

THE EFFECT OF ASSORTATIVE MATING ON THE GENETIC COMPOSITION OF A POPULATION

JAMES F. CROW AND JOSEPH FELSENSTEIN

Laboratory of Genetics, University of Wisconsin,
Madison, Wisconsin; and Institute of Animal
Genetics, Edinburgh, Scotland

There are two classic papers on assortative mating, written by Sewall Wright (1921) and R. A. Fisher (1918). The approach in the two papers is quite different, although the general qualitative conclusions are similar. Fisher's paper is notoriously difficult to read, although this is remedied to some extent by the publication of an annotated version by Moran and Smith (1966).

Our object is mainly a review of these classic results. Included is a derivation of Fisher's main conclusions, using a method rather similar to Wright's. It is thereby possible to obtain Fisher's results using only elementary methods.

COMPARISON OF ASSORTATIVE MATING WITH INBREEDING

Assortative mating means that mated pairs are more similar for some phenotypic trait than would be expected if they were chosen at random from the population. Since individuals with similar phenotypes will usually be somewhat similar in their genotypes, we should expect assortative mating to have generally the same consequences as inbreeding. An excess of consanguineous mating has two consequences: (1) an increase in the average homozygosity; and (2) an increase in the total population variance. Assortative mating should give qualitatively similar results.

Assortative mating may have either of two causes, or some combination of both. The tendency toward phenotypic similarity of mating pairs may be a direct consequence of genetic relationship. For example, in a subdivided population there will generally

be a greater phenotypic similarity among the members of a subpopulation because they share a common ancestry. The genetic consequences in this case are the same as those of inbreeding. On the other hand, there may be assortative mating based on similarity for some trait, and any genetic relationship is solely a consequence of similar phenotypes. For example, there is a high correlation between husband and wife in height and intelligence, probably caused much more by nonrandom marriage associated with the trait itself than by common ancestry.

There are also other situations. For example, there is a considerable correlation in arm length between husband and wife. This is probably a consequence of the fact that those factors, genetic and environmental, that increase height also increase the length of the arm. So, any assortative mating for height will be reflected in a similar assortative mating for arm length, diminished somewhat by the lack of perfect correlation between the two traits. It may be that a population is mating assortatively for some traits and not for others at the same time. There may also be negative assortative, or disassortative, mating if opposites attract, but we shall consider mainly positive assortative mating.

In general, assortative mating causes less increase in homozygosity than inbreeding, especially if the trait involved is determined by several gene loci. On the other hand, assortative mating causes a large increase in the variance of a multifactorial trait, in contrast to that produced by a comparable amount of inbreeding. A further difference is

that inbreeding affects all segregating loci, whereas assortative mating affects only those related to the trait involved.

The variance-enhancing effect of assortative mating is apparent with a simple example. Suppose that an arbitrary quantitative trait is influenced by two loci without dominance. Let each gene with subscript 1 add one unit to the phenotype, whereas each gene with subscript 0 adds nothing. Then the genotype $A_1A_1 B_1B_1$ represents one extreme phenotype and $A_0A_0 B_0B_0$ the other, with $A_1A_1 B_0B_0$, $A_1A_0 B_1B_0$, and $A_0A_0 B_1B_1$ being exactly intermediate. Inbreeding will increase the frequency of all four homozygous genotypes, $A_1A_1 B_1B_1$, $A_0A_0 B_1B_1$, $A_1A_1 B_0B_0$, and $A_0A_0 B_0B_0$. This will increase the variance; in fact, it will exactly double the variance if the population is changed from random mating proportions to complete homozygosity.

On the other hand, with complete assortative mating, the population approaches a state where only the extreme homozygotes, $A_1A_1 B_1B_1$ and $A_0A_0 B_0B_0$, remain. This clearly causes a much greater enhancement of the variance, especially as the number of relevant loci is increased. The variance increase with assortative mating has been shown experimentally in *Nicotiana* (Breese 1956) and *Drosophila* (McBride and Robertson, 1963). The latter authors also found the expected decrease with disassortative mating and demonstrated that the rate of change under selection can be increased with assortative mating.

With inbreeding there is no systematic change in the frequencies of the gamete types, A_1B_1 , A_1B_0 , A_0B_1 , and A_0B_0 . On the other hand, as the example shows, assortative mating causes a change in frequency of the gametic types, increasing two while decreasing the other two. So, another way of describing the effect of assortative mating and of understanding its variance-enhancing effect is to note that it causes

gametic phase (or linkage) disequilibrium. In the words of Breese (1956, p. 342), assortative mating causes a "reassociation of genes in such a way that the excess of combinations giving extreme expression is achieved at the expense of more balanced combinations."

The simplest cases of complete assortative mating were worked out long ago by Jennings (1916) and by Wentworth and Remick (1916). As an example, consider a single locus with two alleles in which all three genotypes are distinguishable and assume that each genotype mates only with a genotype like itself. The genetic consequences are exactly the same as with self-fertilization; heterozygosity is reduced by half each generation and the variance is eventually doubled.

It might be thought from this example that assortative mating leads eventually to complete homozygosity, as do many forms of inbreeding, but this is not the case. Partial assortative mating, like partial self-fertilization, leads to an equilibrium level of heterozygosity other than zero.

In the more general treatment of assortative mating, two cases are of interest. At one extreme the individuals fall into two (or possibly more) discrete phenotypes with preference for mating within a phenotype. For example, deaf persons tend to marry others with the same trait. At the other extreme is a character, like size, for which there is a correlation between mates, but for which the distribution is continuous and determined by multiple genetic and environmental factors.

Before dealing with more complex multifactorial models, we shall first consider a single locus trait.

ASSORTATIVE MATING FOR A SINGLE LOCUS

With inbreeding, the choice of a mathematical model is clear from knowledge of the

relationship and the Mendelian mechanism. With assortative mating, the choice of models is not so obvious, as different behavior patterns can have different consequences.

We shall measure the degree of assortative mating by the product-moment correlation between parents. For a quantitative trait the correlation is directly measurable. For qualitative traits we measure the correlation coefficient as the decrease in the proportion of matings between dissimilar phenotypes divided by that which is expected with random pairs.

As a simple first example, assume that there are two alleles and that each genotype has a distinct phenotype. We let a fraction, r , mate strictly assortatively while the remainder mate at random. In this system, a correlation of one is equivalent to self-fertilization and as mentioned before, decreases heterozygosity at a rate of 50% per generation. Partial assortative mating ($r < 1$) is equivalent to partial self-fertilization. This was first worked out by Wright (1921). The population eventually approaches an equilibrium in which the homozygosity is equivalent to an inbreeding coefficient of $F = r/(2 - r)$. This will be demonstrated later.

A more important example for human genetics is the case where dominance is complete, which we now consider.

Assume that there are only two alleles and, since dominance is complete, there are only two phenotypes. Let P_t be the frequency of AA in generation t , $2Q_t$ be the frequency of heterozygous Aa , and R_t that of the recessive aa . Let r be the correlation between mating individuals; that is to say, a fraction, r , mate strictly assortatively and the rest mate at random with respect to the trait considered.

To see the algebraic relationships, we imagine the population as being divided into three groups: a randomly mating group comprising a fraction $(1 - r)$ of all matings and with the A gene frequency $P + Q = p$;

a recessive assortatively mating group making up a fraction rR of matings and with the A gene frequency zero; a dominant assortative group comprising a fraction $r(1 - R)$ and with the A gene frequency $(P + Q)/(1 - R)$ or $p/(1 - R)$ and recessive gene (a) frequency $Q/(1 - R)$. From the randomly mated group, the fraction of AA , Aa , and aa progeny will be p^2 , $2pq$, and q^2 , where $q = 1 - p$. The contribution of the dominant assortative group to the AA class next generation will be $r(1 - R)[p/(1 - R)]^2$, to the Aa class will be $r(1 - R)2[p/(1 - R)][Q/(1 - R)]$, and to the aa class will be $r(1 - R)[Q/(1 - R)]^2$. The recessive assortative mating group will make its entire contribution, rR , to the aa class.

Putting all this together, the genotype frequencies next generation will be

$$\begin{aligned}
 P(AA) &= P_{t+1} \\
 &= (1 - r)p^2 + r(1 - R_t) \\
 &\times \left(\frac{p}{1 - R_t}\right)^2 \quad (1) \\
 &= (1 - r)p^2 + \frac{r p^2}{1 - R_t} ,
 \end{aligned}$$

$$\begin{aligned}
 P(Aa) &= 2Q_{t+1} \\
 &= (1 - r)2pq + r(1 - R_t) \\
 &\times 2 \frac{p}{1 - R_t} \frac{Q_t}{1 - R_t} \quad (2) \\
 &= 2(1 - r)pq + \frac{2rpQ_t}{p + Q_t} ,
 \end{aligned}$$

$$\begin{aligned}
 P(aa) &= R_{t+1} \\
 &= (1 - r)q^2 + rR_t + r(1 - R_t) \\
 &\times \left(\frac{Q_t}{1 - R_t}\right)^2 = (1 - r)q^2 \quad (3) \\
 &+ r \left[\frac{q^2 + R_t(p - q)}{1 - R_t} \right] .
 \end{aligned}$$

We have written p and q with no subscripts, since they do not change with time. This can be verified by summing (1) and half of (2). Recalling that $p = P + Q$, $q = Q + R$, and $H = 2Q$, this simplifies to $p_{t+1} = p_t$, showing that the gene frequency does not change. As with inbreeding, only the genotype frequencies change, not the gene frequencies.

whose solution gives the equilibrium value, \hat{Q} , in terms of the correlation between mates and the gene frequency. This equilibrium and the rate of approach have been discussed by O'Donald (1960).

Of some interest is the extent to which assortative mating increases the frequency of homozygotes for recessive genes. The equilibrium proportion of recessive homozy-

TABLE 1
PROPORTION OF RECESSIVE HOMOZYGOTES WITH ASSORTATIVE MATING FOR VARIOUS VALUES OF THE RECESSIVE ALLELE FREQUENCY (q) AND THE DEGREE OF ASSORTATIVE MATING (r)*

DEGREE OF ASSORTATIVE MATING (r)	RECESSIVE ALLELE FREQUENCY (q)					
	0.01		0.1		0.5	
	R_1	R_∞	R_1	R_∞	R_1	R_∞
0.000.....	0.00010	0.00010	0.010	0.010	0.250	0.250
0.125.....	0.00011	0.00011	0.011	0.011	0.260	0.261
0.250.....	0.00012	0.00013	0.012	0.013	0.271	0.273
0.500.....	0.00014	0.00020	0.014	0.017	0.292	0.305
0.750.....	0.00017	0.00038	0.016	0.027	0.312	0.352
1.000.....	0.00020	0.01000	0.018	0.100	0.333	0.500

* The values given are the proportion of recessive homozygotes after one generation of assortative mating (R_1) and at equilibrium ($R_\infty = \hat{R}$).

When assortative mating is complete ($r = 1$), (2) becomes

$$2Q_{t+1} = \frac{2pQ_t}{p + Q_t} \tag{4}$$

gotes is given by equating R_{t+1} and R_t . This gives the quadratic

$$\hat{R}^2 - \hat{R}(1 + q^2 - rp^2) + q^2 = 0 \tag{6}$$

with the solution

$$\hat{R} = \frac{1 + q^2 - rp^2 - \sqrt{(1 + q^2 - rp^2)^2 - 4q^2}}{2} \tag{7}$$

This approaches 0 as t increases, but extremely slowly. When $p = \frac{1}{2}$, then $2Q_0 = \frac{1}{2}$, and the frequency of heterozygotes in successive generations follows the simple harmonic series, $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$, and so on, as first shown by Jennings (1916).

With any value of r except 1, the population never attains complete homozygosity, but approaches an equilibrium. We can find the equilibrium heterozygosity by equating Q_{t+1} to Q_t , giving

$$\hat{Q}^2 + p^2(1 - r)\hat{Q} - p^2q(1 - r) = 0, \tag{5}$$

Some numerical examples are given in Table 1.

Several general conclusions emerge from examination of this table. First, with weak assortative mating there is little ultimate increase in homozygous recessive genotypes, as seen in the values near the top of the table. However, the population goes a large fraction of the way to equilibrium in the first generation. On the other hand, as seen in the lower left part of the table, intensive assortative mating with a rare recessive gene

can lead eventually to a considerable increase in recessive homozygotes, but this is approached very slowly.

When q is small and r is not near to 1, the solution of (7) is approximately

$$\hat{R} = q^2 / (1 + q^2 - rp^2) \quad (8)$$

or, alternatively, if r is small a good approximation is

$$\hat{R} = q^2 [1 + rp^2 / (1 - q^2)], \quad (9)$$

where q^2 is the value of R when mating is at random. For example, if red hair were caused by a recessive gene and the tendency to assortative mating for this trait were 25%, the proportion of redheads would eventually be increased by roughly one-third, and most of this increase would be in the first generation.

Assortative mating is quite high for deafness, and it might be thought that this is a major factor in increasing the incidence. It has been estimated (Chung, Robison, and Morton, 1959) that there are at least 35 recessive genes, any one of which can cause deafness when homozygous, and with an average frequency of 0.002. Whatever the amount of assortative mating for deafness as a trait, it would be only about 1/35 of this amount for any one recessive gene—somewhat less because of other causes of deafness. Thus, even with strict assortative mating, the incidence would not be increased by more than 2% or 3%. The increase would be still less, if indeed it were an increase at all, because with assortative mating there would also be a tendency for association between different genes for deafness. To the extent that there is simultaneous homozygosity for two or more genes for deafness there will be fewer deaf persons than if the homozygosity for deafness genes were dispersed in separate individuals. However, the incidence might be enhanced if there were a tendency for consanguineous marriages among the deaf.

ASSORTATIVE MATING FOR A SIMPLE MULTIFACTORIAL TRAIT

Consider a trait determined by n gene loci. At each locus is a gene with frequency p such that the substitution of this gene for its allele adds a constant amount a to the character under consideration. Later, this re-

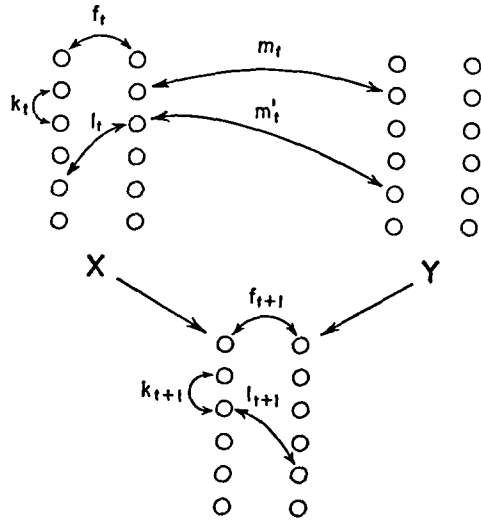


FIG. 1.—Correlations between the values of genes in two parents, X and Y, and their progeny. The circles represent individual genes. Homologous genes are opposite each other and genes from the same gamete are in a single vertical column.

striction to equal gene effects and equal frequencies at all loci will be removed.

Let

- n = the (haploid) number of relevant gene loci;
- f = correlation in value of homologous genes;
- k = correlation of nonhomologous genes in the same gamete;
- l = correlation of nonhomologues in different gametes;
- m = correlation of homologues in different individuals, X and Y;
- m' = correlation of nonalleles in different individuals.

These relations are shown in Fig. 1.

An individual gene has a variance pqa_2 , where $q = 1 - p$. This can be shown as follows: For convenience, let the value of one

allele be a and the other 0 , with frequencies p and q . The mean value is $pa + q0 = pa$. The variance, v , is $p(a - pa)^2 + q(0 - pa)^2 = pqa^2$. Likewise, the covariance, cov , of two genes, each with the same variance, is the variance times the correlation coefficient. For example, the covariance of two homologous genes is pqa^2f .

We can write the variance of the total value of individual X as the sum of the variances of the component genes. Thus

$$V(X) = \Sigma v_i + 2\Sigma cov_{ij}$$

where i and j designate individual genes.

The variance of an individual gene, v_i , is pqa^2 and there are $2n$ of them, so $\Sigma v_i = 2npqa^2$. The covariance of a pair of alleles is $pqfa^2$, and there are n pairs. The covariance between nonalleles from the same gamete is $pqka^2$ and there are $n(n - 1)$ combinations. Likewise, there are $n(n - 1)$ pairs of nonalleles in different gametes with covariance $pqla^2$. Putting all this together the variance of X at time t is

$$\begin{aligned} V(X)_t &= 2npqa^2 + 2npqfa^2 \\ &\quad + 2n(n - 1)pqka^2 \\ &\quad + 2n(n - 1)pqla^2 \quad (10) \\ &= 2npqa^2[1 + f_t + (n - 1) \\ &\quad \times (k_t + l_t)] . \end{aligned}$$

Likewise the covariance of X and Y is

$$\begin{aligned} Cov(X, Y)_t &= 4npqm_a a^2 \\ &\quad + 4n(n - 1)pqmq'_a a^2 . \quad (11) \end{aligned}$$

If the assortative mating is based solely on the phenotype, rather than being a by-product of common ancestry of the mates, and the gene frequencies are the same for all loci, there is no more reason for alleles in mates to be alike than nonalleles. Therefore $m_t = m'_t$ and we can drop the prime in equation (11), leading to

$$Cov(X, Y)_t = 4n^2pqm_a a^2 . \quad (12)$$

From Fig. 1 the following recurrence relations can be seen.

$$f_{t+1} = m_t \quad (13)$$

$$l_{t+1} = m'_t = m_t \quad (14)$$

$$k_{t+1} = (1 - c)k_t + cl_t \quad (15)$$

where c is the proportion of recombination between the two relevant loci. In this case it is an average of the recombination between all pairs of loci concerned with the character, and for man is very nearly one-half, since most pairs of loci are unlinked.

If r is the coefficient of correlation between the phenotypes of the two mates, X and Y , which have the same variance, the covariance is

$$C(X, Y) = rV(X) .$$

Substituting into this from (10) and (11) gives

$$\begin{aligned} 4n^2pqm_a a^2 &= r[1 + f_t + (n - 1) \\ &\quad \times (k_t + l_t)]2npqa^2 . \end{aligned}$$

Now we substitute f_{t+1} for m_t (see [13]) and f_t for l_t ([13] and [14]), which leads after some rearrangement to

$$f_{t+1} = \frac{r}{2n} [1 + nf_t + (n - 1)k_t] . \quad (16)$$

Using this and the relation

$$k_{t+1} = (1 - c)k_t + cf_t \quad (17)$$

(from equations [13], [14], and [15]) we can compute f_t for any generation t , given the starting values, f_0 and k_0 , which would both be 0 for a randomly mating population in gametic phase or "linkage" equilibrium.

At equilibrium there is no distinction between t and $t + 1$, so using carets to designate equilibrium values

$$\hat{f} = \hat{l} = \hat{m} = \hat{k} . \quad (18)$$

Using these equilibrium relations, (16) becomes

$$\hat{f} = \frac{r}{2n} [1 + n\hat{f} + (n - 1)\hat{l}],$$

leading to

$$\hat{f} = \frac{r}{2n(1 - r) + r} \quad (19)$$

as first shown by Wright (1921).

If n is large, \hat{f} is small unless r is very nearly 1. This shows that, unless the number of loci is small or the degree of assortative mating is very intense, there is only a very slight increase in homozygosity.

There is a much larger effect on the variance. From (10), substituting f_i for l_i from (13) and (14),

$$V(X)_t = V_0[1 + nf_t + (n - 1)k_t], \quad (20)$$

where $V_0 = 2npqa^2$, the variance with random mating and linkage equilibrium. At equilibrium under assortative mating, substituting into (20) from (19) and (18),

$$\hat{V}(X) = \frac{V_0}{1 - r(1 - 1/2n)} \quad (21)$$

(Wright, 1921) or, for large n ,

$$\hat{V}(X) \approx \frac{V_0}{1 - r}. \quad (21a)$$

As a numerical example, let $r = 1/4$, which is roughly the correlation in height between husbands and wives. The homozygosity is increased only trivially if n , the number of factors, is large. After one generation $f = 1/8n$ and at equilibrium is $1/(6n + 1)$ or approximately $1/6n$. On the other hand the variance is increased by $1/8$ in the first generation and eventually by $1/3$.

MULTIPLE ALLELES, UNEQUAL GENE EFFECTS, AND UNEQUAL GENE FREQUENCIES

Still assuming no dominance and epistasis and no environmental effects, the assump-

tions of only two alleles with equal effect and equal frequency will be dropped.

Let σ_i^2 be the variance of a gene at the i th locus. Thus $\sigma_i^2 = \sum p_k \alpha_k^2 - M^2$, where p_k and α_k are the frequency and effect on the trait of the k th allele, M is the mean effect of these alleles and the summation is over all alleles at the i th locus. The σ_i 's remain constant under assortative mating since the gene frequencies do not change.

The covariance σ_{ki} between two genes is $\sigma_k \sigma_i r_{ki}$ where r_{ki} is the correlation between the two genes. The correlations f, k, l, m and m' of Fig. 1 are no longer constant for all pairs of genes. Equations (10) and (12) can be written more generally as

$$V(X) = 2 \sum_{i=1}^n \sigma_i^2 + 2 \sum_i \sigma_i^2 f_i \quad (22)$$

$$+ 2 \sum_{i \neq j} k_{ij} \sigma_i \sigma_j + 2 \sum_{i \neq j} l_{ij} \sigma_i \sigma_j$$

and

$$C(X, Y) = 4 \sum_{i, j} m_{ij} \sigma_i \sigma_j. \quad (23)$$

The recurrence relations (13), (14), and (15) still apply to individual gene pairs; that is

$$f_{i, t+1} = m_{ii, t}, \quad (24)$$

$$l_{ij, t+1} = m_{ij, t}, \quad (25)$$

$$k_{ij, t+1} = (1 - c_{ij})k_{ij, t} + c_{ij}l_{ij, t}, \quad (26)$$

so that at equilibrium

$$f_i = m_{ii}, \quad k_{ij} = l_{ij} = m_{ij}. \quad (27)$$

Then, at equilibrium (22) becomes

$$\hat{V}(X) = 4 \sum_{ij} m_{ij} \sigma_i \sigma_j - 2 \sum_i m_{ii} \sigma_i^2 + 2 \sum_i \sigma_i^2. \quad (28)$$

We now let

$$= \sum_{ij} m_{ij} \sigma_i \sigma_j / \sum_i m_{ii} \sigma_i^2. \quad (29)$$

Substituting this into (23) and (28) gives

$$\hat{V}(X) = \hat{C}(X, Y) - \frac{\hat{C}(X, Y)}{2n_e} + V_0 \quad (30)$$

and

$$V_0 = V(X)_0 = 2 \sum_i \sigma_i^2. \quad (30a)$$

V_0 is $V(X)$ before the assortative mating began ($f = l = k = m = 0$ in [22]). Since $\hat{C}(X, Y) = r\hat{V}(X)$, we get

$$\hat{V}(X) = \frac{V_0}{1 - r(1 - 1/2n_e)}. \quad (31)$$

At equilibrium, the average inbreeding coefficient, weighted by the contribution of each locus to the variance, is

$$\hat{f} = \frac{\sum_i m_{ii} \sigma_i^2}{\sum_i \sigma_i^2}.$$

Substituting from (29), (23), and (30a)

$$\begin{aligned} \hat{f} &= \frac{\hat{C}(X, Y)}{2n_e} \cdot \frac{1}{V_0} \\ &= \frac{r}{2n_e} \cdot \frac{\hat{V}(X)}{V_0} \\ &= \frac{r}{2n_e(1 - r) + r}. \end{aligned} \quad (32)$$

Comparing (31) and (32) with (21) and (19) shows the equivalence of n and n_e . When $m_{ii} = m_{ij} = m$, then from (29)

$$n_e = \frac{\sum_{ij} \sigma_i \sigma_j}{\sum_i \sigma_i^2}$$

and if each locus has the same standard deviation ($\sigma_i = \sigma_j = \sigma$) then $n_e = n^2 \sigma^2 / n \sigma^2 = n$. We therefore call n_e the effective number of loci. It will be equal to the true number when there is free recombination and all loci contribute equally to the variance; otherwise it will be less.

Notice that when $n_e = 1$, (32) gives

$$\hat{f} = \frac{r}{2 - r}, \quad (33)$$

the value mentioned earlier when we discussed a single locus. Likewise

$$\hat{V}(X) = V_0 \left(\frac{2}{2 - r} \right). \quad (34)$$

The variance after one generation of assortative mating is readily derived. From equations (22), (23), (24), (25), and (30a) we can write

$$\begin{aligned} V(X)_1 &= V_0 + 2 \sum_i \sigma_i^2 m_{ii,0} \\ &\quad + 2 \sum_{i \neq j} k_{ij,1} \sigma_i \sigma_j \\ &\quad + 2 \sum_{i \neq j} \sigma_i \sigma_j m_{ij,0}. \end{aligned}$$

But $k_{ij,1} = 0$, from (26). Thus

$$\begin{aligned} V(X)_1 &= V_0 + 2 \sum_{i,j} m_{ij} \sigma_i \sigma_j \\ &= V_0 + \frac{1}{2} C(X, Y)_0 \quad (\text{from [23]}) \\ &= V_0 \left(1 + \frac{r}{2} \right) \end{aligned} \quad (35)$$

since $C(X, Y) = rV(X)$.

Table 2 gives some numerical illustrations of the increase in homozygosity and variance after one generation of assortative mating and after equilibrium is reached.

We have not considered the effects of disassortative mating, but there is nothing about these formulae that demands that r be positive. Disassortative mating has opposite effects, a decrease of homozygosity and variance and a building up of linkage disequilibrium in the opposite direction (i.e., an association in the same gamete of genes of opposite effect).

EFFECT OF DOMINANCE AND ENVIRONMENT

In a randomly mating population the variance can be divided into components.

$$V_t = V_g + V_d + V_e \quad (36)$$

or

$$V_t = V_h + V_e \quad (36a)$$

where V_t is the total variance and V_g , V_d , V_h , and V_e are the genic (additive genetic), dominance, genotypic (or total genetic), and environmental components.

The equations above assume that the genetic and environmental factors are independent so that V_e is simply additive to the other components. This is a major limitation to precise quantitative prediction of the

lier, with a large number of genes there is very little change in heterozygosity under assortative mating, and therefore V_d is not expected to change very much.

We let A be the correlation between the genic values of the mates. Thus,

$$A = r \frac{V_g}{V_t} = rH \tag{37}$$

where H is the *heritability*.

TABLE 2
EFFECT OF ASSORTATIVE MATING ON THE AVERAGE INBREEDING COEFFICIENT OF RELEVANT GENES AND THE VARIANCE OF THE TRAIT*

CORRELATION BETWEEN MATES r AND HERITABILITY (H)	GENE LOCI	INBREEDING COEFFICIENT		VARIANCE	
		f_1	f_∞	V_1	V_∞
				V_0	V_0
$r = 1$ } $H = 1$ }	1	0.500	1.000	1.500	2.00
	4	0.125	1.000	1.500	8.00
$r = .5$ } $H = 1$ }	1	0.250	0.333	1.250	1.33
	4	0.063	0.111	1.250	1.77
	∞	0.000	0.000	1.250	2.00
$r = .25$ } $H = 1$ }	1	0.125	0.143	1.125	1.14
	4	0.031	0.040	1.125	1.28
	∞	0.000	0.000	1.125	1.33
$r = .5$ } $H = .5$ }	∞	0.000	0.000	1.063†	1.21

* Subscripts 0 , 1 , and ∞ refer to the randomly mating population, the population after one generation of assortative mating, and the population at equilibrium under assortative mating.

† Exact only if $V_d = 0$.

phenotypic effects of assortative mating, particularly in human populations. We are also ignoring the effects of epistasis. Finally, all the results from here on are only approximate.

According to Fisher (1918), assortative mating will increase V_g , but not V_d and V_e . This is not surprising, since with multiple factors only genic effects contribute to the correlation between parent and offspring (Reeve, 1961). However, it is not strictly true, for V_d does change. But, as noted ear-

After one generation of assortative mating

$$V_t \approx V_g \left(1 + \frac{A}{2} \right) + V_d + V_e \tag{38}$$

from (35), after replacing r with A . Equation (38) will be reasonably accurate if V_d is small. Otherwise, the factor by which V_g is inflated may be appreciably in error. See Reeve (1961) for an exact expression for the two allele case. Equation (38) is strictly correct only in the absence of dominance.

At equilibrium under assortative mating

$$\hat{A} = r \frac{\hat{V}_o}{\hat{V}_i} = r\hat{H}. \quad (39)$$

Substituting into (36) from (31) gives

$$\begin{aligned} \hat{V}_i &= V_o \left(\frac{1}{1 - \hat{A}Q} \right) + V_d + V_e \\ &= V_i + V_o \left(\frac{\hat{A}Q}{1 - \hat{A}Q} \right) \\ &= V_i \left[1 + H \left(\frac{\hat{A}Q}{1 - \hat{A}Q} \right) \right] \end{aligned} \quad (40)$$

where $Q = 1 - 1/2n_e$.

We have used A instead of r because only the genic part of the correlation contributes significantly to the variance of future populations. Furthermore, we use the equilibrium value of A , since even with constant r there will be changes in A as the composition of the population changes.

The object is to express the population variance and the correlation between relatives after equilibrium under assortative mating in terms of quantities that can be measured in the random mating population before assortative mating began. To do this we must have a measure for \hat{A} . Note first the identity

$$V_i - V_o = V_o \left(\frac{1 - H}{H} \right), \quad (41)$$

which follows from the definition, $H = V_o/V_i$. But, since V_d and V_e do not change much with assortative mating,

$$\begin{aligned} V_i - V_o &= \hat{V}_o \left(\frac{1 - \hat{H}}{\hat{H}} \right) \\ &= V_o \left(\frac{1}{1 - \hat{A}Q} \right) \left(\frac{1 - \hat{H}}{\hat{H}} \right). \end{aligned} \quad (42)$$

Equating the right sides of (41) and (42) and recalling that $\hat{H} = \hat{A}/r$, we obtain, after some algebraic rearrangement, Fisher's equation for \hat{A} :

$$Q(1 - H)\hat{A}^2 - \hat{A} + Hr = 0. \quad (43)$$

H can be measured in the randomly mating population. Then, if Q is taken as 1 (i.e., the effective number of genes involved is assumed to be large, as it must be if other assumptions are to be correct), the equation can be solved for \hat{A} , and this value put into (40) to give the equilibrium variance.

As an example, let $H = .5$, $r = .5$, and $Q = 1$. Solving for \hat{A} gives $(2 - \sqrt{2})/2 = .293$. Then, from (40)

$$\begin{aligned} \hat{V}_i &= V_i \left[1 + .5 \left(\frac{.293}{1 - .293} \right) \right] \\ &= 1.207 V_i, \end{aligned}$$

so the population variance is increased after equilibrium under assortative mating of this degree by about 21%.

This value is given in the bottom row of Table 2, along with the increase in variance after one generation of assortative mating. As noted above, the latter value especially may be a poor approximation if V_d is large.

EFFECT OF ASSORTATIVE MATING ON THE CORRELATION BETWEEN RELATIVES

This was first done by Fisher (1918) and we follow his method.

Consider first parent-offspring correlation. The correlation is $V_o/2V_i$ in a randomly mating population. With equilibrium under assortative mating this will change for two reasons. One is that the variances increase, so we must replace V_o and V_i with their equilibrium values. The other reason is that the correlation between the two parents will, to the extent that this is reflected in genetic differences, add to the correlation of offspring with one parent through influences acting through the other.

If the chosen parent deviates by a unit amount from the population average, the other parent will deviate by r because of the correlation between the two mates. The mean deviation of the parents is thus $(1 + r)/2$, and the expected deviation of the children is the genic part of this, or \hat{V}_o/V_i

times the parental mean deviation. Thus, the correlation between the chosen parent and the offspring at equilibrium under assortative mating is

$$\hat{r}_{po} = \frac{1}{2} \frac{\hat{V}_o}{\hat{V}_t} (1 + r), \quad (44)$$

which in terms of the random mating variances is

$$\hat{r}_{po} = \frac{1}{2} \frac{V_o + V_o \hat{K}}{V_t + V_o \hat{K}} (1 + r), \quad (45)$$

$$\hat{K} = \frac{\hat{A}Q}{1 - \hat{A}Q},$$

as given by Fisher (1918) for $Q = 1$.

Fisher also gives the grandparent-child correlation as

$$\hat{r}_{po} = \frac{\hat{V}_o}{\hat{V}_t} \frac{1 + r}{2} \frac{1 + \hat{A}}{2} \quad (46)$$

and each additional descendant multiplies the correlation by $(1 + \hat{A})/2$, as expected since only the genic component is transmitted and therefore \hat{A} replaces r .

With full sibs the problem is more complicated because there are also correlations between the dominance components. Recall first the correlation between sibs under random mating,

$$r_{oo} = \frac{1}{2} \frac{V_o}{V_t} + \frac{1}{4} \frac{V_d}{V_t}. \quad (47)$$

The variance within a sibship with parents chosen at random is

$$\begin{aligned} V_s &= V_t(1 - r_{oo}) \\ &= V_t \left(1 - \frac{1}{2} \frac{V_o}{V_t} - \frac{1}{4} \frac{V_d}{V_t} \right) \\ &= \frac{1}{2} V_o + \frac{3}{4} V_d + V_e. \end{aligned} \quad (48)$$

However, Fisher notes that this is also a good approximation to the variance within a sibship when the parents are mated assortatively, since the variance within a sibship depends only on genes for which the parents

are heterozygous and, as we have learned, with a large number of genes the heterozygosity is only slightly decreased by assortative mating. Considering now the population at equilibrium under assortative mating,

TABLE 3

CORRELATIONS BETWEEN RELATIVES IN A RANDOMLY MATING POPULATION AND IN A POPULATION AT EQUILIBRIUM UNDER ASSORTATIVE MATING WHERE r IS THE PHENOTYPIC CORRELATION BETWEEN MATES, $H = V_o/V_t$, $D = V_d/V_t$, AND $A = Hr^*$

RELATIVES	CORRELATIONS	
	Random mating	Assortative mating
Parent-offspring.....	$\frac{1}{2}H$	$\frac{1}{2}\hat{H}(1+r)$
Grandparent-offspring....	$\frac{1}{4}H$	$\frac{1}{4}\hat{H}(1+r)(1+\hat{A})$
Great grandparent-offspring.	$\frac{1}{8}H$	$\frac{1}{8}\hat{H}(1+r)(1+\hat{A})^2$
Sibs.....	$\frac{1}{2}H + \frac{1}{4}D$	$\frac{1}{2}\hat{H}(1+\hat{A}) + \frac{1}{4}\hat{D}$
Double first cousins.....	$\frac{1}{4}H + \frac{1}{16}D$	$\frac{1}{4}\hat{H}(1+3\hat{A}) + \frac{1}{16}\hat{D}$
Uncle-niece....	$\frac{1}{4}H$	$\frac{1}{4}\hat{H}(1+\hat{A})^2 + \frac{1}{8}\hat{D}\hat{A}$
First cousins...	$\frac{1}{8}H$	$\frac{1}{8}\hat{H}(1+\hat{A})^2 + \frac{1}{16}\hat{D}\hat{A}^2$

* Equilibrium values under continued assortative mating are indicated by carets. The effective number of genes is assumed to be large, so that $(2n_e - 1)/2n_e$ may be regarded as 1.

the correlation is a measure of the reduction of the variance within a sibship.

Thus

$$1 - \hat{r}_{oo} = \frac{\frac{1}{2}V_o + \frac{3}{4}V_d + V_e}{\hat{V}_t}.$$

But, from (40), $\hat{V}_t = V_t + \hat{K}V_o$. Making this substitution and rearranging we obtain

$$\hat{r}_{oo} = \frac{V_o(\hat{K} + \frac{1}{2}) + \frac{1}{4}V_d}{V_t + KV_o} \quad (49)$$

where

$$\hat{K} = \frac{\hat{A}Q}{1 - \hat{A}Q}.$$

\hat{A} may be obtained from (43). Q may be taken as 1.

Correlations for other relatives are given in Table 3.

Fisher applied these methods to data on human stature. The data (obtained from earlier studies by Pearson and Lee) show

$$r = .2804$$

$$r_{po} = .5066$$

$$r_{oo} = .5433 .$$

From (44), we calculate the equilibrium heritability

$$\frac{\hat{V}_g}{\hat{V}_t} = \hat{H} = .791$$

from which

$$\hat{A} = \hat{H}r = .222$$

assuming $Q = 1$.

Assuming the observed correlations represent equilibrium values we can ask what the heritability was before assortative mating began:

$$V_g = \hat{V}_g(1 - \hat{A})$$

$$V_t = \hat{V}_t - V_g \left(\frac{\hat{A}}{1 - \hat{A}} \right)$$

$$H = \frac{V_g}{V_t} = .74.$$

So the assortative mating has increased the heritability from .74 to .79. From the sib correlation (49) we can estimate \hat{V}_d/\hat{V}_t , which turns out to be about the same as $1 - H$. Hence, on the basis of these data, Fisher concluded that environment is of very little importance in determining variance in human stature.

The analysis of variance in a population at equilibrium under assortative mating would be

V_g	62%
V_d	21%
V_t	83%
Effect of assortative mating	17%
\hat{V}_t	100%

Fisher assumed that the environmental similarity between sibs was no greater than that between parent and offspring. This seems quite dubious; it is probable that genes for height are less dominant than he thought and the environmental influence greater.

To make it easier to go from this treatment to Fisher's 1918 paper, here is a list of equivalents:

$$V_g = r^2 \quad \hat{V}_g/\hat{V}_h = c_2$$

$$V_h = \sigma^2 \quad \hat{A} = A$$

$$V_d = e^2 \quad Q = 1$$

$$\hat{V}_h/\hat{V}_t = c_1 \quad r = \mu$$

SUMMARY

Assortative mating, as does inbreeding, causes an increase in homozygosity and an increase in the population variance. However, with multiple factors the increase in homozygosity is very slight while the increase in variance is large. There is an association between genes of like effect and the resulting gametic phase (linkage) disequilibrium explains the large variance increase.

A trait determined by homozygosity for a rare recessive gene eventually has its incidence multiplied approximately by a factor $(1 - rp^2)^{-1}$, where $1 - p$ is the recessive gene frequency and r is the correlation between mates. Exact formulae are given for any generation.

A multifactorial trait with complete heritability (additive gene effects and no environmental influence) has at equilibrium an average inbreeding coefficient

$$f = \frac{r}{2n_e(1 - r) + r}$$

where n_e is the effective number of loci. The variance is increased by a factor $(1 - rQ)^{-1}$ where $Q = (2n_e - 1)/2n_e$. The methods used are similar to those of Wright (1921).

Extensions of these formulae are given to include dominance and environmental ef-

fects for a trait determined by a large number of loci. The effect of assortative mating on the correlation between certain relatives is also given. These were all shown earlier by Fisher (1918), but are derived here by a more elementary method.

ACKNOWLEDGMENTS

We are most grateful to Professor Sewall Wright and Mr. Thomas Wolfe for their useful comments on the manuscript. This paper is Number 1169 from the Genetics Laboratory, University of Wisconsin, Madison, Wisconsin, and was supported in part by the National Institutes of Health (GM 15422).

REFERENCES

- BREESE, E. L. 1956. The genetical consequences of assortative mating. *Heredity* 10:323-343.
- CHUNG, C. S., O. W. ROBISON, and N. E. MORTON. 1959. A note on deaf mutism. *Ann. Hum. Genet.* 23:357-366.
- FISHER, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* 52:399-433.
- JENNINGS, H. S. 1916. The numerical results of diverse systems of breeding. *Genetics* 1:53-89.
- MCBRIDE, G., and A. ROBERTSON. 1963. Selection using assortative mating in *D. melanogaster*. *Genet. Res.* 4:356-369.
- MORAN, P. A. P., and C. A. B. SMITH. 1966. Commentary on R. Fisher's paper on the correlation between relatives on the supposition of Mendelian inheritance. Cambridge University Press, London.
- O'DONALD, P. 1960. Assortative mating in a population in which two alleles are segregating. *Heredity* 15:389-396.
- REEVE, E. C. R. 1961. A note on nonrandom mating in progeny tests. *Genet. Res.* 2:195-203.
- WENTWORTH, E. N., and B. L. REMICK. 1916. Some breeding properties of the generalized Mendelian population. *Genetics* 1: 608-616.
- WRIGHT, S. 1921. Assortative mating based on somatic resemblance. *Genetics* 6:144-161.