An Extension of Wahlund's Principle To Evaluate Mating Type Frequency

NORIKAZU YASUDA[1]

Recent studies on human population structure have tried to predict the frequencies of various types of consanguineous marriage to be expected under random mating (Cavalli-Sforza, 1958; Barrai *et al.*, 1962; Hajnal, 1963; Cavalli-Sforza *et al.*, 1966). An alternative approach is to describe phenotype (or genotype) and mating type frequencies in terms of gene frequencies and factors which depend on the structure of population and mating preference. In this paper, a possible model based upon Wahlund's principle (Wahlund, 1928) is proposed to describe human population structure in terms of gene frequency and the inbreeding coefficient. Application of the model to a population from northeastern Brazil (Morton, 1964) will be communicated elsewhere (Yasuda, 1967*b*).

RANDOM MATING POPULATION

A diploid phase generation begins when gametes from a gene pool are combined by pairs into zygotes in some regular manner; these zygotes experience migration, mutation, and differential mortality and fertility, and the generation terminates with the haploid phase gene pool of the next generation. By *panmixia* or random mating, we mean that uniting gametes are drawn independently from the gene pool, without restrictions due to finite population size, inbreeding, or assortative mating, and are enumerated before differential selection has acted. Accordingly, genotype frequencies and mating type frequencies can be calculated by the Hardy-Weinberg binomial law.

Confusion in the definition of random mating can occur, since there can be random mating with respect to gametes, genotypes, and phenotypes. The latter two as a unit of random mating might occasionally be useful (for example, problems of self-incompatibility or intermixture), but, as far as a single locus is concerned, the gene as the fundamental basis of population genetics is most pertinent both in practice and in theory. Therefore, we will consider only random combination of gametes as random mating.

Received March 13, 1967.

Supported by U.S. Public Health grants GM 10424, GM 12454, and FR 00247 from the National Institutes of Health, with the collaboration of the Universidade de São Paulo and Departamento de Imigração, São Paulo, Brazil. The work presented here is part of a thesis submitted to the graduate school of the University of Hawaii in partial fulfilment of the requirements for the Ph.D. degree. This paper also constitutes Contribution No. 638 from the National Institute of Genetics. Aided in part by a Grant-in-Aid for Fundamental Scientific Research from the Ministry of Education in Japan.

[1] Department of Genetics, University of Hawaii, Honolulu, Hawaii 96822. Present address: Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka-ken, Japan.

THE INBREEDING COEFFICIENT

Several interpretations of the inbreeding coefficient have been attempted (Wright, 1921; Bernstein, 1930; Malecot, 1948) and the same conclusion was reached with respect to zygote frequencies in terms of gene frequencies and the inbreeding coefficient (Table 1). The inbreeding coefficient can be understood as a measure of nonrandomness that also describes zygote frequencies, the correlation between uniting gametes (Wright, 1921), and the probability that two genes are identical by descent (Malecot, 1948). Besides these, it can measure the degree of differentiation in subdivided populations and describe mating type frequencies.

Furthermore, the inbreeding coefficient of each allele f_i, distinguishing from the coefficient f of a locus, may also be defined from: $f = \Sigma f_i p_i$. The quantity f_i may vary among alleles in polymorphic systems (Yasuda, 1966).

TABLE 1

EXPRESSION OF GENOTYPE FREQUENCIES BY DIFFERENT AUTHORS

Genotype	Hardy-Weinberg (1908)	Wright (1921)	Wahlund (1928)	Bernstein (1930)	Malecot (1948)
AA.....	p^2	p^2+pqF	$p^2+\sigma^2$	$p(p+\alpha q)$	$p^2(1-f)+pf$*
Aa.....	$2pq$	$2pq(1-F)$	$2pq-2\sigma^2$	$2pq(1-\alpha)$	$2pq(1-f)$
aa.....	q^2	q^2+pqF	$q^2+\sigma^2$	$q(q+\alpha p)$	$q^2(1-f)+qf$
Total..	$(p+q)^2=1$	1	1	1	1

* The expression is also given by Wright (1943).

NOTE.—Where p and q are frequencies of genes A and a, respectively, F is Wright's inbreeding coefficient, i.e., the correlation coefficient of uniting gametes; f is the probability that two genes are identical by descent; α is the mean inbreeding coefficient; and σ^2 is the gene frequency variance. Note that $F = f = \alpha = \sigma^2/pq$.

WAHLUND'S PRINCIPLE AND ITS EXTENSION

Discrete Model

Suppose that a population is divided into many endogamous panmictic smaller populations (isolates) restricted by geographical, racial, religious, social, and economic barriers. Let $w_i(\Sigma w_i = 1)$ be the relative size of the ith isolate. If a genetic system consists of two alleles A and a with frequencies p_i and q_i in the ith isolate, respectively, then the frequency p of gene A is $p = \Sigma p_i w_i$ and its variance σ^2 in the total population is $\Sigma(p_i - p)^2 w_i = \Sigma p_i^2 w_i - p^2$, where the summation is taken over all isolates. Since the frequencies of AA, Aa, and aa genotypes in the total population are given by $\Sigma p_i^2 w_i$, $\Sigma 2p_i q_i w_i$, and $\Sigma q_i^2 w_i$, respectively, the subdivision results in increasing homozygosity by an amount equal to the gene frequency variance σ^2 (Table 1). Wahlund (1928) discovered this principle and discussed it in the cases with and without dominance.

Comparison of heterozygous frequencies with Wright's result leads to his formula $\sigma^2 = p(1 - p)F$ (Wright, 1943).

All of the above arguments hold for an arbitrary number of alleles, for each of which an inbreeding coefficient can be defined as in the last section, and this leads to an interesting formula:

$$F = \sum_i \sigma_i^2/(1 - p_i),$$

where the summation is taken for all alleles.

It should be borne in mind that an "artificial" subdivision of population does not always result in increasing homozygosity. There would be no change observed whenever a given gene frequency was exactly the same for all isolates; that is, gene frequency variance was zero. This suggests an association between isolate size and probability density of gene frequency.

The Breakup of Isolates

Although the Wahlund principle has been employed to explain why the breakup of isolates decreases homozygosity, there has been no mathematical treatment of how the homozygosity decreases by removing one or more barriers. The following discussion may be helpful.

Suppose that a population consists of three isolates 1, 2, and 3, whose relative sizes are w_1, w_2, and w_3, respectively. In this population, the gene frequency p_{III} and its variance σ^2_{III} are $p_{III} = p_1w_1 + p_2w_2 + p_3w_3$ and $\sigma^2_{III} = p_1^2w_1 + p_2^2w_2 + p_3^2w_3 - p^2_{III}$, respectively. Suppose that the barrier between the isolates 2 and 3 is removed, creating a new isolate in which mating ultimately continues at random (perhaps after a few generations in which a gene cline persists). The relative size W of this new panmictic isolate and its gene frequency P are $W = w_2 + w_3$ and $P = (w_2p_2 + w_3p_3)/W$, respectively.

The gene frequency p_{II} and its variance σ^2_{II} in the total population become, therefore,

$$p_{II} = p_1w_1 + PW$$

$$= p_1w_1 + p_2w_2 + p_3w_3 = P_{III},$$

and

$$\sigma^2_{II} = p_1^2w_1 + P^2W - p^2_{II},$$

or

$$\sigma^2_{II} = \sigma^2_{III} - \frac{w_2w_3}{(w_2 + w_3)}(p_2 - p_3)^2.$$

Apart from mutation, selection, and random genetic drift, the population gene frequency does not change, whereas the gene frequency variance decreases in an amount that depends on the relative sizes and differences of gene frequencies of isolates whose barrier was removed. As a corollary, the change in the inbreeding coefficient is given by using Wright's formula and $p_{II} = p_{III} = p$,

$$F_{II} = F_{III} - F_B,$$

where F_B is a contribution due to the breakup of isolates and, in our terminology,

$$F_B = \frac{w_2w_3}{(w_2 + w_3)} \cdot \frac{(p_2 - p_3)^2}{p(1 - p)}.$$

More general treatment is given in Appendix I.

This elaboration of Wahlund's results can be applied to human populations. For instance, the barrier that was removed might be racial endogamy, and the effect of this on the inbreeding coefficient is immediately apparent. On the other hand, when new barriers are created under a certain circumstance, it is clear from the above discussion that the inbreeding coefficient increases by the amount F_B.

Continuous Model

Although it has been assumed that the barriers are discrete, an actual barrier is usually continuous, or we may not know what type of barrier it is. One of the approaches to bridge the gap is, then, an extension of Wahlund's model to continuous or mixed barriers. Since the result from Wahlund's discussion is described in terms of mean and variance of population gene frequencies, sums can be replaced by integrals. In this continuous model, each individual gene has a "probability density" to contribute to population gene, genotype, or mating type frequencies. Therefore, gene frequency and its variance in the population can be expressed by Lebesgue-Stieltjes integrals (Cramer, 1946),

$$p = \int p_w dW \qquad \text{and} \qquad \sigma^2 = \int p_w^2 dW - p^2 ,$$

where sums are taken for the discrete model and integrals for the continuous case. In the mixed case, the barrier may be separated into discrete and continuous types. Thus, Wahlund's principle covers any type of heterogeneous population. For instance, a continuous model where a population is divided by physical distance has been studied by Holgate (1964). Furthermore, in case of subdivision by time or generation, the probability density may correspond to a solution of the Fokker-Planck equation in population genetics (Wright, 1945). The situation in man is so complicated by factors such as time, space, population size, and human behavior that it may be difficult to find, even approximately, the appropriate probability density function or pattern of subdivision.

It should be emphasized here, however, that Wahlund's principle holds even for unknown density functions, and this extension replaces the concept of "isolate size" by "probability density of gene frequencies." A genetical interpretation of the probability density could be a tendency of genes to combine that would be affected by several genetic barriers.

Moments of a Subdivided Population

Since gene and genotype frequencies of a population are given by the first and the second moments with respect to possible isolates in the population, it seems worthwhile to consider the biological meanings of the moments. The first moment gives the gene frequency and the second moment the genotype frequency. The third and the fourth moments give the mating type frequencies at a sex-linked and an autosomal locus, respectively, since three and four genes are concerned in each gene combination.* More generally, whenever we consider a set of genes, the order of moment corresponds to the number of genes involved. These higher moments appear in studies of linkage, illegitimacy, polyploidy, heritability, and so on, but we shall restrict at-

* Mother-child combinations of autosomal genes will also be obtained from the third moments.

tention to the fourth and lower moments that correspond to mating type frequency for the study of population structure in man, although the results are completely general.

Let us consider a locus with two alleles A and a whose frequencies are p and q, respectively, in a subdivided population with inbreeding coefficient α. Suppose that the difference between gene frequency of an isolate, p_w, and of the population, p, is Δp_w, whose kth moment is expressed by m_k:

$$m_k = \int (\Delta p_w)^k dW = \int (p_w - p)^k dW ,$$

where integrals are understood in the Lebesgue-Stieltjes sense. For the first and the second moments, the following relations hold precisely:

$$m_1 = 0 ,$$

$$m_2 = p(1 - p)\alpha .$$

For the population moment, M_a,

$$M_a = \int p_w^a dW = \int (p + \Delta p_w)^a dW$$

$$= \sum_{r=0}^{a} \left[\binom{a}{r} p^{a-r} \int (\Delta p_w)^r dW \right]$$

$$= \sum_{r=0}^{a} \binom{a}{r} p^{a-r} m_r ;$$

or

$$M_a = p^a + \frac{a(a-1)}{2} p^{a-1}(1 - p)\alpha + O(m_3) .*$$

In the above expression, if the cubic and higher powers of Δp_w are negligible, the term $O(m_3)$ can be ignored. For example,

$$M_1 = p ,$$

$$M_2 = p^2 + p(1 - p)\alpha ,$$

$$M_3 \doteq p^3 + 3p^2(1 - p)\alpha ,$$

$$M_4 \doteq p^4 + 6p^3(1 - p)\alpha .$$

Exact expressions of the moments in terms of gene frequencies and the inbreeding coefficient can be written if a distribution function of isolates is known. For instance, one or two parameter probability functions such as binomial, Poisson, normal, exponential, gamma, and beta distributions have been applied to this case, the beta probability being especially interesting because it corresponds to a steady-state distribution of gene frequency (Wright, 1931) (Appendix II). All these cases indicate that a population moment can be expressed as a polynomial of α, with the quadratic and higher order powers negligible when α is not greater than p or $1 - p$.

* $O(x)$ stands for "any function which is at most of order x."

However, as stated previously, it is extremely difficult to determine the distribution of isolate size in human populations. It is necessary, therefore, to approach this problem without knowing any distribution function. In the general argument above, we assumed that moments higher than the quadratic of Δp_w are negligible. This limiting form is valid, provided that all gene frequencies exceed the inbreeding coefficient, as is certainly the case for the polymorphisms to which this model will be applied. Extensive studies with known distribution forms have suggested that whenever isolate size distributions are symmetrical, then $m_3 = 0$ and $m_4 = O(\alpha^2)$. Even when asymmetric functions such as gamma are assumed, the limiting form holds with sufficient accuracy if the smallest gene frequency is greater than the inbreeding coefficient, which does not exceed 2% in human populations (Wright, 1950).

The population moments as a function of gene frequencies and the inbreeding coefficient can be obtained in the case of more than two alleles at a given locus (Appendix III).

Mating Type Frequency

As mentioned in the preceding section, mating type frequencies of a given genetic system can be obtained from the population moments. It is thus straightforward to evaluate the frequencies in the case of two alleles at an autosomal and a sex-linked locus (for autosomes, reciprocal crosses are grouped together). To illustrate, let us take the intercross $Aa \times Aa$ and its relative frequency fr. In an isolate, the proportion of this type is $4p_w^2(1 - p_w)^2 dW$, so that

$$\begin{aligned} fr &= \int 4p_w^2(1 - p_w)^2 dW \\ &= 4M_2 - 8M_3 + 4M_4 \\ &= 4p^2q^2 + 4pq(1 - 6pq)\alpha \,. \end{aligned}$$

Mating type frequencies and the proportions of the possible children in the limiting form are shown in Table 2 for an autosomal locus and in Table 3 for a sex-linked locus. In the latter case, we assumed that gene frequencies are the same in both sexes. Justification of the moment method to describe mating type frequencies is immediate when frequencies of possible offspring are evaluated as $p^2 + pq\alpha$, $2pq(1 - \alpha)$, and $q^2 + pq\alpha$ for genotype AA, Aa, and aa, respectively, at the autosomal locus without dominance as well as in the other cases. Mating type frequencies, when a distribution is assumed, are also calculated for autosomal and for sex-linked loci (Yasuda, 1966). (Incidentally, normal and rectangular distributions give exactly the same frequencies for sex-linked mating types.) When α approaches unity, incross frequencies go to corresponding gene frequencies with beta and binomial distributions, as happens also for genotype frequencies. For the other distributions examined, a convergency of incross frequencies to the gene frequencies fails when $\alpha \to 1$, since the distribution must condense into two poles at 0 and 1. This cannot be represented by one of these distributions. As our purpose is to describe human population structure, we are not going to consider further the case when $\alpha \to 1$.

Dominance does not create any difficulty in obtaining phenotype mating type

frequencies, since it requires simple additions of terms of genotype mating type frequencies whose phenotypes are the same (Tables 2 and 3).

The main effect of inbreeding on frequencies of zygotes is a decrease in heterozygosity. The effects of inbreeding are greatest when both genes are of equal frequency, but of course the maximum relative increase is seen as $p \to 0$.

This presentation will extend to mating type frequencies. Only a two-allelic locus will be discussed here, since the essential features of the inbreeding effect can be il-

TABLE 2

FREQUENCY OF MATING TYPES AND THEIR OFFSPRING
(TWO ALLELES AT AN AUTOSOMAL LOCUS)

A. NO DOMINANCE

MATING TYPE	FREQUENCY OF MATING TYPE	FREQUENCY OF OFFSPRING		
		AA	Aa	aa
$AA \times AA$	$p^4+6p^3q\alpha$	$p^4+6p^3q\alpha$	.....	.....
$AA \times Aa$	$4p^3q+12p^2q(1-2p)\alpha$	$2p^3q+6p^2q(1-2p)\alpha$	$2p^3q+6p^2q(1-2p)\alpha$	.....
$Aa \times Aa$	$4p^2q^2+4pq(1-6pq)\alpha$	$p^2q^2+pq(1-6pq)\alpha$	$2p^2q^2+2pq(1-6pq)\alpha$	$p^2q^2+pq(1-6pq)\alpha$
$AA \times aa$	$2p^2q^2+2pq(1-6pq)\alpha$	.....	$2p^2q^2+2pq(1-6pq)\alpha$	.....
$Aa \times aa$	$4pq^3+12pq^2(1-2q)\alpha$	.....	$2pq^3+6pq^2(1-2q)\alpha$	$2pq^3+6pq^2(1-2q)\alpha$
$aa \times aa$	$q^4+6pq^3\alpha$	.....	.....	$q^4+6pq^3\alpha$
Total	1	$p^2+pq\alpha$	$2pq-2pq\alpha$	$q^2+pq\alpha$

B. COMPLETE DOMINANCE

MATING TYPE	FREQUENCY OF MATING TYPE	FREQUENCY OF OFFSPRING	
		$A-$	aa
$A- \times A-$	$p^2(1+q)^2-2pq(1-3q^2)\alpha$	$p^2(1+2q)-3pq(1-2q)\alpha$	$p^2q^2+pq(1-6pq)\alpha$
$A- \times aa$	$2pq^2(1+q)+2pq(1-6q^2)\alpha$	$2pq^2+2pq(1-3q)\alpha$	$2pq^3+6pq^2(1-2q)\alpha$
$aa \times aa$	$q^4+6pq^3\alpha$	.....	$q^4+6pq^3\alpha$
Total	1	$1-q^2-pq\alpha$	$q^2+pq\alpha$

NOTE.—Where A and a are alleles with frequency p and q ($p+q=1$), respectively, and α is the inbreeding coefficient. It is assumed that $p, q > \alpha$.

lustrated by this case. Since mating type frequencies can be written in the form $R + I\alpha$ in the neighborhood of $\alpha = 0$, where R is the frequency in a panmictic population and I is the inbred component, we will examine I as a function of p in order to visualize inbreeding effects. Figures 1–4 are valid only when $p > \alpha$ and $q > \alpha$, although the graphs are presented for all possible values of p.

Autosome without dominance. Figure 1 gives the general features of the inbreeding effects on six different types of mating. Both incrosses ($AA \times AA$ and $aa \times aa$) are always increasing. Interestingly, both backcrosses ($AA \times Aa$ and $aa \times Aa$) decrease if the gene frequency is small and are compensatory to all other types of mating if $p < 0.212$ or $q < 0.212$. It is clear that the inbreeding effect is more striking in back-

TABLE 3

FREQUENCY OF MATING TYPES AND THEIR OFFSPRING
(TWO ALLELES AT A SEX-LINKED LOCUS)

A. NO DOMINANCE

MATING TYPE	FREQUENCY OF MATING TYPE	FREQUENCY OF FEMALE OFFSPRING			FREQUENCY OF MALE OFFSPRING	
		AA	Aa	aa	A	a
$AA \times A$	$p^3+3p^2q\alpha$	$p^3+3p^2q\alpha$	.....	.....	$p^3+3p^2q\alpha$	.....
$Aa \times A$	$2p^2q+2pq(1-3p)\alpha$	$p^2q+pq(1-3p)\alpha$	$p^2q+pq(1-3p)\alpha$	.....	$p^2q+pq(1-3p)\alpha$	$p^2q+pq(1-3p)\alpha$
$AA \times a$	$p^2q+pq(1-3p)\alpha$	.....	$p^2q+pq(1-3p)\alpha$	.....	$p^2q+pq(1-3p)\alpha$	.....
$Aa \times a$	$2pq^2+2pq(1-3q)\alpha$	.....	$pq^2+pq(1-3q)\alpha$	$pq^2+pq(1-3q)\alpha$	$pq^2+pq(1-3q)\alpha$	$pq^2+pq(1-3q)\alpha$
$aa \times A$	$pq^2+pq(1-3q)\alpha$	.....	$pq^2+pq(1-3q)\alpha$	.....	.....	$pq^2+pq(1-3q)\alpha$
$aa \times a$	$q^3+3pq^2\alpha$	.....	.....	$q^3+3pq^2\alpha$	.....	$q^3+3pq^2\alpha$
Total	1	$p^2+pq\alpha$	$2pq-2pq\alpha$	$q^2+pq\alpha$	p	q

B. COMPLETE DOMINANCE

MATING TYPE	FREQUENCY OF MATING TYPE	FREQUENCY OF FEMALE OFFSPRING		FREQUENCY OF MALE OFFSPRING	
		$A-$	aa	A	a
$A- \times A$	$p^2(1+q)+pq(2-3p)\alpha$	$p^2(1+q)+pq(2-3q)\alpha$	.....	$p^2+pq\alpha$	$p^2q+pq(1-3p)\alpha$
$A- \times a$	$pq(1+q)-3pq^2\alpha$	$pq-pq\alpha$	$pq^2+pq(1-3q)\alpha$	$pq-pq\alpha$	$pq^2+pq(1-3q)\alpha$
$aa \times A$	$pq^2+pq(1-3q)\alpha$	$pq^2+pq(1-3q)\alpha$	.....	.....	$pq^2+pq(1-3q)\alpha$
$aa \times a$	$q^3+3pq^2\alpha$	.....	$q^3+3pq^2\alpha$	.....	$q^3+3pq^2\alpha$
Total	1	$1-q^2-pq\alpha$	$q^2+pq\alpha$	p	q

NOTE.—Where A, a and p, q ($p+q=1$) are alleles and their corresponding frequencies respectively, α is the inbreeding coefficient. It is assumed that $p, q > \alpha$.

crosses than incrosses. As an extreme case, when $p = 0.18$, the I value for backcross $aa \times Aa$ reaches a minimum of -0.93. If we take $\alpha = 0.006$, which is a rather high value in man, the decreasing proportion due to inbreeding is $0.93 \times 0.006 = 0.0056$, which is $0.0056/0.3970 = 0.014$, or about 2% of the mating type frequency calculated from the Hardy-Weinberg law. Thus, an assumption of random mating for the estimation of gene frequencies in polymorphic systems might be justified as a good approximation.

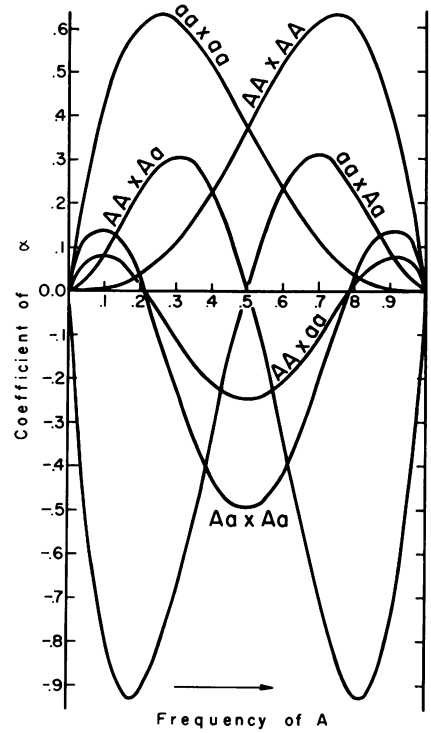

FIG. 1.—Effect of inbreeding on mating type (autosome). Two alleles, A and a, without dominance.

Autosome with complete dominance. In Figure 2, $A-$ denotes the dominant phenotype. The effect of inbreeding is largest when the frequency of the dominant gene is nearly 0.25. The cross of both dominant phenotypes ($A- \times A-$) compensates for the other two matings if $p > 0.577$.

Sex-linked without dominance. The general tendency of inbreeding effects is similar to that of an autosomal locus without dominance. The effects are rather weaker in a sex-linked than in an autosomal locus (Fig. 3).

Sex-linked with complete dominance. The effect of inbreeding is symmetrical when mating type frequencies in the population are classified by male or hemizygote (Fig. 4).

In summary, the frequency of incrosses is always enhanced by inbreeding, as are homozygotes. The effect on the other type of crosses depends on the gene frequency.

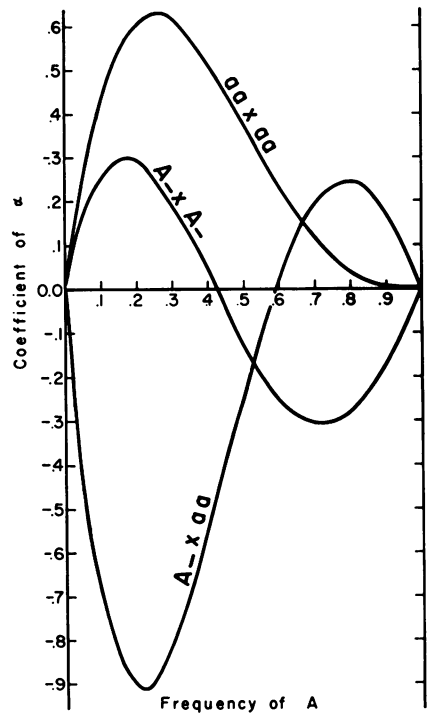

FIG. 2.—Effect of inbreeding on mating type (autosome). Two alleles, *A* and *a*, with complete dominance.

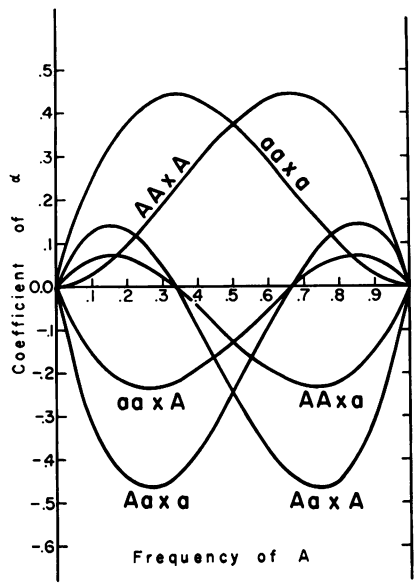

FIG. 3.—Effect of inbreeding on mating type frequencies (sex-linked). Two alleles, *A* and *a*, without dominance.

Roughly speaking, the effects of inbreeding or subdivision of population on mating type frequencies are magnified when the gene frequency is nearly 0.25 or 0.75 instead of 0.5. These predictions can be tested immediately with mating type frequencies in man in the case of antigenic polymorphisms and serum variations (Yasuda, 1966).

When there are three alleles at a locus, the numbers of possible genotypes and mating types are six and 21, respectively; and, with ten alleles, the corresponding values become 55 and 1,540. If a, g, and m stand for the numbers of alleles, genotypes, and

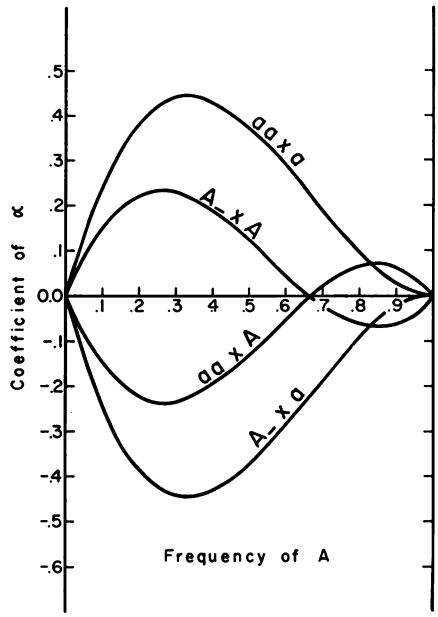

FIG. 4.—Effect of inbreeding on mating type frequencies (sex-linked). Two alleles, A and a, with complete dominance.

mating types, respectively, then $m = g(g + 1)/2 = a(a + 1)(a^2 + a + 2)/8$ for autosomal, and $m = ag = a^2(a + 1)/2$ for sex-linked loci. In these large numbers of mating types, however, there are only seven functionally different types which can be made when one has a set of at least four multiple alleles at an autosomal locus and four types with at least three alleles at a sex-linked locus. They are tentatively called incross ($AA \times AA$), backcross ($AA \times AB$), intercross ($AB \times AB$), outcross ($AA \times BB$), 3-way-intercross ($AB \times AC$), 3-way-outcross ($AA \times BC$), and 4-way-intercross ($AB \times CD$) for an autosomal locus; and incross ($AA \times A$), outcross ($AA \times B$), backcross ($AB \times A$), and intercross ($AB \times C$) at a sex-linked locus, where A, B, C, and D denote different alleles. Dominance between alleles will diminish the number of mating types. These mating type frequencies are also derived from the population moments (Tables 4 and 5).

DISCUSSION AND PROBLEMS IN ASCERTAINING THE MEAN INBREEDING COEFFICIENT IN MAN

The most common way to ascertain the inbreeding coefficient in human populations is to classify marriages into known degrees of inbreeding and to take their average weighted by the observed numbers (Haldane and Moshinsky, 1939). This method is called a "pedigree study;" it requires a complete knowledge of pedigrees and assumes the nominal coefficient of consanguinity to be equal to the inbreeding coefficient.

TABLE 4

FREQUENCY OF THE SEVEN FUNCTIONALLY DIFFERENT MATING TYPES (AUTOSOMAL)

Mating Type	Frequency
Incross ($AA \times AA$)	$p_A^4 + 6p_A^3(1 - p_A)\alpha$
Backcross ($AA \times AB$)	$4p_A^3p_B + 12p_A^2p_B(1 - 2p_A)\alpha$
Outcross ($AA \times BB$)	$2p_A^2p_B^2 + 2p_Ap_B(p_A + p_B - 6p_Ap_B)\alpha$
Intercross ($AB \times AB$)	$4p_A^2p_B^2 + 4p_Ap_B(p_A + p_B - 6p_Ap_B)\alpha$
3-Way outcross ($AA \times BC$)	$4p_A^2p_Bp_C + 4p_Ap_Bp_C(1 - 6p_A)\alpha$
3-Way intercross ($AB \times AC$)	$8p_A^2p_Bp_C + 8p_Ap_Bp_C(1 - 6p_A)\alpha$
4-Way intercross ($AB \times CD$)	$8p_Ap_Bp_Cp_D(1 - 6\alpha)$

NOTE.—A, B, ... and p_A, p_B, ... are alleles and their corresponding frequencies, respectively ($p_A + p_B + \ldots = 1$); α is the inbreeding coefficient. It is assumed that $p_A, p_B, \ldots > \alpha$.

TABLE 5

FREQUENCY OF FOUR FUNCTIONALLY DIFFERENT MATING TYPES (SEX-LINKED)

Mating Type	Frequency
Incross ($AA \times A$)	$p_A^3 + 3p_A^2(1 - p_A)\alpha$
Outcross ($AA \times B$)	$p_A^2p_B + p_Ap_B(1 - 3p_A)\alpha$
Backcross ($AB \times A$)	$2p_A^2p_B + 2p_Ap_B(1 - 3p_A)\alpha$
Intercross ($AB \times C$)	$2p_Ap_Bp_C(1 - 3\alpha)$

NOTE.—A, B, C and p_A, p_B, p_C are alleles and their corresponding frequencies, respectively ($p_A + p_B + p_C + \ldots = 1$); α is the inbreeding coefficient. It is assumed that $p_A, p_B, \ldots > \alpha$.

However, this does not cover unrecognized remote consanguinity. For instance, under favorable circumstances, the ascertainment of consanguinity can extend several generations into the past. In some areas, records of Roman Catholic marriage dispensations go back hundreds of years (Moroni, 1962). Under these conditions, ascertained consanguinity is likely to account for a large fraction of the total inbreeding coefficient. Formally, we may represent the situation as

$$\alpha_T = \alpha + \alpha_R , \qquad (1)$$

where α_T, α, and α_R denote the inbreeding coefficient due to total consanguinity, ascertained consanguinity, and undetected remote consanguinity, respectively. Unfortunately, as we go backward in time, the proportion of ancestors who were migrants increases, so that ascertainment of consanguinity will always be incomplete even for populations with extensive marriage records.

Although we may hope that α_R/α is small, doubt arises even in the most favorable cases. For example, birth records in the Alpine village of Bosco-Gurin permit recon-

struction of pedigrees for ten generations (Moor-Jankowski and Huser, 1957). There was little migration into the village. We might expect that all important consanguinity had been ascertained. But, in fact, history shows that the villages migrated into the area in the thirteenth and fourteenth centuries from the Valais. It is likely that inbreeding during the ages before the birth records began had effects on gene frequencies which are still appreciable and contribute to the α_R of Switzerland.

Since no system of records, however complete, can ascertain the total inbreeding coefficient, we must look for other ways. There are two approaches to pursue the remote inbreeding coefficient: use of a biological indicator (*bioassay*) and of migration and inbreeding functions with distance (*correlation method*). In both methods, the remote inbreeding coefficient is calculated as the difference of the total inbreeding and the close inbreeding ascertained by pedigree analysis. In this connection, we define remote consanguinity as a relationship more distant than a first cousin once removed ($F < 1/32$).

The mean inbreeding coefficient can be estimated from individual phenotype and mating type frequencies. Differential selection, illegitimacy, and misclassification are the main sources to disturb an accurate estimate of the inbreeding coefficient, and, generally speaking, they affect phenotype frequencies more than mating type frequencies. Differential selection, especially against homozygotes, might tend to give smaller or even a negative estimate of the inbreeding coefficient. Illegitimacy or misclassification has, in a statistical sense, the same effect on the biological indicator as selection does. And genes whose frequencies are relatively small are excluded from the probability models for mating types and should be pooled with more common alleles to meet the restriction $p > \alpha$.

Sanghvi (1955) reported the insensitiveness of inbreeding on genotype frequencies, and Schull (1965) pointed out the instability of phenotype frequencies used to estimate the inbreeding coefficient. Regardless of these statistical difficulties which have been shown mathematically (Yasuda, 1967*a*), there is no such trouble in estimating the inbreeding coefficient from mating type data (Yasuda, 1967*b*).

Use of migration functions, $m(x)$, defined as the probability among all marriages that the marital distance is x, requires determination of the function and evaluation of the genetic correlation, $f(x)$, of children whose parents had a marital distance x. If these two functions with distance are found, the mean inbreeding coefficient is calculated by

$$\alpha = \int_0^\infty f(x)\, m(x)\, dx .$$

Human migration does not follow a normal distribution expected from dispersion of genes by a diffusion process (Cavalli-Sforza, 1958). This is not surprising because of the many barriers which prevent random combination of gametes. Thus, at present, a choice of a migration function is not completely specified except (1) the function is leptokurtic, (2) the proportion of near-zero distances should be finite, and (3) the function should be mathematically and statistically simple. Under these conditions, some promising functions are exponential, square root exponential (Cavalli-Sforza, 1958), log normal, beta, bi-exponential, etc. A gamma function that includes an

exponential distribution as a special case has been fitted to a northern Italian population (Cavalli-Sforza, 1962). The fit is good, but the estimate of the dimension parameter is always less than one, so that $m(0)$ tends to be infinite. This is unrealistic. It is expected, however, that no distribution would fit well because of a practical difficulty in estimating frequency near zero distance.

The genetic correlation with distance, $f(x)$, is more intricate. This can be derived if the migration function is known (Malecot, 1948; M. Kimura, personal communication), but it seems that the assumption of a migration function is not necessary (Malecot, 1950; Kimura and Weiss, 1964). A difficulty in practice is the fact that $f(x)$ depends on the dimension of human migration. Fortunately, $f(x)$ can be determined empirically as a gene frequency correlation with respect to locality:

$$f(x) = \frac{\Sigma(p_y - p)(p_{x+y} - p)}{p(1-p)} \qquad \text{(Malecot, 1955)},$$

where the summation is taken over y location with gene frequencies p_y. This method has been applied to Switzerland (Yasuda and Morton, 1966).

The separation of the total inbreeding coefficient into contributions due to ascertained and remote consanguinity involves an important concept of population structure. Wahlund's principle tells that, if random mating is assumed within isolates, the inbreeding coefficient due to barriers is always positive, since the coefficient is defined with respect to gene frequency variance. The more barriers there are, the higher is the inbreeding coefficient expected. However, all barriers would not be ascertained in practice. If F_i designates the ascertained inbreeding coefficient by the ith level procedure (for instance, the first level may be due to ascertainment of close consanguinity less distant than second cousin once removed, the second level up to known consanguinity, etc.), then the total inbreeding coefficient a_T can be obtained from

$$a_T = \Sigma F_i .$$

However, the assumption of random mating within isolates may not be justified in a particular situation. For example, suppose an isolate consists of two types of homozygotes, AA and aa, and mating occurs only between different genotypes. Obviously, the inbreeding coefficient for the isolate is not zero but minus one in the sense of a negative correlation between uniting gametes. This leads to

$$a_T = \Sigma F_i + r , \qquad (2)$$

where r is the correlation coefficient due to nonrandom mating in isolates, and the following relation holds:

$$-1 \leqq r \leqq 0 \leqq \Sigma F_i \leqq 1 .$$

In other words, all positive correlations of uniting gametes are considered due to genetic barriers which might have been generated by random genetic drift and geographical, sociological, and other factors. In practice, however, the ascertainment of ΣF_i is dependent on technique, so that

$$a_T = \sum_A F_i + \left(\sum_{T-A} F_i + r \right),$$

where

$$\sum_{A} \quad \text{and} \quad \sum_{T-A}$$

mean summations of ascertained inbreeding coefficients with respect to the level of procedure and of unascertained positive correlation between uniting gametes, respectively. This is equivalent to (1) if we put

$$a = \sum_{A} F_i$$

for the ascertained inbreeding and

$$a_R = \left(\sum_{T-A} F_i + r \right)$$

for the remote consanguinity. a_T and a_R can be negative if the negative correlation in isolates is high.

An alternative model has been proposed by Wright (1943) for consideration of breeds of cattle. If a population has hierarchic structure, the total inbreeding coefficient, F_{IT}, is related to the inbreeding coefficient within a subpopulation, F_{IS}, and due to subdivision, F_{ST}, in the following manner:

$$1 - F_{IT} = (1 - F_{IS})(1 - F_{ST}) , \tag{3}$$

which can be extended into any degree of hierachic structure or

$$1 - F_{IT} = \prod_{i} (1 - F_{Si}) .$$

This relation can be deduced from the moments of the population (Appendix IV).

It is thus obvious that a hierarchic pattern of barriers is specified in (3), whereas no such scheme is made in (2). Genetic barriers are hard to recognize in human populations, while it is rather easy to set up such a model in experimental populations like cattle. F_{ST} and F_{IS} should be always positive with respect to genetic barriers and from a probabilistic viewpoint, but Wright (1951, 1965) stated that F_{IS} could be negative. This is true only when mating is not random within a "basic" unit of population, since the size is small. In this situation, F_{IS} corresponds to r in our terminology, and, whenever the F value becomes negative, the independence between the system of mating and gene frequency breaks down, since no homozygote frequency can be less than zero. Therefore r must be near zero in human populations. This implies that each gene in an isolate has potentially an equal probability to unite with its neighbors in the sense of probability density. On the other hand, the identification of isolates is almost impossible in man without knowledge of the "original composition of the population." At present we do not have any method to evaluate r, so we must assume it to be zero for human populations. Further research is desirable.

The hierarchic discription may be a good approximation of population structure. Taking the logarithm of a general form of (3) and expanding in series, we obtain approximately

$$F_{IT} = \Sigma F_{Si} + F_{IS} ,$$

where F_{IS} is zero if mating in isolates is at random, otherwise: $-1 \leqq F_{IS} < 0$.

Ultimately, the inbreeding coefficient should be determined by balance between mutation and reproduction of genes in the course of evolution, following the direction set by Malecot (1948).

SUMMARY

A new theory for describing human population structure has been proposed by replacing the concept of isolate size, in Dahlberg's sense, or neighborhood size, in Wright's sense, by an idea of probability density of a gene, or a tendency that a gene shall combine with its neighbors in order to form the genotype, the mating type, and other gene combinations. These genetic quantities can be described in terms of moments of the population whose order corresponds to the number of genes combined. The main result when the inbreeding coefficient is not greater than the smallest gene frequency is that mating type frequencies are given as a function of gene frequencies and the inbreeding coefficient both at autosomal and sex-linked loci.

A method of estimating the total, ascertained, and remote inbreeding coefficients has been discussed. Two components in describing a system of mating in terms of the correlation coefficient between uniting gametes should be distinguished: positive and negative correlations. The positive correlation, which measures effects of genetic barriers on combinations of genes, consists of ascertained and unascertained consanguinity. The negative correlation, which may be observed in a small population, is also included in the unascertained inbreeding coefficient. The ascertained inbreeding coefficient consists of positive correlations, and the remote inbreeding may include both components. The total inbreeding coefficient is thus due to contributions from both close and remote consanguinity.

Comparison with Wright's hierarchic structure of population is also discussed.

ACKNOWLEDGMENT

The author would like to express his sincere thanks to Dr. Newton E. Morton for his interest, supervision, and encouragement throughout this work and for his helpful advice and stimulating discussion in the preparation of the manuscript. He also wishes to thank Dr. Ming-Pi Mi for reading the manuscript.

APPENDIX I

BREAKUP OF ISOLATE

The conclusion in the text is not altered when we consider more than three isolates. Although several models may be developed, we shall discuss only two of them: (1) breakup of isolates in a part of the total population and (2) a hierarchic model of removing barriers.

1. Suppose that a population consisted of n isolates, and k of the $n(k < n)$ isolates were grouped into a new panmictic isolate by removing barriers so that the population now consists of $n - k$ isolates.

At the first phase, the population is characterized by:

$$\sum_{i=1}^{n} w_i = 1 ,$$

$$p_N = \sum_{i=1}^{n} p_i w_i ,$$

and

$$\sigma_N^2 = \sum_{i-1}^{n} p_i^2 w_i - p_N^2 ,$$

where w_i and p_i are the relative size and gene frequency of the ith isolate, respectively, and also, p_N and σ_N^2 are the gene frequency and its variance in the total population. Let us assume that the first k isolates are grouped. The present population is now specified by

$$W + \sum_{i=k+1}^{n} w_i = 1 ,$$

$$p_{N-K} = PW + \sum_{i=k+1}^{n} p_i w_i ,$$

and

$$\sigma_{N-K}^2 = P^2 W + \sum_{i=k+1}^{n} p_i^2 w_i - p_{N-K}^2 ,$$

where

$$W = \sum_{i=1}^{k} w_i , \qquad P = \sum_{i=1}^{k} p_i w_i / W ,$$

and p_{N-K} and σ_{N-K}^2 stand for gene frequency and its variance in the present population.

Comparison of the two phases results in

$$p_N = p_{N-K} (= p) ,$$

and

$$\sigma_{N-K}^2 = P^2 W + \sum_{i=k+1}^{n} p_i^2 w_i - p_{N-K}^2$$

$$= \sigma_N^2 - \left[\left(\sum_{i=1}^{k} w_i \right) \left(\sum_{i=1}^{k} p_i^2 w_i \right) - \left(\sum_{i=1}^{k} p_i w_i \right)^2 \Big/ \sum_{i=1}^{k} w_i \right] ,$$

or

$$\sigma_{N-K}^2 = \sigma_N^2 - \sum_{i>j}^{k} w_i w_j (p_i - p_j)^2 \Big/ \sum_{i=1}^{k} w_i .$$

The relation among the inbreeding coefficients will be

$$F_{N-K} = F_N - F_B ,$$

where

$$F_B = \sum_{i>j}^{k} \frac{w_i w_j}{W} \frac{(p_i - p_j)^2}{p(1-p)} .$$

2. Although the breakup of isolates has occurred in several parts of the population, some barriers still remain, so that the population consists of a number of new isolates. We can easily verify the relation among the inbreeding coefficients between the two phases of population: $F_{N-\ldots-X} = F_N - F_B$, where $F_{N-\ldots-X}$ stands for the inbreeding coefficient at the second phase, and

$$F_B = \sum \sum_{i>j} \left(\frac{w_i w_j}{W}\right) \frac{(p_i - p_j)^2}{p(1-p)}.$$

The first summation is taken for new isolates.

Example. The comparison of the endogamy coefficient in three racial ancestral populations with the inbreeding coefficient of a triracial mixture population: Suppose that three racial groups are Indian, Negro, and Caucasian whose relative sizes I, N, and C are p_i, p_n, and p_c, respectively. Let F_M and F_E be the inbreeding coefficient of the triracial mixture population and the endogamy coefficient in the ancestral populations. We obtain

$$\begin{aligned} F_B &= F_E - F_M \\ &= \frac{IC(p_i - p_c)^2 + CN(p_c - p_n)^2 + NI(p_n - p_i)^2}{p(1-p)}, \end{aligned}$$

since $p = Ip_i + Np_n + Cp_c$ and $I + N + C = 1$. Consideration of several genes will provide more information about F_B.

In the above discussion, we assumed no mutation, selection, and accidents of sampling which might change the mean gene frequency of the population. Since it is always difficult to estimate gene frequencies in ancestral populations, F_B may be taken as a first approximation. Examining the mean gene frequency of populations at different stages, we may justify the method if the difference from zero is not significant.

APPENDIX II

MOMENTS OF A SUBDIVIDED POPULATION, GIVEN A DISTRIBUTION OF ISOLATE SIZE

A general idea to obtain the moments knowing a distribution function of isolates can be demonstrated by the beta distribution in a locus with two alleles, since this case might represent a steady state distribution of gene frequency under Wright's island model, where the population consists of isolates of equal size and each isolate constantly exchanges individuals with the total gene pool. Only the third and fourth moments will be given in the other distributions for the sake of comparison. For the higher moments, a method of moment generating function would be helpful.

Beta Distribution

Suppose that a density function is given by

$$dW = \frac{(a+b-1)!}{(a-1)!(b-1)!} p_w^{a-1} q_w^{b-1} dp_w \qquad (p_w + q_w = 1),$$

where a and b are distribution parameters and p_w is the gene frequency in the neighborhood of point w. The moment of population will be

$$M_k = \int p_w^k dW$$

$$= \frac{(a+b-1)!}{(a-1)!(b-1)!} \int_0^1 p_w^{a+k-1}(1-p_w)^{b-1} dp_w$$

$$= (a+b-1)!(a+k-1)!/(a-1)!(a+b+k-1)!$$

$$(k = 0, 1, \ldots),$$

which gives

$$M_1 = a/(a+b) \qquad \text{and} \qquad M_2 = M_1(a+1)/(a+b+1).$$

Since the first and second moments correspond to the population gene frequency p and homozygous frequency $p^2 + pq\alpha$, where α is the inbreeding coefficient, the parameters can be written in terms of gene frequency and the inbreeding coefficient or

$$a/(a+b) = p,$$

$$(a+1)/(a+b+1) = p + q\alpha,$$

so that $a = p(1-\alpha)/\alpha$ and $b = q(1-\alpha)/\alpha$.

We obtain, therefore,

$$M_3 = \frac{1}{(1+\alpha)}\,[p^3 + p^2(1+2q)\alpha + pq(1+q)\alpha^2],$$

and

$$M_4 = \frac{1}{(1+\alpha)(1+2\alpha)}\,[p^4 + 3p^3(1+q)\alpha + p^2(2+6q+3q^2)\alpha^2 + pq(1+q)(2+q)\alpha^3],$$

or, if α is small (say less than 2%),

$$M_3 = p^3 + 3p^2q\alpha - pq(1+3q)\alpha^2 + 2q(1-2q)\alpha^3 + \ldots,$$

and

$$M_4 = p^4 + 6p^3q\alpha - p^2q(8-19q)\alpha^2 + 2pq(7-27q+23q^2)\alpha^3 + \ldots.$$

In the island model, $a = 4Nmp$ and $b = 4Nmq$, where N is the effective size of isolates and m is migration rate. Thus $\alpha = 1/(1+4Nm)$ (Wright, 1931).

The following results are straightforward. (The form of the distribution function may be found in Mood and Graybill, 1963.)

Binomial Distribution

$$M_3 = p^3 + 3p^2q\alpha - pq(1-2q)\alpha^2,$$

$$M_4 = p^4 + 6p^3q\alpha - p^2q(4-11q)\alpha^2 + pq(1-6q+6q^2)\alpha^3.$$

Poisson Distribution

$$M_3 = p^3 + 3p^2q\alpha + pq^2\alpha^2,$$

$$M_4 = p^4 + 6p^3q\alpha + 7p^2q^2\alpha^2 + pq^3\alpha^3.$$

Rectangular Distribution

$$M_3 = p^3 + 3p^2q\alpha ,$$

$$M_4 = p^4 + 6p^3q\alpha + (9/5)p^2q^2\alpha^2 .$$

Normal Distribution

$$M_3 = p^3 + 3p^2q\alpha ,$$

$$M_4 = p^4 + 6p^3q\alpha + 3p^2q^2\alpha^2 .$$

Gamma Distribution ($p < q$)

$$M_3 = p^3 + 3p^2q\alpha + 2pq^2\alpha^2 ,$$

$$M_4 = p^4 + 6p^3q\alpha + 11p^2q^2\alpha^2 + 6pq^3\alpha^3 .$$

Thus the square and higher powers of α may be ignored when $|\alpha|$ is not greater than the smallest gene frequency.

APPENDIX III

DERIVATION OF A GENERAL FORMULA FOR THE MOMENT OF POPULATION

When the number of alleles increases beyond two, the covariance moments, which are given in Wahlund's principle as the frequency of the heterozygote, become important.

Let p and q be population gene frequencies and p_w and q_w be those of an isolate. (It is not required that $p + q = 1$ and $p_w + q_w = 1$.) Denoting differences in gene frequencies between the isolate and the population by Δp_w and Δq_w, their covariance moment is given by

$$m_{ij} = \int (\Delta p_w)^i (\Delta q_w)^i dW \equiv E(\Delta p_w)^i (\Delta q_w)^j ,$$

where E is an operational symbol denoting expectation. For example,

$$m_{10} = m_{01} = 0 ,$$

$$m_{20} = p(1 - p)\alpha , \qquad m_{11} = -pq\alpha \qquad \text{and} \qquad m_{02} = q(1 - q)\alpha ,$$

where α is the inbreeding coefficient. The moment of population is now

$$\begin{aligned} M_{a,b} &= E(p_w^a p_w^b) \\ &= E(p + \Delta p_w)^a (q + \Delta q_w)^b \\ &= \sum_{r,s} \binom{a}{r}\binom{b}{s} p^{a-r} q^{b-s} m_{rs} \\ &= p^a q^b + a\, p^{a-1} q^b m_{10} + b\, p^a q^{b-1} m_{01} + \binom{a}{2} p^{a-2} q^b m_{20} + a\, b\, p^{a-1} q^{b-1} m_{11} \\ &\qquad + \binom{b}{2} p^a q^{b-2} m_{02} + \cdots , \end{aligned}$$

or, ignoring the higher terms of $m_{r,s}(r+s>2,\ r\geqq 0 \text{ and } s\geqq 0)$,

$$M_{a,b}=p^aq^b+\left[\frac{a(a-1)}{2}\,p^{a-1}(1-p)\,q^b+\frac{b(b-1)}{2}\,p^aq^{b-1}(1-q)-ab\,p^aq^b\right]\alpha .$$

Justification to ignore the higher moments, m_{rs}, is also seen in a series of calculations for moments assuming distribution functions. The sufficient condition is again that the smallest gene frequency is greater than the inbreeding coefficient.

Generalization is now straightforward. The moment of population is

$$M_{a_1,\ldots,a_n}=E\left[\prod_{i=1}^{n}(p_i+\Delta p_{w_i})^{a_i}\right],$$

where p_i is the frequency of ith allele. By expansion of binomial product terms and by replacing m_{rs} in terms of gene frequencies and the inbreeding coefficient, we obtain

$$M_{a_1,\ldots,a_n}=\prod_{i=1}^{n}p_i^{a_i}+\left(\prod_{i=1}^{n}p_i^{a_i-1}\right)\left[\sum_{i=1}^{n}\binom{a_i}{2}(1-p_i)\left(\prod_{j\neq i}p_j\right)\right.$$

$$\left.-\left(\prod_{i=1}^{n}p_i\right)\left(\sum_{i>j}a_ia_j\right)\right]\alpha .$$

The assumptions for deriving the general formula are the same as the previous arguments.

APPENDIX IV

A PROOF THAT WRIGHT'S HIERARCHIC STRUCTURE CAN BE CONSIDERED AS A SPECIAL CASE OF THE GENERALIZED WAHLUND'S PRINCIPLE

Suppose that a population consists of isolates whose size and gene frequency are

$$w_{ij}\left(\sum_i\sum_j w_{ij}=1\right)$$

and p_{ij}, respectively, and within which mating is at random. Then, the total frequency of homozygotes in question is

$$\sum_i\sum_j p_{ij}^2w_{ij}=p^2+p(1-p)F_{IT},$$

where

$$p=\sum_i\sum_j p_{ij}w_{ij}.$$

On the other hand, when we consider barriers with respect to i, within the ith aggregate of isolates, the homozygote frequency is then

$$\sum_j p_{ij}^2w_{ij}=[\,p_i^2+p_i(1-p_i)F_i\,]w_i,$$

where

$$p_i = \sum_j p_{ij} w_{ij}/w_i , \qquad w_i = \sum_j w_{ij} ,$$

and F_i is the inbreeding coefficient of the ith aggregate. The total homozygote frequency is therefore

$$\sum_i [\, p_i^2 + p_i(1 - p_i)F_i \,] w_i = \Sigma p_i^2 w_i + \Sigma p_i(1 - p_i)F_i w_i ,$$

where

$$\Sigma p_i^2 w_i = p^2 + p(1 - p)F_{ST} ,$$

so that

$$p^2 + p(1 - p)F_{IT} = p^2 + p(1 - p)F_{ST} + \Sigma p_i(1 - p_i)F_i w_i ,$$

or

$$F_{IT} = F_{ST} + \frac{\Sigma F_i p_i(1 - p_i)w_i}{p(1 - p)} .$$

If the inbreeding coefficients for all aggregates are the same or $F_i = F_{IS}$ for all i, we obtain

$$F_{IT} = F_{ST} + F_{IS}(1 - F_{ST}) ,$$

which is equivalent to (3) in the text.

The corresponding result can also be shown with respect to heterozygote frequency.

REFERENCES

BARRAI, I., CAVALLI-SFORZA, L. L., and MORONI, A. 1962. Frequencies of pedigrees of consanguineous marriages and mating structure of the population. *Ann. Hum. Genet.* (London) **25**:347–376.

BERNSTEIN, F. 1930. Fortgesetzte Untersuchungen aus der Teorie der Blutgruppen. *Z. indukt. Abstamm. Vererb. Lehr.* **56**:223–273.

CAVALLI-SFORZA, L. L. 1958. Some data on the genetic structure of human populations. *Proc. Tenth Int. Cong. Genet.* **1**:389–407.

CAVALLI-SFORZA, L. L. 1962. The distribution of migration distance: models and applications to genetics. Pp. 139–166 in J. SUTTER (ed.), *Les deplacements humains, entretien de Monaco en sciences humaines*. Hachette, Paris.

CAVALLI-SFORZA, L. L., KIMURA, M., and BARRAI, I. 1966. The probability of consanguineous marriages. *Genetics* **45**:37–60.

CRAMER, H. 1946. *Mathematical methods of statistics*. Princeton Univ. Press, Princeton, N.J.

HAJNAL, J. 1963. Concepts of random mating and the frequency of consanguineous marriages. *Proc. Roy. Soc.* (London), B. **159**:125–177.

HALDANE, J. B. S., and MOSHINSKY, P. 1939. Inbreeding in Mendelian populations with special reference to human cousin marriage. *Ann. Eugen.* (Cambridge) **9**:321–340.

HOLGATE, P. 1964. Genotype frequencies in a section of a cline. *Heredity* **19**:507–509.

KIMURA, M., and WEISS, G. H. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**:561–576.

MALECOT, G. 1948. *Les Mathématique de l'hérédité*. Masson, Paris.

MALECOT, G. 1950. Quelques schémas probabilistes sur la variabilité des populations naturelles. *Ann. Univ. Lyon* 13. Sec. A. Pp. 37–60 in *Sciences mathématiques et astronomie*. Masson, Paris.

MALECOT, G. 1955. The decrease of relationship with distance. *Cold Spring Harbor Symp. Quant. Biol.* **20**:52–53.

MOOD, A. M., and GRAYBILL, F. A. 1963. *Introduction to the theory of statistics*, 2d ed. McGraw-Hill, New York.

MOOR-JANKOWSKI, J. K., and HUSER, H. J. 1957. Sero-anthropological investigations in the Walser and Romansh isolates in the Swiss Alps and their methodological aspects. *Acta Genet. Stat. Med.* **6**:527–531.

MORONI, A. 1962. Sources, reliability and usefulness of consanguinity data with special reference to Catholic records. Paper presented at *U.N./W.H.O.* seminar on use of vital and health statistics for genetics and radiation studies.

MORTON, N. E. 1964. Genetic studies of northeastern Brazil. *Cold Spring Harbor Symp. Quant. Biol.* **29:** 69–79.

SANGHVI, L. D. 1955. Inbreeding, genes and phenotypes. *Amer. Nat.* **89**:247–248.

SCHULL, W. J. 1965. Estimation of genetic parameters in population studies. Pp. 45–60 in *Genetics and the epidemiology of chronic diseases*. J. V. NEEL, M. W. SHAW, and W. J. SCHULL (eds.), U.S. Dept. of Health, Education and Welfare, Washington.

WAHLUND, S. 1928. Zusammensetzung von Populationen und Korrelationserscheinungen von Standpunkt der Vererbungslehre aus Betrachet. *Hereditas* **11**:65–106.

WRIGHT, S. 1921. System of mating. *Genetics* **6**:111–178.

WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97–159.

WRIGHT, S. 1943. Isolation by distance. *Genetics* **28**:114–138.

WRIGHT, S. 1945. The differential equation of the distribution of gene frequencies. *Proc. Nat. Acad. Sci.* **31**:382–389.

WRIGHT, S. 1950. Discussion on population genetics and radiation. *J. Cell .Comp. Physiol.* **35** (Suppl. 1): 187–205.

WRIGHT, S. 1951. The genetical structure of population. *Ann. Eugen.* (London) **15**:323–354.

WRIGHT, S. 1965. The interrelation of population structure by F-statistics with special regard to system of mating. *Evolution* **19**:395–420.

YASUDA, N. 1966. The genetical structure of northeastern Brazil. Ph.D. thesis. Univ. Hawaii.

YASUDA, N. 1967*a*. Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. (In preparation.)

YASUDA, N. 1967*b*. The genetical structure of northeastern Brazil. (In preparation.)

YASUDA, N., and MORTON, N. E. 1966. Studies on human population structure. Pp. 98–104 *in* J. F. CROW and J. V. NEEL (eds.), *Third International Congress of Human Genetics*. Johns Hopkins, Baltimore, Md.