermixture rates.*

In the former cases the procedure was carried out for a single allele at a time. For a solution to be obtained in this third type of problem data on frequencies of at least as many independent alleles as there are intermixing populations are required. A further factor has also to be considered. Determinations of gene frequencies are subject to sampling error, so if there are more than the minimum available the estimate of intermixture derived from one set of allele frequencies may not coincide with that from a different set. What is required is the most probable estimate of intermixture. The best estimates derive from a group of independent loci, at which the allele frequencies in the parent populations are well separated. The Rh genes are rather too heavily represented in the examples which follow.

For N intermixing populations, the frequencies of k loci are known in generations n and o. Form from k vectors q_n an N x k matrix $\underset{\sim}{Q}_n$, and similarly form from k vectors q_o another matrix $\underset{\sim}{Q}_o$, so that

$\underset{\sim}{Q}_n = \underset{\sim}{M}^n\underset{\sim}{Q}_o$ (corresponding to equation 30)

$\underset{\sim}{Q}_n\underset{\sim}{Q}_o' = \underset{\sim}{M}^n\underset{\sim}{Q}_o\underset{\sim}{Q}_o'$ (where $\underset{\sim}{Q}_o'$ is the transpose of $\underset{\sim}{Q}_o$)

and $\underset{\sim}{M}^n = \underset{\sim}{Q}_n\underset{\sim}{Q}_o' [\underset{\sim}{Q}_o\underset{\sim}{Q}_o']^{-1}$

$\underset{\sim}{M}^n$ can now, if its latent roots are distinct, be spectrally resolved into N matrices $\underset{\sim}{M}_i$ (i = 1, 2, . . . , N), and, if the latent roots corresponding to these spectral matrices are indicated by λ_i^n,

$$M^n = \Sigma\, \lambda_i^n M_i \qquad (35)$$

TABLE 3.

$$\underset{\sim}{M}^n = \begin{matrix} & M & S & d & B \end{matrix}$$

$$\begin{bmatrix} .5750 & .2254 & .1782 & .1270 \\ .5670 & .1505 & .1759 & .1155 \\ .5047 & .1477 & .2523 & .0941 \end{bmatrix} \begin{bmatrix} .5803 & .5786 & .4900 \\ .2559 & .1197 & .1501 \\ .1766 & .1566 & .2709 \\ .1324 & .1158 & .0894 \end{bmatrix} \left\{ \begin{bmatrix} .5803 & .2559 & .1766 & .1324 \\ .5786 & .1197 & .1566 & .1158 \\ .4900 & .1501 & .2709 & .0894 \end{bmatrix} \begin{bmatrix} .5803 & .5786 & .4900 \\ .2259 & .1197 & .1501 \\ .1766 & .1566 & .2709 \\ .1324 & .1158 & .0894 \end{bmatrix} \right\} - 1$$

$$= \begin{bmatrix} .7639 & .1811 & .0549 \\ .1963 & .6691 & .1346 \\ .0194 & .1468 & .8337 \end{bmatrix} \begin{bmatrix} 67.05 & -46.90 & -26.25 \\ -46.90 & 77.80 & -27.98 \\ -26.25 & -27.98 & 60.90 \end{bmatrix} = \begin{bmatrix} .7651 & .1810 & .0538 \\ .1963 & .6695 & .1341 \\ .0177 & .1461 & .8363 \end{bmatrix}$$

$$= (1.000) \begin{bmatrix} .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \end{bmatrix} + (.7727) \begin{bmatrix} .4196 & .0589 & -.4785 \\ .1628 & .0229 & -.1857 \\ -.4889 & -.0686 & .5576 \end{bmatrix} + (.4984) \begin{bmatrix} .2780 & -.3876 & .1095 \\ -.4652 & .6485 & -.1833 \\ .1866 & -.2601 & .0735 \end{bmatrix}$$

The 20th roots of the latent roots, i.e., $\sqrt[20]{1} = 1$, $\sqrt[20]{(0.7727)} = .9869$ and $\sqrt[20]{(0.4984)} = .9658$, are most conveniently obtained by using logarithms, so that the expression for $\underset{\sim}{M}$ is

$$\underset{\sim}{M} = (1.000) \begin{bmatrix} .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \end{bmatrix} + (.9869) \begin{bmatrix} .4196 & .0589 & -.4785 \\ .1628 & .0229 & -.1857 \\ -.4889 & -.0686 & .5576 \end{bmatrix} + (.9658) \begin{bmatrix} .2780 & -.3876 & .1095 \\ -.4652 & .6485 & -.1833 \\ .1866 & -.2601 & .0735 \end{bmatrix}$$

and, summing the expressions on the right hand side,

$$\underset{\sim}{M} = \begin{bmatrix} .9850 & .0125 & .0025 \\ .0138 & .9775 & .0087 \\ 0 & .0098 & .9902 \end{bmatrix}$$

[In the above computation ten decimal places were retained but only four are shown here for economy of space.]

From this relation the matrix $\underset{\sim}{M}$ may be determined uniquely, since each latent root λ_i^n is a real positive number [2] and therefore has a single real positive root which will be called λ_i. Since the resolution of $\underset{\sim}{M}^n$ exists and is unique for distinct latent roots of this matrix there is a single matrix $\underset{\sim}{M}$ whose nth power satisfies (35), and it may be obtained from

$$\underset{\sim}{M} = \Sigma\lambda_i M_i \qquad (36)$$

where the λ_i are the nth roots of the latent roots λ_i^n. It may be noted that this is the "least squares" solution for $\underset{\sim}{M}^n$ given $\underset{\sim}{Q}_n$ and $\underset{\sim}{Q}_o$. In the situation however in which two latent roots of $\underset{\sim}{M}^n$ are equal or nearly so, the nth root (i.e., $\underset{\sim}{M}$ the migration rate matrix) cannot be obtained by the above method and $\underset{\sim}{M}^n$ is the only answer available, representing the accumulated component from each source in each population;[3] in such a case, however, it may be possible to obtain an algebraic solution, as in footnote 4.

In the estimation of $\underset{\sim}{M}^n$ by this procedure, sampling error in Q_o or Q_n may cause the row sums of $\underset{\sim}{M}^n$ to deviate from unity and the migration rates to fall outside the range 0 to 1. (Indeed it may be that such error will be large enough to cause $\underset{\sim}{M}^n$ not to have positive latent roots, in which case derivation of $\underset{\sim}{M}$ by the above method will not be possible.) Before actually carrying out the resolution, therefore, the matrix obtained should be modified first by adding the largest negative element in a row to all elements in that row (to make all elements positive), and then by dividing each modified element by the new row sum. Until such time as the errors of the estimates involved in the present procedure have been established, departure from unity of the unmodified row sums of $\underset{\sim}{M}^n$ may be taken as a guide to the reliability of the elements of $\underset{\sim}{M}^n$; there would be no departure at all if there were no error in $\underset{\sim}{Q}_n$ and $\underset{\sim}{Q}_o$.

The working of this type of example may be demonstrated, for purposes of illustration only, from the data obtained in example (b) above; n is taken as 20 generations, $\underset{\sim}{Q}_o$ and $\underset{\sim}{Q}_n$ are, respectively, the left and right sections of table 2, so that the calculations are as in table 3.

II. *From the American Negro and Indian*

The third type of example dealt with above may be applied to the American Negro. Roberts (1955) pointed out that the average rate of European gene flow into the American Negro of .02 — .025 per generation, estimated from a model involving two parental populations, would be an overestimate if the American Indian had made an appreciable contribution to the hybrid. Glass (1955) argued that there had been no significant American Indian contribution to the American Negro gene pool. As the former paper demonstrated,

[2] The latent roots are real positive numbers if the sum of the elements on the principal diagonal of the matrix is greater than unity ($m_{11} + m_{22} + m_{33} > 1$). In the model envisaged where most mating occurs within each population, this condition is always satisfied.

[3] A further inequality must also be satisfied for the latent roots to be distinct, viz., $(m_{22} - m_{12})(m_{33} - m_{13}) > (m_{32} - m_{12})(m_{23} - m_{13})$. This condition holds when there is some migration into all three populations, but not for instance if there is migration into only one of them, i.e., if some of the off diagonal elements in the matrix are zero. See examples.

there is considerable variation in the estimate of admixture according to which African populations are regarded as providing the original frequencies for the American Negro. So too will there be variation according to which frequencies are taken as representative of the original American Indian. For the present analysis, however, figures from those used by Glass are applied, i.e., the pooled gene frequencies for all relevant African populations, the pooled figures for the present American Negro and those selected to represent the American Indian original frequencies (table 4), and data on a further allele (Kell) have been

TABLE 4.

	Q_o Original frequencies			Q_m Present frequencies		
	West African	White	American Indian	American Negro	White	American Indian
R^o	.5512	.0279	0000	.4332	.0279	0000
S	.1344	.3374	.3416	.1708	.3374	.3416
Jk^b	.217	.482	.230	.269	.482	.230
M	.4788[1]	.540	.718	.489	.540	.718
R'	.0692	.4203	.3367	.1582	.4203	.3367
R^2	.0861	.1499	.5303	.1088	.1499	.5303
r	.2111	.3842	0000	.2637	.3842	0000
k	.9912[2]	.9336[3]	1.0000[4]	.9823[5]	.9336[3]	1.0000[4]

[1]Modified from Glass (1955).
[2]Barnicot and Lawler (1953).
[3]Wiener and Gordon (1951).
[4]Chown and Lewis (1953).
[5]Miller, Rosenfield and Vogel (1951).

added. The Negro is assumed to have made no contribution to the gene pool of the European or the Indian and the Indian no contribution to the European. From the frequencies of the eight characters set out in table 4 it can be calculated that in the American Negro there is a total white component of 23.2 per cent and no Indian component. In this case two of the latent roots of $\underset{\sim}{M}^n$ are equal so M cannot be obtained by the present method; however, the admixture rates may be calculated algebraically by the procedure set out in the footnote[4], giving a gene flow from White to Negro of .026 per generation over 10 generations.

A second estimate may be obtained using the frequencies for the original American Indian population suggested by Chown and Lewis (1953), again on the assumption that gene flow was in one direction only into the Negro from White and Indian. Again the calculations indicate in the American Negro a white component of 23.4 per cent and an Indian of zero.

It is possible that estimates of admixture employing gene frequencies of

[4]Suppose

$$M = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_{31} & m_{32} & m_{33} \end{Bmatrix} \qquad M^n = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_{31}^n & m_{32}^n & m_{33}^n \end{Bmatrix}$$

then it can be shown that

$$m_{11} = \sqrt[n]{m_{11}^{(n)}};\ m_{12} = \frac{m_{12}^{(n)}\ (1 - m_{11})}{(1 - m_{11}^n)};\quad m_{13} = \frac{m_{13}^{(n)}\ (1 - m_{11})}{(1 - m_{11}^n)}.$$

Indian groups as far away from the major centers of Negro influx as the Chippewa may be inaccurate on account of regional variations in American Indian gene frequencies. Recent studies of remnant groups of southeastern Indians have made up for the former lack of information on them, and a third estimate of admixture has been obtained using Pollitzer's unpublished data on Cherokee frequencies. Assuming that the frequencies observed in present day Cherokee "fullbloods" are those of the parental Indian population, and again that gene flow was in one direction only, into the Negro from white and Indian, the calculations indicate in the American Negro a white component of 26.3 per cent and an Indian of zero. Again, taking the present Cherokee fullblood frequencies as those of the ancestral Indian population, the frequencies in all Cherokee samples as those of the present Indian population (allowing for gene flow in two directions, into the Negro and Indian but not into the White), the same admixture figures appear for the American Negro while the present Cherokee are seen to be 6.1 per cent white and zero Negro. The Cherokee phenotypic frequencies do not allow differentiation of the Rh gene combinations cDe, and cde, and in the above calculations all such (.0385 fullblood, .0460 whole group) were assigned to cde and none to cDe. If, however, as a less satisfactory alternative all are assigned to cDe and none to cde, then the present Cherokee are 1.1 per cent white and zero Negro, and the present Negro are 26.0 per cent white and zero Indian.

A further example in which the frequencies in all three populations were changing may be drawn from American populations again employing eight alleles (table 5). The West African and American Negro frequencies were

TABLE 5.

	Q_o			Q_n		
	West African	European	Full Chippewa	American Negro	Minnesota White	Present Chippewa
R^o	.5512	.0206	0	.4332	.0573	.0354
S	.1344	.3278	.3416	.1708	.2841	.3555
M	.4788	.5470	.7205	.4890	.5983	.5940
R^1	.0692	.4235	.3367	.1582	.4127	.3841
R^2	.0861	.1483	.5303	.1088	.1294	.4287
r	.2111	.3907	.0803	.2637	.3866	.1204
Fy^b	.8914	.5872	.1364	.8602	.5656	.2917
k	.9912	.9539	.9252	.9823	.9434	.9448

as above; the original American Indian frequencies were taken to be those of Matson, Koch and Levine (1954) for fullblood Chippewa and the present Indian frequencies those of Matson's complete Chippewa sample (his sample only included those of a quarter or more Indian ancestry as given in Agency records); for original frequencies for the white population Norwegian figures of Hartman (1944 and unpublished) and Heisto (1953) were applied, though in the absence of Scandinavian frequencies of S and Duffy the English figures of Race and Sanger (1950) were incorporated; for the present population the Minnesota White frequencies of Matson *et al.* (1954) were used. Unpublished

data of Ikin *et al.* were used for the Duffy frequencies in West Africa and England, respectively. The amount of accumulated admixture shows the present Negro to be 26.1 per cent white with no Indian component, the present Indian is 25.6 per cent white with a Negro component of 4.1 per cent while the present white frequencies may be explained by a slight (4.6 per cent) Negro and Indian (3.5 per cent) admixture.

If such admixture had been continuing at a constant rate for 6 generations then the rate of gene flow per generation would be as follows:

		From Negro	From White	From Indian
Into	Negro	.950	.051	0
	White	.009	.984	.007
	Indian	.008	.050	.942

Lest it be argued that it is unlikely that the Negro frequencies would have remained until 6 generations ago the same as in the original West African populations, figures halfway between those in columns 1 and 4 of table 4 were arbitrarily taken as representing the Negro frequency at that time. On this basis the rates of gene flow from Negro into Minnesota White and Indian would have been very little different from those above. Some of the apparent Negro admixture may be due to the choice of Scandinavian samples to represent the original white frequencies instead of allowing for a component from further south in Europe. Indeed a survey of a township in southern central Minnesota in 1910 showed in that particular town a preponderance of non-Scandinavian ancestry, with Germans (30.8 per cent) the majority group. Alternatively, instead of the apparent Negro component deriving from ancient European gene frequency variation, it may have been derived through Negro admixture in the antecedents of that section of the population of Minnesota which settled there from other parts of the United States. Such difficulties of interpretation serve to emphasize the arbitrary manner of selecting parental population frequencies upon which of course the actual numerical results depend. All estimates confirm Glass's argument of the absence of an Indian component in the American Negro, though local populations of Negroes in parts of the South still require examination.

SUMMARY

Procedures are described for calculating the change in gene frequency when two or more populations are intermixing. The methods in the case of three populations are illustrated by examples from Africa and North America.

ACKNOWLEDGMENTS

Acknowledgment is gratefully made to Dr. W. S. Pollitzer for permission to use his unpublished Cherokee data; to the director and staff of the Oxford University Computing Laboratory for use of the Mercury computer and help in connection therewith; to all those who read and criticized the text and particularly Dr. D. G. Kendall, and Dr. M. T.

Newman; and to members of the Department of Human Genetics, Ann Arbor, Michigan, for their kind antenatal influence on some of the ideas now brought forth.

REFERENCES

AITKEN, A. C. 1956. *Determinants and matrices.* London: Oliver & Boyd.

ALSTRÖM, C. H. 1958. First cousin marriages in Sweden 1750-1844. *Acta Genet.* (Basel) 8: 296-369.

BARNICOT, N. A., AND LAWLER, S. D. 1953. A study of the Lewis, Kell, Lutheran and P blood group systems and the ABH secretion in West African Negroes. *Amer. J. Phys. Anthrop.* 11: 83-90.

BOYD, W. C. 1939. Blood groups of American Indians. *Amer. J. Phys. Anthrop.* 25: 215-235.

CHOWN, B., AND LEWIS, M. 1953. The blood groups and the secretor status of the Blackfoot Indians of Alberta. *Amer. J. Phys. Anthrop.* 11: 369-383.

CROW, J. F. 1958. Some possibilities for measuring selection intensity in man. *Hum. Biol.* 30:1-13.

FISHER, R. A. 1930. *The genetical theory of natural selection.* London: Oxford University Press.

GLASS, B. 1955. On the unliklihood of significant admixture of genes from the North American Indians in the present composition of the Negroes of the United States. *Amer. J. Hum. Genet.* 7: 368-385.

GLASS, B., AND LI, C. C. 1953. The dynamics of racial intermixture: an analysis based on the American Negro. *Amer. J. Hum. Genet.* 5: 1-20.

HALDANE, J. B. S. 1924-32. A mathematical theory of natural and artificial selection. *Proc. Camb. Phil. Soc.* 23: 19, 158, 363, 838; 26: 220; 27: 131; 28: 244.

HARTMANN, O. AND LUNDEVALL, J. U. 1944. *Blood group distribution in Norway.* Skrift. Norske Vidensk. Akad. No. 2.

HEISTO, H. 1953. The frequency of the Kell antigen in the population of Oslo. *Acta Path. Microbiol. Scand.* 33: 433-434.

IKIN, E. W., MOURANT, A. E., *et al.* Unpublished.

KIMURA, M. 1956. Rules for testing stability of a selective polymorphism. *Proc. Nat. Acad. Sci.* 42: 336.

LASKER, G. W. 1952. Mixture and genetic drift in ongoing human evolution. *Amer. Anthrop.* 44: 433.

MALECOT, G. 1948. *Les mathematiques de l'heredite.* Paris: Masson.

MATSON, G. A., KOCH, E. A., AND LEVINE, P. 1954. A study of the hereditary blood factors among the Chippewa Indians of Minnesota. *Amer. J. Phys. Anthrop.* 12:413-426.

MILLER, E. B., ROSENFIELD, R. E., AND VOGEL, P. 1951. On the incidence of some of the new blood agglutinogens in Chinese and Negroes. *Amer. J. Phys. Anthrop.* 9: 115-126.

OTTENSOOSER, F. 1944. Calculo do grau de mistura racial atravis dos grupos sanguineos. *Rev. Bras. Biol.* 4: 531-537.

OWEN, A. R. G. 1959. Mathematical models for selection, in *Natural selection in human populations,* D. F. Roberts and G. A. Harrison, eds. London: Pergamon Press.

RACE, R. R., AND SANGER, R. 1950. *Blood groups in man,* 3rd edition. Oxford: Blackwell Scientific Publications.

ROBERTS, D. F. 1955. The dynamics of racial intermixture in the American Negro: some anthropological considerations. *Amer. J. Hum. Genet.* 7: 361-367.

ROBERTS, D. F. 1956. A demographic study of a Dinka village. *Hum. Biol.* 28: 323-349.

ROBERTS, D. F. 1956. Some genetic implications of Nilotic demography. *Acta Genet.* (Basel) 6: 446.

ROBERTS, D. F., IKIN, E., AND MOURANT, A. E. 1955. Blood groups of the Northern Nilotes. *Ann. Hum. Genet.* 20: 135-154.

SALDANHA, P. H. 1957. Gene flow from white into Negro populations in Brazil. *Amer. J. Hum. Genet.* 9: 299-309.

SILVA, E. M. DA 1948. Blood groups of whites, Negroes and mulattos from the State of Maranhao, Brazil. *Amer. J. Phys. Anthrop.* 6: 423.

SILVA, E. M. DA 1949. Blood groups of Indians, whites and white-Indian mixtures in Southern Matto Grosso, Brazil. *Amer. J. Phys. Anthrop.* 7: 575-586.

STEINBERG, A. G., STAUFFER, R., AND BOYER, S. H. 1960. Evidence for a Gm^{ab} allele in the Gm system of American Negroes. *Nature* 188: 169-170.

STEVENS, W. L. 1952. Statistical analysis of the ABO system in mixed populations. *Hum. Biol.* 24: 12-24.

WIENER, A. S., AND GORDON, E. B. 1951. Répartition des facteurs de groupes sanguins dans une population de New York City. *Rev. Hémat.* 6: 45-50.

WRIGHT, S. 1931. Evolution in mendelian populations. *Genetics* 16: 97.