3 INBREEDING

Inbreeding occurs when mates are more closely related than they would be if they had been chosen at random from the population. Related individuals have one or more ancestors in common, so the extent of inbreeding is related to the amount of ancestry that is shared by the parents of the inbred individuals. Alternatively stated, the degree of inbreeding of an individual is determined by the proportion of genes that his parents have in common.

An immediate consequence of this sharing of parental genes is that the inbred individual will frequently inherit the same gene from each parent. Thus inbreeding increases the amount of homozygosity. So one observable effect of inbreeding is that recessive genes, previously hidden by heterozygosity with dominant alleles, will be expressed. Since most such genes are harmful in one way or another, inbreeding usually leads to a decrease in size, fertility, vigor, yield, and fitness. There are also likely to be loci segregating in a population where a heterozygote is fitter than either of the two

corresponding homozygotes. In this case, too, inbreeding leads to a decreased fitness.

Another consequence of consanguineous mating within the population is greater genetic variability, since similar genes tend to be concentrated in the same individuals. Usually, because of the correlation between genotype and phenotype, this leads to an increase in phenotypic variability.

Inbreeding may follow either of two patterns. There may be a certain amount of consanguineous mating within a population, with the consequences just mentioned. On the other hand, the inbreeding may be such as to break the population into subgroups. An extreme example is continued self-fertilization in which the population (if it is of constant size and each parent contributes equally to the next generation) is divided into a set of subpopulations of one individual each. Likewise, a pattern of repeated sib mating could lead to a series of isolated populations of size 2. As a third example, there may be a natural population which is divided into isolated subpopulations, within each of which mating is random or nearly so. The effect will be that each subpopulation becomes more homozygous, and therefore the whole population does. The individual subpopulations become more uniform genetically; but, since they become homozygous for different genes, the population as a whole becomes more variable. Of course there may be only partial isolation, with intermediate consequences.

A point that at first seems paradoxical is that within a subpopulation there is an increase in homozygosity despite the fact that mating within this group is random. The reason, as will be discussed in Section 3.11, is that there are random changes in the frequencies of the individual alleles and these, on the average, lead to a decrease in heterozygosity. As an extreme example, self-fertilization can be regarded as random mating (i.e., random combination of gametes) within a population of one. The gene frequencies at different, previously heterozygous loci change from 1/2 to 0 or 1.

Whether inbreeding leads to subdivision or not, it can be measured in the same way—by Wright's (1922) coefficient of inbreeding, f, which measures the proportion by which the heterozygosity has been decreased. As we shall show later, other population properties can also be related to f. However, before discussing f, we shall illustrate with two simple examples the effect of continued inbreeding.

3.1 Decrease in Heterozygosity with Inbreeding

The qualitative effect of continued inbreeding can be seen by examining the most extreme form, self-fertilization. In a self-fertilized population the progeny of homozygotes are like their parents, whereas the progeny of heterozygotes are 1/2 heterozygotes and 1/4 each of the two homozygous types.

Thus, in each generation the proportion of heterozygous loci is reduced by half and the homozygous types are correspondingly increased. This is illustrated in Table 3.1.1.

Table 3.1.1. The changes in the probabilities of different genotypes with continued self-fertilization. D, H, and R stand for the initial proportions of dominant, heterozygous, and recessive types.

GENERATION	FREQUENCY OF GENOTYPE		
	AA	*Aa*	*aa*
0	D	H	R
1	$D + H/4$	$H/2$	$R + H/4$
2	$D + 3H/8$	$H/4$	$R + 3H/8$
3	$D + 7H/16$	$H/8$	$R + 7H/16$
4	$D + 15H/32$	$H/16$	$R + 15H/32$
Limit	$D + H/2$	0	$R + H/2$

If H_0 is the initial proportion of heterozygotes, the proportion after t generations of self-fertilization is $H_0/2^t$. If the original population were panmictic, with *AA*, *Aa*, and *aa* genotypes in the proportions p^2, $2pq$, and q^2 $(p + q = 1)$, the individual lines eventually become homozygous. The probability of being *AA* is $D + H/2 = p^2 + pq = p(p + q) = p$; likewise the probability of being *aa* is q. Thus, the population becomes broken into separate lines, each homozygous for one or the other of the genes in the ratio of their original frequencies in the population. Notice one other fact: There has been no change in the gene frequency. Inbreeding per se does not change the proportions of the various genes, only the way they are combined into homozygous and heterozygous genotypes.

With less extreme forms of inbreeding the results are similar, though the change in heterozygosity is less rapid. The results for continued brother-sister mating are shown in Table 3.1.2. Again there is a decrease in the proportion of heterozygotes, with the amount deducted being divided equally and added on to the two homozygous types.

These results may be obtained by writing out all the possible matings generation after generation, as was done by the early investigators (Fish, 1914; Jennings, 1916). This and several other systems of recurrent inbreeding were worked out by these authors. The papers are now mainly of historical interest since more general methods are available. We shall discuss them

Table 3.1.2. The decrease in heterozygosity with successive generations of brother-sister mating.

GENERATION	RELATIVE HETEROZYGOSITY	DECREASE IN HETEROZYGOSITY	RATES OF CHANGE IN HETEROZYGOSITY	
t	$\frac{H_t}{H_0} = P$	$\frac{H_0 - H_t}{H_0} = f$	$\frac{H_t}{H_{t-1}}$	$\frac{H_{t-1} - H_t}{H_{t-1}}$
0	1	0	1	0
1	2/2	0	1	0
2	3/4	1/4	3/4 = .750	.250
3	5/8	3/8	5/6 = .833	.167
4	8/16	8/16	8/10 = .800	.200
5	13/32	19/32	13/16 = .812	.188
6	21/64	43/64	21/26 = .808	.192
Limit	0	1	λ = .809	.191

in Sections 3.4 and 3.8, where the results of this table will appear as a special case.

In this example the heterozygosity follows a simple rule. The numerator in successive generations is given by the Fibonacci series in which each term is the sum of the two preceding terms, while the denominator doubles each generation. The number 1 in the second row is written as 2/2 to make the sequence more obvious. The reduction in heterozygosity, expressed as a fraction of the initial heterozygosity, is the same regardless of the initial gene frequencies and, as we shall show later, the number of alleles.

The relative heterozygosity, H_t/H_0, has been called by Wright (1951) the *panmictic index*, for which he used the letter P. $1 - P$ is the inbreeding coefficient, for which Wright has used the letter F. (We shall use the lower case f in order to reserve F for multiple-locus inbreeding effects.)

The last two columns give the rate of change in heterozygosity. Notice that the ratio H_t/H_{t-1} after a few oscillations rapidly approaches a constant value. The limiting value of the ratio of heterozygosity to that in the previous generation is usually designated by λ (Fisher, 1949).

3.2 Wright's Inbreeding Coefficient, *f*

Wright's (1922) original derivation of the inbreeding coefficient, f, was through correlation analysis. An alternative approach using only probability rules has been developed by Haldane and Moshinsky (1939), Cotterman (1940), and

Malécot (1948). They distinguish between two ways in which an individual can be homozygous for a given locus. The two homologous genes may be: (1) alike in state, that is to say, indistinguishable by any effect they produce (or perhaps, when molecular genetics has become sufficiently precise, alike in their nucleotide sequence), and (2) identical by descent, in that both are derived from the same gene in a common ancestor.

We follow the notation of Cotterman in designating an individual whose two homologous genes are identical by descent as *autozygous*. If the two alleles are of independent origin (as far as known from our pedigree information), the individual is *allozygous*. The effect of inbreeding is to increase that part of the homozygosity that is due to autozygosity. (Notice that an individual can be homozygous without being autozygous, if the two homologous genes are alike in state but not identical by descent. Conversely, an autozygous individual can be heterozygous for this locus if one of the two alleles has mutated since their common origin, although this is negligibly rare if only a small number of generations is being considered.)

The inbreeding coefficient, f, is defined as the probability that the individual is autozygous for the locus in question. Alternatively stated, it is the probability that a pair of alleles in the two gametes that unite to form the individual are identical by descent.

An individual with inbreeding coefficient f has a probability f that the two genes at a particular locus are identical and a probability $1 - f$ that they are not identical, and therefore independent. If they are independent the frequencies of the genotypes will be given by the binomial formula. If they are identical, the frequencies of the gene pairs will be simply the frequencies of the alleles in the population. Thus, for two alleles, A_1 and A_2, with frequencies p_1 and p_2 ($p_1 + p_2 = 1$), the genotype frequencies are:

	ALLOZYGOUS		AUTOZYGOUS	
Homozygous, A_1A_1:	$p_1^2(1-f)$	+	$p_1 f$	
Heterozygous, A_1A_2:	$2p_1p_2(1-f)$			3.2.1
Homozygous, A_2A_2:	$p_2^2(1-f)$	+	$p_2 f$	
Total	$1-f$		f	

Notice that when f is 0 these formulae reduce to the usual Hardy–Weinberg proportions. When $f = 1$ the population is completely homozygous. Thus f ranges from 0 in a randomly mating population to 1 with complete homozygosity. How to compute f from a pedigree will be shown later.

Multiple alleles introduce no difficulty. The genotype frequencies are a natural extension of the results for two alleles. The frequencies are

$$A_iA_i : p_i^2(1-f) + p_i f \qquad 3.2.2$$

for homozygous genotypes, and

$$A_iA_j : 2p_ip_j(1-f) \qquad 3.2.3$$

for heterozygous genotypes.

The expected proportion of heterozygous genotypes with inbreeding coefficient f, H_f, is given by

$$H_f = \sum_{i \neq j} p_i p_j(1-f) = H_0(1-f); \quad f = \frac{H_0 - H_f}{H_0}, \qquad 3.2.4$$

where H_0 is a constant equal to the proportion of heterozygotes expected with random mating ($f = 0$). The summation is over all combinations of values of i and j except when these are equal.

This proves the assertion made earlier that the inbreeding coefficient measures the fraction by which the heterozygosity has been reduced. We have written the formula as if, when $f = 0$, the population is in Hardy–Weinberg proportions. However, for any measured f (as determined, for example, from a pedigree), the heterozygosity, H, is $H_0(1-f)$, where H_0 is whatever the heterozygosity would have been in the absence of the observed inbreeding. To be concrete, the inbreeding coefficient for the child of a cousin marriage is 1/16 (as we shall show later); therefore the child of such a marriage is 15/16 as heterozygous as if his parents had the same relationship as a random pair in this population.

There is a simple relationship between the correlation coefficient, r, and the inbreeding coefficient, f. If we assign numerical values to each allele, then the inbreeding coefficient, f, is the correlation between these values in a pair of uniting gametes. In fact, Wright's original derivation of the inbreeding coefficient was through correlation methods.

The relationship between r and f can be shown in the following way. For convenience we assign the value 1 to allele A_1 and 0 to allele A_2, though we would get the same result with any values. The calculations are shown in Table 3.2.1.

Since the sum of the genotype frequencies is equal to 1, the weighted sum and the mean of any value are the same. For example, the sum (and mean) of the egg value, X, is $[p_2^2(1-f) + p_2 f](0) + [p_1p_2(1-f)](1) + [p_2p_1(1-f)](0) + [p_1^2(1-f) + p_1 f](1)$, which after some algebraic simplification reduces to p_1. The other calculations are given in the table, using the standard formula for calculation of r given in A.4.3.

The calculations in Table 3.2.1. are made by assuming that there are only two alleles and letting them have the values 0 and 1. The correlation interpretation of f, however, is completely general. Table 3.2.2 gives the same

Table 3.2.1. Demonstration of the equivalence of the inbreeding coefficient, f, and the coefficient of correlation, r_{xy}, between the genetic values of the uniting gametes.

EGG	SPERM	FREQUENCY OF THIS COMBINATION	VALUE OF EGG X	VALUE OF SPERM Y	X^2	Y^2	XY
A_2	A_2	$p_2^2(1-f)+p_2 f$	0	0	0	0	0
A_1	A_2	$p_1p_2(1-f)$	1	0	1	0	0
A_2	A_1	$p_2p_1(1-f)$	0	1	0	1	0
A_1	A_1	$p_1^2(1-f)+p_1 f$	1	1	1	1	1
Sum or Mean		1	p_1	p_1	p_1	p_1	$p_1^2(1-f)+p_1 f$

$$\bar{X} = p_1p_2(1-f) + p_1^2(1-f) + p_1 f = p_1.$$

Likewise, $\bar{Y} = \overline{X^2} = \overline{Y^2} = p_1$.

$$r_{xy} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sqrt{(\overline{X^2} - \bar{X}^2)(\overline{Y^2} - \bar{Y}^2)}} = \frac{p_1^2(1-f) + p_1 f - p_1^2}{p_1 - p_1^2} = f.$$

demonstration without restriction as to number of alleles and letting the contribution of the alleles differ. Furthermore, if the genic values are summed over k loci the covariance will be

$$f \sum_k \sum_i p_{ik} a_{ik}^2,$$

where p_{ik} and a_{ik} are the frequency and value of the ith allele at the kth locus. This is f times the variance. Hence f is the expected value of the correlation between the genetic values of two uniting gametes, regardless of the number of loci and number of alleles under consideration.

The equivalence of r and f suggests an interpretation of the correlation coefficient. If a measurement can be thought of as being the sum of a number of elements, then the correlation coefficient is the measure of the fraction of these elements that are common to the two measurements, the other elements being chosen at random. This interpretation is useful in many branches of science. In quantitative genetics the elements can obviously be interpreted as cumulatively acting genes.

The computation of f will be discussed in Section 3.4.

Table 3.2.2. Demonstration of the equivalence of the inbreeding coefficient and the correlation between the genetic value of the uniting gametes regardless of the contribution of the individual genes and the number of alleles. The contribution, or value, of allele A_i is assumed to be a_i, measured as a deviation from the mean value.

EGG	SPERM	FREQUENCY OF THIS COMBINATION	VALUE OF		X^2	Y^2	XY
			EGG X	SPERM Y			
A_i	A_i	$p_i^2(1-f)+p_i f$	a_i	a_i	a_i^2	a_i^2	a_i^2
A_i	A_j	$p_i p_j(1-f)$	a_i	a_j	a_i^2	a_j^2	$a_i a_j$

$$V_X = \sum_i p_i a_i^2,$$

since the variance of the egg value is the sum of the squares of the allele values, each weighted by its frequency. V_Y is the same.

$$\mathrm{Cov}_{XY} = (1-f)\Big[\sum_i p_i^2 a_i^2 + \sum_{i \neq j} p_i p_j a_i a_j\Big] + f \sum_i p_i a_i^2.$$

But the quantity in brackets is equal to $[\sum p_i a_i]^2$ which is equal to 0, because the sum of the deviations from the mean is 0. Therefore,

$$\mathrm{Cov}_{XY} = f \sum p_i a_i^2.$$

The correlation coefficient, being the ratio of the covariance to the geometric mean of the two variances (which in this case are the same), is f, as was to be shown.

$$r_{XY} = \mathrm{Cov}_{XY}/V_X = f.$$

3.3 Coefficients of Consanguinity and Relationship

We have used the inbreeding coefficient of an individual I, f_I, to give the probability that two homologous genes in that individual are identical by descent. Or, as just shown, this is the correlation between the genetic value of the two gametes that united to produce the individual. Since inbreeding of the progeny depends on the consanguinity of the parents we can use the inbreeding coefficient as a measure of this.

We define the *coefficient of consanguinity*, f_{IJ}, of two individuals I and J as the probability that two homologous genes drawn at random, one from each of the two individuals, will be identical. The answer to this is clearly the same as the inbreeding coefficient of a progeny produced by these two indi-

viduals. Hence the inbreeding coefficient of an individual is the same as the coefficient of consanguinity of its parents (Malécot, 1948).

There is a bewildering plethora of alternative names for this coefficient. Malécot, who introduced the idea, called it the *coefficient de parenté*. Falconer (1960) calls it the *coancestry*. Kempthorne (1957) translated *parenté* into *parentage*. Malécot himself has, on at least one occasion, translated it into *kinship*. We shall use either consanguinity or kinship.

A different measure of relatedness, introduced much earlier and still widely used, is Wright's (1922) *coefficient of relationship*, r_{IJ}, defined as:

$$r_{IJ} = \frac{2f_{IJ}}{\sqrt{(1 + f_I)(1 + f_J)}}. \qquad 3.3.1$$

For two individuals that are not inbred, the coefficient of relationship is exactly twice the coefficient of consanguinity.

As we shall show later, the coefficient of relationship is the correlation between the genic, or genetic, values of the two individuals. If the genes act without dominance or epistasis, and there is no effect of the environment on the trait being measured, this is the expected correlation. We shall also show later the effect of dominance on the correlation between relatives (Section 4.3).

3.4 Computation of *f* from Pedigrees

The procedure for computing the inbreeding or consanguinity coefficient from a pedigree follows directly from the definition of f. Consider the pedigree in Figure 3.4.1.

In this pedigree individual I is inbred because both his parents are descended from a single common ancestor, A. All unrelated ancestors, which are irrelevant to the inbreeding of I, are omitted from the pedigree. We ask for the probability that I is autozygous; i.e., that the homologous genes contributed to I by gametes b and e are both descended from the same gene in ancestor A. We shall use the notation Prob($c = b$) to mean the probability that c and b carry identical genes for the locus under consideration.

Prob($c = b$) = 1/2, since the gene in b has an equal chance of having come from C or from B's other parent. Likewise, Prob($c = a$) = 1/2. The probability that a and a' carry identical genes may be obtained as follows:

Let the two alleles in A be called W and Z. Then there are four equally likely possibilities for gametes a and a': (1) W and W, (2) Z and Z, (3) W and Z, and (4) Z and W. In the first two cases they are identical, so the probability is 1/2 that a and a' get the same gene from A. However, there is an additional possibility if ancestor A is inbred, for in this case the two alleles W and Z

may both be descended from some more remote ancestor not shown in the figure. The probability that A is autozygous, is, by definition, the inbreeding coefficient of A, f_A. Altogether, if A is inbred, $\text{Prob}(a = a') = \frac{1}{2} + \frac{1}{2}f_A = \frac{1}{2}(1 + f_A)$; if A is not inbred $\text{Prob}(a = a') = 1/2$.

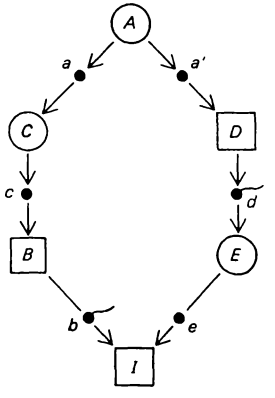


Figure 3.4.1. A simple pedigree with inbreeding. Circles and squares denote females and males respectively. Eggs and sperms are designated by small letters. Ancestors that do not contribute to the inbreeding of I are omitted.

Continuing around the path $BCADE$, $\text{Prob}(a' = d) = \text{Prob}(d = e) = 1/2$. Summarizing, b and e will carry identical genes only if b, c, a, a', d, and e do so. Therefore, since all these probabilities are independent

$$f_I = f_{BE} = \text{Prob}(b = e) = \underset{b=c}{\tfrac{1}{2}} \times \underset{c=a}{\tfrac{1}{2}} \times \underset{a=a'}{\tfrac{1}{2}(1 + f_A)} \times \underset{a'=d}{\tfrac{1}{2}} \times \underset{d=e}{\tfrac{1}{2}}$$

$$= (\tfrac{1}{2})^5(1 + f_A).$$

If A is not inbred (and according to information given in this pedigree she is not) the inbreeding coefficient of I is simply $(1/2)^5$. Notice that whether B, C, D, and E is inbred is irrelevant, since, for example, the probability that c and b are identical is independent of the gene contributed by B's other parent. The general rule is that the contribution of a path of relationship

through a common ancestor is $(1/2)^n(1 + f_A)$ where n is the number of individuals in the path from one parent to the ancestor and back through the other parent.

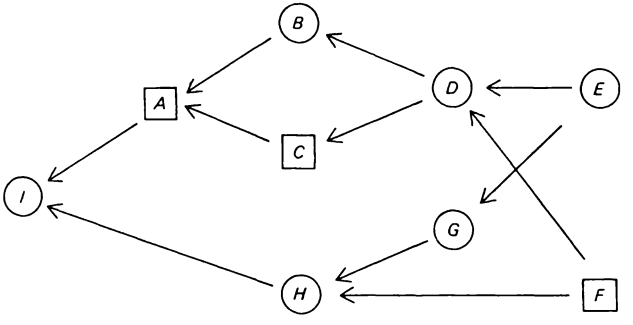


Figure 3.4.2. A more complicated pedigree; $f_I = 3/32$.

In more complicated pedigrees there may be multiple paths through an ancestor or more than one common ancestor. Consider the pedigree in Figure 3.4.2. The contributions to the inbreeding coefficient of I from the various paths are as follows. The common ancestor in a path is underlined.

PATH	CONTRIBUTION TO f
$ABD\underline{E}GH$	$(1/2)^6 = 1/64$
$ACD\underline{E}GH$	$(1/2)^6 = 1/64$
$ABD\underline{F}H$	$(1/2)^5 = 1/32$
$ACD\underline{F}H$	$(1/2)^5 = 1/32$

As we are considering only a single locus, the paths are all mutually exclusive; if I is autozygous for a pair of genes inherited through one path it cannot at the same time be autozygous for a pair inherited through another. Therefore the total probability for autozygosity is the sum of the probabilities for the separate paths, in this case 3/32.

This pedigree was not complicated by the common ancestor of any path being inbred. Individual A is inbred, but this is irrelevant since A is not a common ancestor. Only inbreeding of E or F would matter. This complication arises in the pedigree in Figure 3.4.3, where there are several inbred individuals in the pedigree and two of these, B and D, are common ancestors of one or more paths.

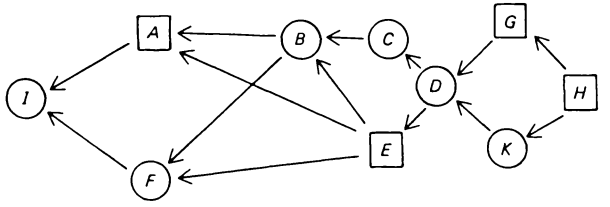


Figure 3.4.3. A still more complicated pedigree; $f_I = f_{AF} = .428$.

We begin by noting the inbreeding coefficients of D and B. The inbreeding coefficient of D, f_D, is $(1/2)^3$; $f_B = (1/2)^3(1 + f_D) = 9/64$. The components of f_I through the various paths are:

$A\underline{E}F$:	$(1/2)^3$	$= .1250$
$A\underline{B}F$:	$(1/2)^3(1 + f_B)$	$= .1426$
$AB\underline{E}F$:	$(1/2)^4$	$= .0625$
$A\underline{E}BF$:	$(1/2)^4$	$= .0625$
$ABC\underline{D}EF$:	$(1/2)^6(1 + f_D)$	$= .0176$
$AE\underline{D}CBF$:	$(1/2)^6(1 + f_D)$	$= .0176$
	$f_I = f_{AF}$	$= .4278$

Notice that the inbreeding of B is taken into consideration in path ABF where B is a common ancestor, but ignored in the other paths where B is not the common ancestor. A path such as $ABCDEBF$ is not included because B enters twice; the contribution to this path is included in path ABF by the term $(1 + f_B)$.

To summarize: The inbreeding coefficient of an individual I, or the coefficient of consanguinity of his parents, J and K, is the sum of a series of terms, one for each path leading from a parent to a common ancestor and back through the other parent. The general formula is

$$f_I = f_{JK} = \Sigma[(1/2)^n(1 + f_A)], \qquad \textbf{3.4.1}$$

where the summation is over all possible paths, n is the number of individuals in the path (counting J and K, but not I) and f_A is the inbreeding coefficient of the common ancestor at the apex of this path.

A path cannot pass through the same individual twice. No reversal of direction is permitted except at the common ancestor; always go against the arrows in going from one parent to the ancestor, and with them coming back

through the other. It is helpful in avoiding counting the same path twice to adopt the convention of starting all paths with the same parent (the male, say) and ending with the other.

In the earlier literature the procedure given for computing f was to count the number of steps between individuals in a path rather than the number of individuals. The results are of course the same either way. Formula 3.4.1 was first given in the present form by Wright (1951). We use it because it follows more naturally from our derivation than the earlier form, and because it is easily adapted to X-linked genes.

An X-chromosome gene that is in a gamete produced by a male must be the same as was in the egg from which this male came. Therefore the probability of identity by descent in these two gametes is 1, rather than 1/2 as it would be for a female or an autosomal locus. Hence each male in a path multiplies the probability of identity through this path by 1 rather than 1/2, and the effect is as if the males were not counted at all. Furthermore, a male does not receive an X-chromosome from his father, so a path involving two successive males makes no contribution to the probability of identity of X-chromosomal loci.

Therefore the rule for obtaining the inbreeding coefficient for a sex-linked locus in females is: Proceed as usual except that only females in a path are counted and any path with two successive males is omitted entirely.

As examples, consider again the three pedigrees in Figure 3.4.1, 3.4.2, and 3.4.3. There is no meaning to the inbreeding coefficient of a male since he has only one X chromosome; so the pedigree in 3.4.1 is not of interest. In Figure 3.4.2, f_I is 1/32 (path $ABD\underline{E}GH$) + 1/8($ABD\underline{F}H$) = 5/32. Paths $ACD\underline{F}H$ and $ACD\underline{E}GH$ have successive males and are omitted. In the same manner, in Figure 3.4.3, $f_D = 0$, $f_B = 1/4$, and $f_I = 5/8$.

3.5 Phenotypic Effects of Consanguineous Matings

In Section 2.2 the frequency of the recessive gene causing phenylketonuric feeble-mindedness was given as approximately 1/100. Therefore with random mating the frequency of persons homozygous for the gene is the square of this, or 1/10,000. We now inquire how much this is enhanced with consanguineous marriage.

The probability of an affected child as given by 3.2.1 is $p^2(1-f)+pf$, where p is the frequency of the recessive allele and f is the inbreeding coefficient of the child. If the parents are cousins their coefficient of consanguinity or the inbreeding coefficient of their child is 1/16, as computed from Figure 3.5.1. With $p = 1/100$ and $f = 1/16$ the expected frequency of homozygous recessives is 115/160,000, or approximately 7/10,000, a 7-fold increase compared with the risk when the parents are unrelated.

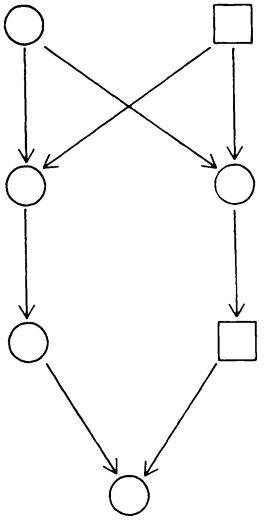


Figure 3.5.1. Pedigree of an individual whose parents are cousins. $f = \frac{1}{2}^5 + \frac{1}{2}^5 = 1/16$.

Table 3.5.1 illustrates the way in which the relative frequency of a disease in the children of cousin marriages increases with the rarity of the gene.

We can also ask the question the other way around: What proportion of the persons affected with recessive traits come from consanguineous marriages?

This proportion, K, may be obtained by dividing the number of affected from consanguineous marriages by the total number of affected. If c is the

Table 3.5.1. The proportion of homozygous-recessive individuals in a population with random mating and when the parents are first cousins for various frequencies of the recessive allele.

GENE FREQUENCY	FREQUENCY OF AFFECTED WHEN		RATIO
	$f = 0$	$f = 1/16$	
0.1	0.01	0.016	1.6
0.01	0.0001	0.00072	7.2
0.005	0.000025	0.000335	13.4
0.001	0.000001	0.000063	63

proportion of consanguineous marriages in the population of size N, the number of affected from consanguineous marriages is $Nc[p^2(1-f)+pf]$, where f is the coefficient of consanguinity of the parents and p is the recessive gene frequency. The total number of affected in the population is $N[p^2(1-\bar{f})+p\bar{f}]$, where $\bar{f}$ is the mean inbreeding coefficient in the population. Therefore

$$K = \frac{c[p^2(1-f)+pf]}{p^2(1-\bar{f})+p\bar{f}} = \frac{c[p+(1-p)f]}{p-p\bar{f}+\bar{f}}, \qquad 3.5.1$$

or, since $p\bar{f}$ is usually very small

$$K = \frac{c[p+(1-p)f]}{p+\bar{f}} \qquad 3.5.2$$

approximately.

The most common consanguineous marriage is between first cousins. For $f = 1/16$, 3.5.1 becomes

$$K = \frac{c(1+15p)}{16[p+\bar{f}(1-p)]},$$

or approximately

$$\frac{c(1+15p)}{16(p+\bar{f})}. \qquad 3.5.3$$

Some representative values of K are given in Table 3.5.2.

This shows that, even though consanguineous marriages are very rare, a substantial fraction of diseases caused by recessive genes comes from such marriages if p is .01 or less. Consanguinity of the parents is one of the strongest kinds of evidence of recessive inheritance.

On the other hand, if the recessive gene is common, the increased incidence with consanguinity is very slight. Cystic fibrosis of the pancreas appears to be due to a simple recessive factor, yet there is no appreciable rise in incidence from consanguineous marriages, because the allele frequency is so high. The parental-consanguinity rate is much higher for recessive traits where the gene frequency is low.

Most recessive genes are carried concealed in the heterozygous condition. We can get some idea of the total number of such genes carried by normal persons through a study of consanguineous marriages. A study by Jan A. Book showed that about 16% of the children of first cousin marriages in Sweden had a genetic disease, and if diseases of more doubtful etiology were included the number rose to 28%. The corresponding figures for the control population with unrelated parents were 4% and 6%. Thus cousin marriage, by these

Table 3.5.2. The proportion, K, of cases of recessive conditions expected from first-cousin marriage for various values of gene frequency (p), frequency of cousin marriage (c), and population inbreeding coefficient (f).

c	f	p	K
0.005	0.0005	0.1	0.008
0.005	0.0005	0.03	0.015
0.005	0.0005	0.01	0.034
0.005	0.0005	0.003	0.093
0.005	0.0005	0.001	0.212
0.005	0.0005	0.0003	0.392
0.01	0.001	0.1	0.015
0.01	0.001	0.03	0.029
0.01	0.001	0.01	0.065
0.01	0.001	0.003	0.163
0.01	0.001	0.001	0.317
0.01	0.001	0.0003	0.483
0.02	0.002	0.1	0.031
0.02	0.002	0.03	0.057
0.02	0.002	0.01	0.120
0.02	0.002	0.003	0.262
0.02	0.002	0.001	0.423
0.02	0.002	0.0001	0.546

data, entails an increased risk of 12% to 22% of having a child with a detectable genetic defect. Since the child of a cousin marriage has an inbreeding coefficient of 1/16, we reason that a completely homozygous individual would have 16 times as many diseases, or approximately 2.0 to 3.5. This is the number of recessive factors per gamete (since a homozygous individual may be regarded as a doubled gamete), so the number per zygote is between 4 and 7. These figures are based on rather limited data, but they furnish a rough idea of the magnitude. The conclusion is that the average human carries hidden the equivalent of some half a dozen deleterious recessive genes that, if made homozygous, would cause a detectable disease.

We can also estimate the amount of genetic weakness that is carried hidden in a heterozygous individual, but which would be expressed as inviability if he were made homozygous. Sutter and Tabah (1958 and earlier) found from a demographic study in two rural provinces in France that children of cousin marriages died before adulthood about 25% of the time,

whereas the death rate from unrelated parents was about 12%. Thus, in this environment, cousin marriage increased the risk of death by about 0.13. Making the same calculations as above (i.e., multiplying by 16 × 2) we estimate that the average individual in this population carries 32 × .13 or about 4 hidden "lethal equivalents." We say "lethal equivalents" because one cannot distinguish between 4 full lethals and 8 genes with 50% probability of causing death, or any system where the product of the number of genes and the average effect of each is 4. For a more sophisticated treatment of this subject, making use of all degrees of relationship rather than just cousins, see Morton, Crow, and Muller (1956).

The data on human inbreeding effects have not been very reproducible. The large body of data from Japan show significant heterogeneity effects from city to city. There is danger of confounding inbreeding effects with the effects of social concomitants of consanguineous marriages. For all these reasons, we cannot place too much reliance on the numerical values of the previous paragraph.

It is also to be expected that what is lethal in one environment may be only detrimental in a better one. In much of the world there has been a substantial rise in the standard of living and a decrease in the death rate. This means that the number of lethal equivalents is decreasing.

In Drosophila, where the measures are precise and reproducible, there are about two lethal equivalents per fly. About 2/3 of the viability depression from inbreeding is attributable to monogenic lethals; the rest is the cumulative effect of a much larger number of genes with individually small effects.

3.6 The Effect of Inbreeding on Quantitative Characters

We consider first a theoretical model that is applicable to any measurable trait, such as height, weight, yield, survival, or fertility. For initial simplicity, a single locus with only two alleles is assumed. The model is summarized as follows:

GENOTYPE	A_1A_1	A_1A_2	A_2A_2
FREQUENCY	$p_1^2(1-f)+p_1 f$	$2p_1p_2(1-f)$	$p_2^2(1-f)+p_2 f$
PHENOTYPE	$Y-A$	$Y+D$	$Y+A$

In this model Y is the residual phenotype when the A locus is not considered. Genotype A_2A_2 adds an amount A to the phenotype, and A_1A_1 subtracts an equal amount. (We could just as well assume that both genotypes add to the residual, or that both subtract; this model is chosen arbitrarily and for algebraic simplicity. The same result would be obtained in any case.)

If there were no dominance, the phenotype A_1A_2 would be Y. Under this circumstance, the amount by which the phenotype is changed by substituting an A_2 for an A_1 is always A, which we can call the additive effect of the A locus. D is a measure of dominance. When $D = 0$, there is no dominance; when $D = A$, A_2 is completely dominant; when $D = -A$, A_2 is completely recessive, or A_1 dominant; when $D > A$, there is overdominance, the heterozygote having a higher phenotypic value than either homozygote.

We now obtain an expression for the mean phenotype $\bar{Y}$. This will be given by summing the products of each phenotype and its frequency. (Since the frequencies add up to one, the sum is the same as the mean.) Therefore,

$$\bar{Y} = (Y - A)[p_1^2(1 - f) + p_1 f] + (Y + D)2p_1p_2(1 - f) + (Y + A)[p_2^2(1 - f) + p_2 f],$$

which leads after some algebraic rearrangement to

$$\begin{aligned} \bar{Y} &= Y + A(p_2 - p_1) + 2p_1p_2 D - 2p_1p_2 Df \\ &= G - Hf, \end{aligned} \qquad \text{3.6.1}$$

where $G = Y + A(p_2 - p_1) + 2p_1p_2 D$ and $H = 2p_1p_2 D$. G is the average phenotype with random mating and $G - H$ is the average with complete homozygosity. Notice that H is positive if D is positive.

This equation brings out two important facts about the phenotypic consequences of inbreeding. The first is that in the absence of dominance ($D = 0$), there is no mean change with inbreeding. The second fact is that the equation is a linear function of f. This means that whatever the level of dominance, as measured by D, the change with inbreeding is proportional to the inbreeding coefficient. As long as D is positive (i.e., the heterozygote has a larger phenotype than the mean of the two homozygotes), inbreeding will produce a decline.

There are two possible causes of the inbreeding decline that is so universally observed: (1) Favorable genes tend to be dominant or partially dominant ($0 < D < A$), and (2) the heterozygote has a higher phenotype than either homozygote, ($0 < A < D$). Notice that the observation of a linear decline in a quantitative trait cannot discriminate between these possibilities, for it would be expected with either type of gene action, or any mixture of the two. To discriminate between them will require other kinds of evidence.

The extension to more than two alleles is straightforward and will not be given here. We shall consider the extension to more than one locus. Consider a model with two loci. each with two alleles. We shall let A stand

for the additive effect of the A locus and B for that of the B locus. If there is no interaction between loci A and B, the model is as follows:

GENOTYPE		A_1A_1	A_1A_2	A_2A_2
	FREQUENCY	$p_1^2(1-f)+p_1 f$	$2p_1p_2(1-f)$	$p_2^2(1-f)+p_2 f$
B_1B_1	$r_1^2(1-f)+r_1 f$	$Y-A-B$	$Y+D_A-B$	$Y+A-B$
B_1B_2	$2r_1r_2(1-f)$	$Y-A+D_B$	$Y+D_A+D_B$	$Y+A+D_B$
B_2B_2	$r_2^2(1-f)+r_2 f$	$Y-A+B$	$Y+D_A+B$	$Y+A+B$

If the A and B loci are independent and in gametic phase equilibrium, the frequency of any of the nine classes in the table is given by the product of the frequencies at the borders. We get the mean phenotype as before, by multiplying each phenotype value within the table by its frequency and summing the nine products. After simplification this leads to

$$\bar{Y} = Y + A(p_2 - p_1) + 2p_1p_2\,D_A + B(r_2 - r_1) + 2r_1r_2\,D_B - 2(p_1p_2\,D_A + r_1r_2\,D_B)f. \qquad 3.6.2$$

As before, there is a linear relation to the inbreeding coefficient (unless $D_A = D_B = 0$), as might be expected from knowledge that this is true for either locus by itself. We now complicate the model by assuming an interaction between the two loci. In population genetics the word *epistasis* is given a meaning broader than its classical meaning so as to include all levels of nonadditive effects between loci. Any circumstance where a substitution at the A locus has a different effect depending on the genotype at the B locus is an example of epistasis. A simple way to construct such a model is to add interaction terms to each of the values already given so that the phenotypes are now:

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	$Y-A-B+I$	$Y+D_A-B-L$	$Y+A-B-I$
B_1B_2	$Y-A+D_B-K$	$Y+D_A+D_B+J$	$Y+A+D_B+K$
B_2B_2	$Y-A+B-I$	$Y+D_A+B+L$	$Y+A+B+I$

Notice that there are nine parameters, Y, A, B, D_A, D_B, I, J, K, and L, which correspond to the nine phenotypes so there is a complete specification of the phenotypes when the parameters are given, and vice versa.

The formula for $\overline{Y}$ may be written as

$$\overline{Y} = G - Hf + Mf^2, \qquad 3.6.3$$

where

$$\begin{aligned} G = {} & Y + A(p_2 - p_1) + B(r_2 - r_1) + 2D_A p_1 p_2 + 2D_B r_1 r_2 \\ & + I(r_1 - r_2)(p_1 - p_2) + 2Kr_1 r_2(p_2 - p_1) + 2Lp_1 p_2(r_2 - r_1) \\ & + 4Jp_1 p_2 r_1 r_2, \\ M = {} & 4p_1 p_2 r_1 r_2 J, \end{aligned}$$

and

$$\begin{aligned} H = {} & 2[p_1 p_2 D_A + r_1 r_2 D_B + p_1 p_2(r_2 - r_1)L \\ & + r_1 r_2(p_2 - p_1)K + 4p_1 p_2 r_1 r_2 J]. \end{aligned}$$

In this model, A and B represent the additive effects of the two loci, and D_A and D_B the dominance effects, as before. I is the effect of pure epistasis without dominance; in other words, the interaction of the additive effect of A with the additive effect of B. K is a measure of interaction and dominance; it is the effect of the A locus on the dominance of the B locus. Likewise, L is the effect of the B locus on the dominance of the A locus. Another way of saying this is that this is the interaction of the additive effect of the B locus with the dominance effect of the A. Finally, J is the epistatic effect of the two dominances; it is the interaction of dominance at the A locus with dominance at the B locus.

From equation 3.6.3 we see that if all the terms involving dominance effects—D_A, D_B, J, K, and L—are 0, there is no inbreeding effect since the coefficients of f and f^2 are 0. Thus epistasis alone, without dominance, does not produce an inbreeding decline.

If $J = 0$ the inbreeding change is linear in f. In order for the coefficient of f^2 to be other than 0, there must be interaction between the two dominance effects.

If A, B, D_A, D_B, K, and L are all positive, then the genes with subscript 2 are associated with increased performance (or yield, or fitness, or whatever is being measured). This also generally means that the alleles with subscript 2 will be more frequent than those with subscript 1. Then if J is positive there will be *diminishing* epistasis during inbreeding. By this, we mean that the curve is concave upward and that homozygosity for two loci reduces performance by less than the sum of the individual effects. Contrariwise if J is negative (assuming $H > 0$), the epistasis is *reinforcing*. That is, the deleterious

effect of two loci is more than cumulative. This is sometimes called synergistic.

These general types of epistasis are illustrated in Figure 3.6.1.

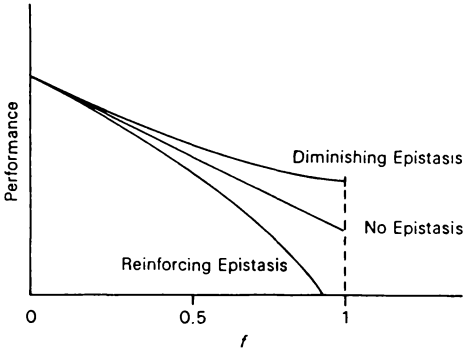


Figure 3.6.1. Decline in performance (or any trait of interest) with inbreeding under no epistasis, diminishing epistasis, and reinforcing epistasis.

3.7 Some Examples of Inbreeding Effects

Table 3.7.1 gives some numerical examples and illustrates the way in which specific levels of dominance and epistasis may be constructed by choosing appropriate values for the parameters. In these examples the gene frequencies are all 1/2. The values are contrived so that fully homozygous individuals ($f = 1$) have an average phenotype of 10.

Models 1 and 2 show that, whether or not there is epistasis, there is no inbreeding effect without dominance; the phenotype is 10, independent of f.

Models 3 and 4 show that the rate of decline with inbreeding by itself cannot distinguish among different levels of dominance. Model 5 is an extreme form of reinforcing epistasis; model 6 is an extreme form of diminishing epistasis. Models 6 and 9 show that quite different systems of gene action can give the same quadratic equation.

Models 7 and 8 show that epistasis, though it is a necessary condition for a nonlinear inbreeding effect, is not sufficient. In both there is epistasis, yet the equations are the same; moreover they are the same as models 3 and 4. Only when J is not 0 is there a nonlinear effect, as 5, 6, and 9 show.

The examples in Table 3.7.1 all assume equal frequency for the two alleles at each locus. Usually this is not the case. Let us assume that the higher phenotypic value is desirable, which means that for most of the models

Table 3.7.1. Some numerical illustrations of the phenotypic values under various models of dominance and epistasis. Also given is the average phenotype, $\bar{Y}$, when $p_1 = p_2 = r_1 = r_2 = .5$.

MODEL	DESCRIPTION	PHENOTYPIC VALUE	AVERAGE PHENOTYPE
1.	No dominance, no epistasis $Y = 10, A = B = 1$ $D_A = D_B = I = J = K = L = 0$	8 9 10 9 10 11 10 11 12	$\bar{Y} = 10$
2.	Epistasis, but no dominance $Y = 10, A = B = 2, I = 1$ $D_A = D_B = J = K = L = 0$	7 8 9 8 10 12 9 12 15	$\bar{Y} = 10$
3.	Complete dominance, no epistasis $Y = 10, A = B = D_A = D_B = 1$ $I = J = K = L = 0$	8 10 10 10 12 12 10 12 12	$\bar{Y} = 11 - f$
4.	Overdominance, no epistasis $Y = 10, A = B = 0, D_A = D_B = 1$ $I = J = K = L = 0$	10 11 10 11 12 11 10 11 10	$\bar{Y} = 11 - f$
5.	Complete dominance, complementary recessive genes $Y = 10, A = B = D_A = D_B = 1$ $I = J = K = L = -1$	7 11 11 11 11 11 11 11 11	$\bar{Y} = \frac{43}{4} - \frac{1}{2}f - \frac{1}{4}f^2$
6.	Complete dominance, complementary dominant genes $Y = 10, A = B = D_A = D_B = 1$ $I = J = K = L = 1$	9 9 9 9 13 13 9 13 13	$\bar{Y} = \frac{45}{4} - \frac{3}{2}f + \frac{1}{4}f^2$
7.	Dominance and epistasis $Y = 10, A = B = D_A = D_B = 1$ $I = K = L = 1, J = 0$	9 9 9 9 12 13 9 13 13	$\bar{Y} = 11 - f$
8.	Dominance and epistasis $Y = 10, A = B = D_A = D_B = I = 1$ $J = K = L = 0$	9 10 9 10 12 12 9 12 13	$\bar{Y} = 11 - f$
9.	Overdominance and epistasis $Y = 10, A = B = 0, D_A = D_B = 1,$ $I = K = L = 0, J = 1$	10 11 10 11 13 11 10 11 10	$\bar{Y} = \frac{45}{4} - \frac{3}{2}f + \frac{1}{4}f^2$

A_2 and B_2 will tend to be more frequent than their alleles. Consider as an example that $p_1 = r_1 = .1$. Then the equations for models 5 and 6 become:

Model 5, complementary recessive genes: $\overline{Y} = G - .007f - .032f^2$,
Model 6, complementary dominant genes: $\overline{Y} = G - .356f + .032f^2$.

Model 5 shows that with a trait depending on simultaneous homozygosity for two rare recessives, there will be very little inbreeding effect when f is small—because the coefficient of f is small, and f^2 is very small. Only when the inbreeding is sufficient for f^2 to be appreciable will its larger coefficient become important. Thus, to the extent that there are detrimental traits depending on multiple homozygosity, inbreeding effects will tend to be nonlinear, with very little effect of slight inbreeding, but with an accelerating effect at very high levels of inbreeding.

With model 6, on the other hand, the quadratic term never becomes important and the linear term dominates for all values of f. Thus for rare genes with duplicate effects the inbreeding effect is linear for all practical purposes.

When the trait considered is survival it is often more natural to measure epistasis as deviations from independence rather than from additivity. Survival probabilities are multiplicative if the genes act independently. It is often advantageous to transform to logs, or to measure fitness in Malthusian parameters.

We summarize this section by stating three conclusions, all of which are apparent from 3.6.3 and are illustrated by the numerical examples in Table 3.7.1.

1. If there is no dominance, there is no mean change in phenotype with inbreeding regardless of the amount of epistasis (models 1 and 2.). If D_A, D_B, J, K, and L are 0, $\overline{Y}$ is not a function of f.

2. If there is dominance, but no epistasis, the effect of inbreeding on the phenotype is linear in f. Usually inbreeding leads to a change in quantitative measures and if there is no epistasis this change is proportional to f (models 3 and 4.) If $I = J = K = L = 0$, the term in f^2 drops out.

3. If there is both dominance and epistasis the inbreeding effect may be quadratic in f (or higher order if more than two loci are involved). With reinforcing type epistasis the inbreeding effect is greater than if the loci were additive; with diminishing epistasis the effect is less (models 5 and 6). However, the change in average phenotype with inbreeding is not necessarily quadratic (models 7 and 8); as long as $J = 0$, the inbreeding effect is linear.

Reliable data on the results of inbreeding uncomplicated by the effects of selection are rare. There is also a difficulty in choosing an appropriate scale of measurement if the linearity of the inbreeding effect is to be tested. Some of the best data come from maize and Table 3.7.2 shows an example.

The yield was measured for inbred lines, crosses, and randomly mated progeny from fields of crosses.

Table 3.7.2. The average yield in bushels per acre of randomly pollinated maize derived from hybrids between inbred lines. Expected yields are based on assumption of no epistasis. Data from Neal (1935).

	HYBRID AVERAGE $f=0$ (G)	INBRED AVERAGE $f=1$ $(G-H)$	RANDOMLY POLLINATED		
			f	EXPECTED YIELD $(G-Hf)$	OBSERVED YIELD
10 two-way hybrids	62.8	23.7	.500	43.3	44.2
4 three-way hybrids	64.2	23.8	.375	49.1	49.3
10 four-way hybrids	64.1	25.0	.250	54.3	54.0

Three kinds of crosses were tested, two-way, three-way, and four way (or double-cross) hybrids. Let A, B, C, and D stand for four lines that have been self-fertilized long enough to be regarded as completely homozygous. Two-way crosses are first generation hybrids between two lines, e.g., $A \times B$; three-way crosses are between a hybrid and a different inbred, e.g., $(A \times B) \times C$; four-way crosses are between two different hybrids, e.g., $(A \times B) \times (C \times D)$.

If a field of two-way hybrids is allowed to pollinate at random the probability that two alleles have come from the same parental inbred line is 1/2. Assuming the parent line to be autozygous, two alleles from the same line are identical; hence the progeny from random pollination have an inbreeding coefficient of 1/2. For a four-way cross the probability of two alleles from the same line is 1/4. For the three-way cross, $(A \times B) \times C$, the probability of two alleles both coming from C is 1/4, from A is 1/16, and from B is 1/16; f is the sum, or 3/8.

The data in Table 3.7.2. show the close agreement with the expected values. Since the inbreeding effect is so nearly linear with the inbreeding coefficient, this implies that epistasis is not very important in corn yield. Either the genes at different loci act additively on yield or opposite interactive effects cancel each other.

Figure 3.7.1 shows data from Drosophila gathered in a different way. In these experiments the cultures were maintained with as little natural selection as possible and the chromosome number 2 was kept heterozygous.

Then after several generations the chromosome, with its accumulated recessive mutations, was made homozygous. That there is an appreciable synergistic effect is clear from the graph. The mutants presumably accumulate linearly with time, but the homozygous viability decreases somewhat more than linearly. The graph shows only the influence of mutations with small effects; chromosomes with lethal mutations are not included.

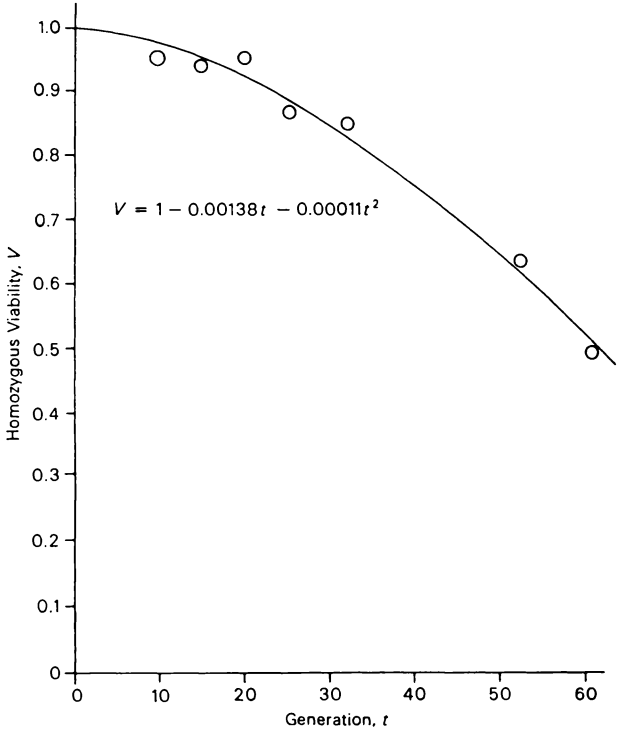


Figure 3.7.1. The viability of Drosophila homozygous for the second chromosome as a function of the number of generations during which mutants were permitted to accumulate. All lethal chromosomes are omitted. There is clearly some reinforcing epistasis for the deleterious effects of the accumulated mutations. Data from T. Mukai.

3.8 Regular Systems of Inbreeding

In Section 3.4 we derived methods for computing the coefficients of inbreeding and consanguinity from pedigrees. The same methods can be used to derive recurrence relations for f in successive generations with regular systems of

inbreeding. A detailed treatement of this subject was first given by Wright (1921) and recent summaries are available (Wright, 1951; Li, 1955*a*). We shall illustrate only some of the simpler, but most important cases.

1. Self-fertilization From Figure 3.8.1 and the principles already

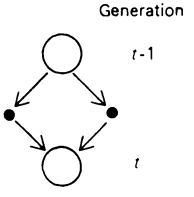


Figure 3.8.1. A diagram of self-fertilization. Zygotes are shown in large circles; gametes by dots.

discussed in Section 3.4,

$$f_t = \tfrac{1}{2}(1 + f_{t-1}), \qquad \textbf{3.8.1}$$

where f_t is the inbreeding coefficient in generation t and f_{t-1} is the coefficient one generation earlier. To obtain the change in heterozygosity we utilize the relation $f_t = (H_0 - H_t)/H_0$ from equation 3.2.4, where H_t and H_0 are the proportions of heterozygosity in generation t and initially ($t = 0$) .This leads to

$$\begin{aligned} H_t &= \tfrac{1}{2}H_{t-1} \\ &= (\tfrac{1}{2})^2 H_{t-2} = (\tfrac{1}{2})^t H_0 . \end{aligned} \qquad \textbf{3.8.2}$$

This confirms the result stated in Section 3.1; with self-fertilization the amount of heterozygosity is reduced by one-half each generation. After 10 generations, only 1/1024 of the loci that previously were heterozygous remain heterozygous.

For a matrix treatment of this, see A.7.

2. Sib Mating Recalling previous definitions, the inbreeding coefficient f_I is the probability that two homologous genes in an individual I are identical by descent; the coefficient of consanguinity f_{IJ} is the probability that two homologous genes, one chosen at random from individual I and the other from individual J are identical.

We let f_t be the inbreeding coefficient of an individual in generation t and g_t the coefficient of consanguinity of two individuals (necessarily sibs after the first generation). The inbreeding coefficient of an individual is the same as the coefficient of consanguinity of his parents; that is, $f_t = g_{t-1}$.

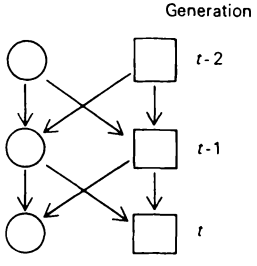


Figure 3.8.2. Continued sib mating.

Two genes in different individuals in generation t came from the same individual in generation $t-1$ with probability 1/2, in which case their probability of identity is $(1+f_{t-1})/2$. They came from different individuals also with probability 1/2, in which case the probability of identity is g_{t-1}. Putting these together, we have

$$\begin{aligned} f_t &= g_{t-1}, \\ g_t &= \tfrac{1}{4}(1+f_{t-1}) + \tfrac{1}{2}g_{t-1}. \end{aligned} \qquad \text{3.8.3}$$

We can, if we wish, eliminate the g's from these two equations and write a relation involving only the inbreeding coefficients.

$$f_t = \tfrac{1}{4}(1 + 2f_{t-1} + f_{t-2}). \qquad \text{3.8.4}$$

Letting $f_t = (H_0 - H_t)/H_0$, we can obtain a recurrence relationship for the heterozygosity

$$H_t = \tfrac{1}{2}H_{t-1} + \tfrac{1}{4}H_{t-2}. \qquad \text{3.8.5}$$

If $H_0 = H_1 = 1/2$, we obtain the sequence 3/8, 5/16, 8/32, 13/64, etc., as given in Table 3.1.2.

In order to get a more general expression for the heterozygote frequency, H_t, in terms of the initial frequency, H_0, we let $f_t = 1 - h_t$ and $g_t = 1 - k_t$. Then equations 3.8.3 become

$$\begin{aligned} h_t &= \quad\quad k_{t-1}, \\ k_t &= \tfrac{1}{4}h_{t-1} + \tfrac{1}{2}k_{t-1}, \end{aligned} \qquad \text{3.8.6}$$

or, in matrix form,

$$\begin{pmatrix} h_t \\ k_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} h_{t-1} \\ k_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}^t \begin{pmatrix} h_0 \\ k_0 \end{pmatrix}. \qquad \text{3.8.6a}$$

The rules for matrix multiplication are given in A.7. However, there is a standard procedure for obtaining h_t directly without having to proceed generation by generation. For an explanation using this same example see A.8.

The characteristic equation for this matrix is

$$\begin{vmatrix} 0-\lambda & 1 \\ \frac{1}{4} & \frac{1}{2}-\lambda \end{vmatrix} = 0,$$

which leads to

$$\lambda^2 - \tfrac{1}{2}\lambda - \tfrac{1}{4} = 0, \qquad \text{3.8.7}$$

of which the largest (and only positive) root is

$$\lambda = \frac{1+\sqrt{5}}{4} = 0.809. \qquad \text{3.8.8}$$

Using both roots we can write an exact expression for h_t for any generation, t. This is shown with this same example in the appendix starting with equation A.8.13.

We are especially interested in the limiting rate of change in heterozygosity, which is given by the largest root, 3.8.8. Thus, after t is sufficiently large, h_t is approximately λh_{t-1}, which is equivalent to

$$H_t \approx \lambda H_{t-1} \approx 0.809 H_{t-1}. \qquad \text{3.8.9}$$

Comparison with the actual values in Table 3.1.2 shows that this equation becomes exceedingly accurate after about five to six generations.

We can also obtain λ directly and simply from 3.8.5. Dividing both sides of the equation by H_{t-1}, and letting $H_t/H_{t-1} = H_{t-1}/H_{t-2} = \lambda$ (which of course implies a constant rate of decrease in heterozygosity), we obtain equation 3.8.7. directly. This procedure has been regularly used by Wright (1933*a*, 1951).

If we desire a system for comparison of two mating systems in regard to their ultimate progress toward homozygosity, we can ask how many generations are required by the two systems to achieve the same reduction in heterozygosity. We can do this by equating $(\lambda_1)^{t_1}$ to $(\lambda_2)^{t_2}$ where t is the number

of generations to achieve a given level of change in heterozygosity and the subscripts refer to the two mating systems. Writing

$$\lambda_1^{t_1} = \lambda_2^{t_2} \qquad \text{3.8.10}$$

and taking logarithms of both sides, we obtain

$$\frac{t_1}{t_2} = \frac{\log \lambda_2}{\log \lambda_1}. \qquad \text{3.8.11}$$

For example, with self-fertilization $\lambda = .5$ while with sib mating $\lambda = .809$. This leads to

$$\frac{t_1}{t_2} = \frac{\log .500}{\log .809} = \frac{-.3010}{-.0921} = 3.27.$$

So we can say that asymptotically one generation of self-fertilization is equal to a little more than three generations of sib mating.

This method of comparing mating systems was first used by Fisher (1949). There are of course many properties of a mating system other than the rate of change in heterozygosity. These can be studied by writing out all possible matings and noting how their frequencies change in successive generations. This quickly becomes unmanageable unless matrix methods are used. The matrix procedure first introduced by Haldane (1937*a*) and further developed by Fisher (1949) can be used to obtain a general solution.

3. More Complicated Systems Figure 3.8.3 shows a pedigree in which, starting with the third generation, all matings are between double first cousins.

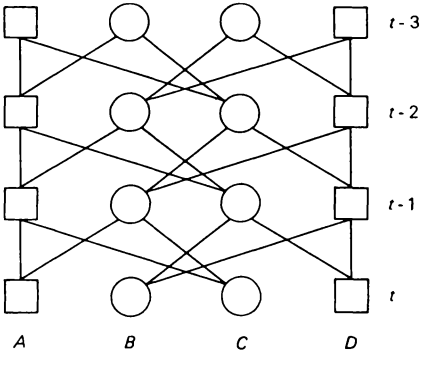


Figure 3.8.3. Repeated mating of double first cousins. This is the maximum avoidance of inbreeding in a population of size 4. $\lambda = .920$.

To write equations for this system we let f_t be the coefficient of inbreeding of an individual in generation t, as before. Notice that there are two kinds of relationship in generation t. We let g_t be the coefficient of consanguinity of nonsibs such as A and B, and j_t be that of sibs such as A and C. Thus

$$g_t = f_{AB} = f_{BC} = f_{CD} = f_{AD},$$
$$j_t = f_{AC} = f_{BD}.$$

From these definitions it follows that:

$$\begin{aligned} f_t &= g_{t-1}, \\ g_t &= \tfrac{1}{2}g_{t-1} + \tfrac{1}{2}j_{t-1}, \\ j_t &= \tfrac{1}{2}(\tfrac{1}{2} + \tfrac{1}{2}f_{t-1}) + \tfrac{1}{2}g_{t-1}. \end{aligned} \qquad \text{3.8.12}$$

Letting $f = 1 - h$, $g = 1 - k$, and $j = 1 - m$ in 3.8.12, we obtain

$$\begin{aligned} h_t &= \quad k_{t-1}, \\ k_t &= \tfrac{1}{2}k_{t-1} + \tfrac{1}{2}m_{t-1} \\ m_t &= \tfrac{1}{4}h_{t-1} + \tfrac{1}{2}k_{t-1}. \end{aligned} \qquad \text{3.8.13}$$

Table 3.8.1. Decrease of heterozygosity with four mating systems, starting from a randomly mating population. The numbers are the ratio of the heterozygosity in generation t to the original heterozygosity, $H_t/H_0 = h_t$.

GENERATION t	SELF-FERTILIZATION $N = 1$	SIB MATING $N = 2$	DOUBLE FIRST-COUSIN MATING $N = 4$	CIRCULAR HALF-SIB MATING $N = 4$
0	1.000	1.000	1.000	1.000
1	.500	1.000	1.000	1.000
2	.250	.750	1.000	.875
3	.125	.625	.875	.813
4	.063	.500	.813	.750
5	.031	.406	.750	.695
6	.016	.328	.688	.644
10	.001	.141	.492	.477
15		.048	.324	.327
20		.017	.213	.224
30		.002	.092	.105
50			.017	.023
λ	.500	.809	.920	.927

These are now homogeneous and can be written in matrix form

$$\begin{pmatrix} h_t \\ k_t \\ m_t \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} h_{t-1} \\ k_{t-1} \\ m_{t-1} \end{pmatrix}. \qquad \text{3.8.13a}$$

The characteristic equation is

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 0 & \frac{1}{2} - \lambda & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\lambda \end{vmatrix} = 0, \qquad \text{3.8.14}$$

which, upon expansion, becomes

$$\lambda^3 - \tfrac{1}{2}\lambda^2 - \tfrac{1}{4}\lambda - \tfrac{1}{8} = 0, \qquad \text{3.8.14a}$$

and the largest root is $\lambda = .9196$.

Some numerical values for heterozygosity with this mating system are given in Table 3.8.1.

Notice that for a population of size 4 this is the system of mating in which mated pairs are least related. A corresponding system in a population of size 8 would be quadruple second-cousin mating. Wright (1921) designated such systems as having maximum avoidance of inbreeding.

Such systems do, in fact, minimize the rate of approach to homozygosity during the initial generations, but somewhat surprisingly there are systems of mating that ultimately have a slower rate of decrease in heterozygosity. An example, for a population of 4, is half-sib mating, or circular mating, as illustrated in Figure 3.8.4.

Letting g_t be the coefficient of consanguinity of individuals one position apart and j_t be that for individuals two positions apart,

$$g_t = f_{AB} = f_{BC} = f_{CD} = f_{AD}$$

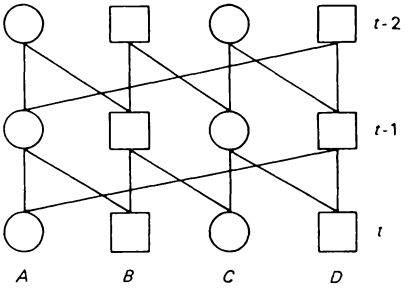


Figure 3.8.4. Half-sib mating in a population of 4, or circular mating. $\lambda = .927$.

and

$$j_t = f_{AC} = f_{BD},$$

we have

$$\begin{aligned} f_t &= g_{t-1}, \\ g_t &= \tfrac{1}{4}(\tfrac{1}{2} + \tfrac{1}{2}f_{t-1}) + \tfrac{1}{2}g_{t-1} + \tfrac{1}{4}j_{t-1}, \\ j_t &= \tfrac{1}{2}g_{t-1} + \tfrac{1}{2}j_{t-1}. \end{aligned} \qquad \text{3.8.15}$$

Substituting $f = 1 - h$, $k = 1 - g$, and $m = 1 - j$ as before leads to

$$\begin{pmatrix} h_t \\ k_t \\ m_t \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{8} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} h_{t-1} \\ k_{t-1} \\ m_{t-1} \end{pmatrix} \qquad \text{3.8.16}$$

with the characteristic equation

$$\lambda^3 - \lambda^2 + \tfrac{1}{16} = 0 \qquad \text{3.8.17}$$

and the largest root is

$$\lambda = .9273.$$

Notice that the eventual rate of decrease in heterozygosity is less in this system than with double first-cousin mating. Referring to Table 3.8.1, we see that the heterozygosity curves for the two systems cross at about the fifteenth generation. The general principle is that more intense inbreeding produces a lower ultimate rate of decrease in heterozygosity, provided that there is no permanent splitting of the population into isolated lines. Conversely, a system that avoids mating of relatives for as long as possible does so at the expense of a more rapid final approach to homozygosis. The breeder therefore may choose a different system of mating if he is more interested in maximum heterozygosity during the initial generations than in the long-time future population. An extension of the procedures of this section to larger populations than $N = 4$ has been given by Kimura and Crow (1963). Robertson (1964) and Wright (1965*a*) have shown that many of these results can be brought together very generally under a single point of view. For other types of mating systems see Wright (1921, 1951). Many of Wright's earlier results are summarized by Li (1955).

4. Partial Self-fertilization All the examples discussed thus far lead eventually to complete homozygosity. This is not always the case, and we shall now consider one such example. This is the simple, yet important, case where a certain fraction each generation are self-fertilized and the remainder are mated at random, a situation found in several plant species.

Let S be the fraction of the population that is produced by self-fertilization; then $1 - S$ is the fraction that is produced by random mating. From 3.8.1 we can write the expected recurrence relation for f as

$$f_t = S[(1 + f_{t-1})/2] + (1 - S)(0) = \frac{S}{2}(1 + f_{t-1}). \qquad \text{3.8.18}$$

This assumes that the plants to be self-fertilized each generation are a random sample of the population; for example, there is no tendency for the progeny of self-fertilized plants to be self-fertilized.

Substituting $f_t = (H_0 - H_t)/H_0$ from 3.2.4 into 3.8.18 we get

$$H_t = H_0(1 - S) + \frac{S}{2} H_{t-1}. \qquad \text{3.8.18a}$$

Subtracting $2(1 - S)H_0/(2 - S)$ from both sides and simplifying,

$$\begin{aligned} H_t - \frac{2(1-S)}{2-S} H_0 &= \frac{S}{2}\left[H_{t-1} - \frac{2(1-S)}{2-S} H_0\right] \\ &= \left(\frac{S}{2}\right)^2\left[H_{t-2} - \frac{2(1-S)}{2-S} H_0\right] \\ &= \left(\frac{S}{2}\right)^t\left[H_0 - \frac{2(1-S)}{2-S} H_0\right]. \end{aligned} \qquad \text{3.8.19}$$

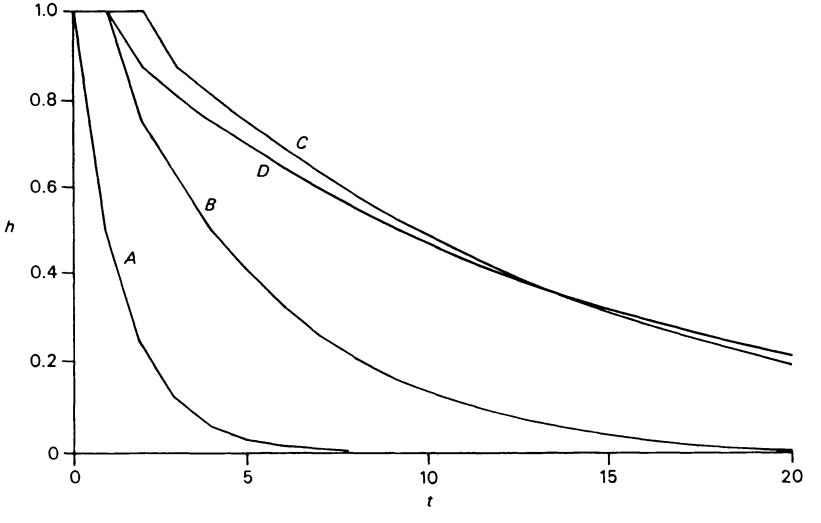


Figure 3.8.5. Change in heterozygosity with four mating systems. A. Self-fertilization; B. Sib mating; C. Double first-cousin mating; D. Circular half-sib mating. The ordinate is the heterozygosity relative to the starting population; the abscissa is the time in generations.

Since $(S/2)^t$ approaches 0 as t becomes large, the heterozygosity approaches a limit where the heterozygosity is a fraction $2(1 - S)/(2 - S)$ of its original value. The rate of approach is such that the departure from the equilibrium value is decreased by a fraction $1 - S/2$ each generation. Notice that when $S = 1$ we get the usual formula for self-fertilization.

This situation is striking in that unless S is large there is almost no cumulative effect; most of the effect occurs in the first generation. For example, with 10% self-fertilization, the initial heterozygosity is reduced by 5% in the first generation, but even when equilibrium is reached the reduction is only 5.3%!

5. Repeated Backcrossing to the Same Strain Frequently a plant breeder may wish to introduce one or more dominant genes from an extraneous source into a standard variety. For example he may have a highly desirable variety, except for its being susceptible to some disease. The resistant gene may exist in another strain which is less desirable in other respects. He can introduce this gene by crossing the two strains and then repeatedly crossing resistant plants to the susceptible strain. In this way the resistant gene is inserted into a genetic background that becomes more like the susceptible strain with each backcross. As another example, a mouse breeder may wish to introgress a new histocompatibility gene into a standard inbred strain.

It is clear that in recurrent backcrossing the number of loci that contain genes from both strains is reduced by half each generation. Thus, after t generations, a fraction equal to $1 - .5^t$ are from the recurrent parental strain. After seven generations less than 1% of the loci contain a gene from the other parental strain. If the recurrent parental strain is homozygous, the heterozygosity will reduce by half each generation, as with self-fertilization.

However, genes that are linked to the resistance or histocompatibility gene will tend to remain heterozygous. The question of how large a linked region will remain after a certain number of generations of backcrossing has been investigated by Haldane (1936) and Fisher (1949).

Consider a chromosome segment on one side of the selected factor and let the length of the segment be $100x$ map units in length (see Figure 3.8.6).

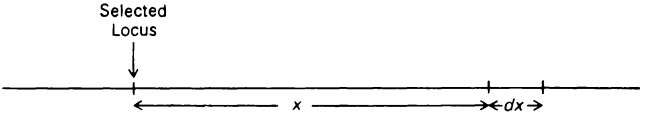


Figure 3.8.6. A chromosome segment. One locus is selected during recurrent backcrossing and the problem is to determine the length of chromosome to the right of this locus that will be intact after t generations.

If there is no interference, the probability of no crossover in this interval in one generation is e^{-x} (see Appendix A.5.6). The probability of no crossover in t generations is e^{-tx}. The chance of a crossover in the small interval x to $x + dx$ is dx, if we take this interval small enough that multiple crossovers can be ignored. The probability of a crossover in the interval dx sometime during t generations is tdx. Thus the probability after t generations of having had a crossover in the interval dx but not in the interval x is $e^{-tx}tdx$. Then the mean value of the intact interval x is

$$\bar{x} = \int_0^\infty e^{-tx}tx\,dx = \frac{1}{t}, \qquad \text{3.8.20}$$

or $100/t$ map units.

For example, after 20 generations the average segment remaining intact would be 5 units on each side of the selected locus, or 10 units altogether. The derivation has assumed no interference, but for short regions the interference pattern makes very little difference.

There is a closely related question that can be answered very simply. We ask: What is the mean number of backcross generations that a gene will remain linked to the selected locus when the recombination probability is r? The number may be derived as follows. The probability that the genes will remain linked for t generations and then recombine in the next generation is $(1 - r)^t r$. The average number of generations until they become separated is

$$\begin{aligned}\bar{t} &= r + 2(1 - r)r + 3(1 - r)^2 r + 4(1 - r)^3 r + \cdots \\ &= r(1 + 2y + 3y^2 + 4y^3 + \cdots),\end{aligned}$$

where $y = 1 - r$.

Notice that the quantity in parentheses is the derivative of the series $1 + y + y^2 + y^3 + y^4 + \cdots = 1/(1 - y)$. Therefore

$$\bar{t} = r\frac{d}{dy}\left(\frac{1}{1 - y}\right) = r\frac{1}{(1 - y)^2} = \frac{1}{r}. \qquad \text{3.8.21}$$

The value of x (the map distance in 3.8.20) and r (the recombination value in 3.8.21) will not in general be the same, but will become more and more similar as x becomes small enough that multiple crossovers can be neglected.

3.9 Inbreeding with Two Loci

The idea of the inbreeding coefficient can be extended to cover multiple loci. The principal item of interest is that inbreeding may cause the association of two or more recessive traits. This can happen in either of two ways: (1) with

unlinked loci but nonuniform inbreeding, and (2) with uniform inbreeding and linkage—and, of course, because of both linkage and nonuniform inbreeding.

Our treatment follows very closely the method of Haldane (1949). We consider first unlinked loci in linkage equilibrium.

Let p_i be the frequency of allele A_i and r_k that of the independent allele B_k. The frequency of A_iA_i homozygotes (from 3.2.2) is $p_i^2(1-f) + p_i f = p_i^2 + p_i(1-p_i)f$ with a similar formula for B_kB_k, $r_k^2 + r_k(1-r_k)f$.

Since the loci are in linkage equilibrium the frequency of $A_iA_iB_kB_k$ is the product, or

$$(p^2 + pqf)(r^2 + rsf) = p^2r^2 + (pqr^2 + p^2rs)f + pqrsf^2,$$

where $q = 1 - p$ and $s = 1 - r$, and the subscripts have been dropped for simplicity since we are discussing only one allele at each locus.

In a population with different values of f from individual to individual the frequency of the double homozygote, $P(AABB)$, is

$$p^2r^2 + (pqr^2 + p^2rs)\bar{f} + pqrs\overline{f^2}.$$

But, by the definition of the variance

$$\overline{f^2} = \bar{f}^2 + V_f.$$

Hence

$$P(AABB) = (p^2 + pq\bar{f})(r^2 + rs\bar{f}) + pqrsV_f. \qquad 3.9.1$$

Suppose that in the human population two recessive genes each have a frequency of .01, and that 1% of the marriages are between cousins while the rest are random. Then $\bar{f} = 1/1600$, $V_f = 1/25600 - (1/1600)^2 = 3.87 \times 10^{-5}$, $(p^2 + pq\bar{f})(r^2 + rs\bar{f}) = (.00011)^2 = 11.28 \times 10^{-9}$, and $pqrsV_f = 3.79 \times 10^{-9}$. Hence the frequency of double homozygotes in the population is 15.07×10^{-9}, just about 4/3 what it would be if all families had the same inbreeding coefficient ($V_f = 0$).

To consider the second case we extend the inbreeding and consanguinity coefficient, f, to two loci. Let F be the probability that both of two loci carry identical alleles. F will be a function, not only of the pedigree, but also of the amount of recombination between the two loci.

A pair of uniting gametes can be: (1) identical for both loci with probability F; (2) identical for only the A locus, probability $f - F$; (3) identical for only the B locus, also $f - F$; or (4) identical for neither, with probability $1 - 2f + F$.

If the population has reached equilibrium with respect to the linkage phases, the frequency of AB gametes will be the product of the frequencies of the A and B genes.

Thus

$$\begin{aligned} P(AABB) &= prF + pr^2(f - F) + p^2r(f - F) + p^2r^2(1 - 2f + F) \\ &= pr(Fqs + fqr + fps + pr) \\ &= pr[(qsf^2 + fqr + fps + pr) + Fqs - f^2qs] \\ &= (p^2 + fpq)(r^2 + frs) + \phi pqrs, \end{aligned} \qquad 3.9.2$$

where $\phi = F - f^2$. The frequency of any other genotype is obtainable the same way. For a summary table see Haldane (1949).

The quantity ϕ measures the degree of association in identity beyond that which would occur if the loci were independent. A comparison of 3.9.1 and 3.9.2 shows that ϕ and V_f enter the formula the same way. The effects are approximately additive, so that,

$$P(AABB) = (p^2 + fpq)(r^2 + frs) + (\phi + V_f)pqrs \qquad 3.9.3$$

with analogous expressions for other genotypes.

There remains the problem of computing F (or ϕ) from a pedigree. There is no simple algorithm for F comparable to that for f. However, Denniston (1967) has discovered a method for computing F for any degree of relationship. The simple relationships are not hard to do by strong-arm methods.

Consider first two gametes produced from the same individual. They will be identical if both sets of alleles are derived from the same chromosome, there having been no recombination between A and B, or if the A's come from one chromosome and the B's from the other, which demands that each gamete be the result of a recombination. Letting c be the frequency of recombinant gametes, and d the frequency of nonrecombinant ($c + d = 1$), the probability of double autozygosity is

$$F = \tfrac{1}{2}(c^2 + d^2),$$
$$f = \tfrac{1}{2},$$
$$\phi = \tfrac{1}{4}(c - d)^2.$$

With parent-offspring mating, in order to have double autozygosity one of the identical gametes from the common ancestor must pass intact through another generation, which has probability $d/2$. Therefore

$$F = \tfrac{1}{4}(c^2 + d^2)d,$$
$$f = \tfrac{1}{4},$$
$$\phi = \tfrac{1}{16}(d - c)(3d^2 + c^2).$$

For half-sibs, each of two gametes must pass intact through a generation, which has probability $d^2/4$. Thus

$$F = \tfrac{1}{8}(c^2 + d^2)d^2,$$
$$f = \tfrac{1}{8},$$
$$\phi = \tfrac{1}{64}(8d^2 + 8c^2d^2 - 1).$$

With full-sib mating a complication arises since there must be consideration of the possibility that the A's come from one grandparent and the B's from the other. For the details we refer to Haldane's paper; his results for several kinds of relationship are given in Table 3.9.1.

Table 3.9.1. Values of the inbreeding coefficient for two loci, where c is the proportion of recombination and $d = 1 - c$. The parents are not inbred and are in linkage equilibrium. From Haldane (1949).

RELATIONSHIP OF PARENTS	F	f	$\phi = F - f^2$
Identical	$\frac{1}{2}(c^2 + d^2)$	$\frac{1}{2}$	$\frac{1}{4}(c - d)^2$
Parent-offspring	$\frac{1}{4}(c^2 + d^2)d$	$\frac{1}{4}$	$\frac{1}{16}(d - c)(3d^2 + c^2)$
Full sibs	$\frac{1}{8}(2d^4 + 2c^2d^2 + c^2)$	$\frac{1}{4}$	$\frac{1}{16}(4d^4 + 4c^2d^2 + 2c^2 - 1)$
Half-sibs	$\frac{1}{8}(c^2 + d^2)d^2$	$\frac{1}{8}$	$\frac{1}{64}(8c^2d^2 + 8d^4 - 1)$
Uncle-niece	$\frac{1}{16}(2d^4 + 2d^2c^2 + c^2)d$	$\frac{1}{8}$	$\frac{1}{64}(8d^5 + 8c^2d^3 + 4c^2d - 1)$
First cousins	$\frac{1}{32}(2d^4 + 2c^2d^2 + c^2)d^2$	$\frac{1}{16}$	$\frac{1}{256}(16d^6 + 16c^2d^4 + 8c^2d^2 - 1)$
Double half-cousins	$\frac{1}{128}(8d^6 + 8c^2d^4 + c^2)$	$\frac{1}{16}$	$\frac{1}{256}(16d^6 + 16c^2d^4 + 2c^2 - 1)$

It is interesting that several relationships with the same f may have different F's. Compare, for example, ordinary first cousins with double half-cousins.

As a numerical example, assume $p = r = .01$. The frequency of AA or BB in the progeny of cousins is .00072 compared with .00010 with random mating. In the absence of linkage the frequency of $AABB$ is $(7.2 \times 10^{-4})^2 = 5.2 \times 10^{-7}$; with 10% recombination the frequency is $5.2 \times 10^{-7} + \phi pqrs = 34.6 \times 10^{-7}$.

However, such a sharp rise is expected only with rare genes and close linkage. Some association of rare recessive traits should be expected as a consequence of inbreeding and linkage, but probably not of sufficient magnitude to be detected in ordinary circumstances.

The effect of inbreeding on quantitative traits involving two loci when epistasis and linkage disequilibrium are present may be obtained by using expressions like 3.9.3. for all the genotypes. Nei (1965) did this for the mean fitness, as an extension of 3.6.3, which was derived by assuming linkage equilibrium.

3.10 Effect of Inbreeding on the Variance

In Section 3.6 we considered the effect of inbreeding on quantitative traits, especially as the mean is affected. In addition to its effect on the population mean, inbreeding also has an effect on the variance.

Consider again the single-locus model of Section 3.6. It is convenient to let $Y = 0$, as this does not change the conclusion and saves some troublesome algebra. To assess the total effect we consider first a single locus.

GENOTYPE	A_1A_1	A_1A_2	A_2A_2
FREQUENCY	$p_1^2(1-f)+p_1 f$	$2p_1p_2(1-f)$	$p_2^2(1-f)+p_2 f$
CONTRIBUTION TO TOTAL PHENOTYPE	$-A$	D	A

We write the mean as a function of f,

$$\begin{aligned}\overline{Y}_f &= (1-f)(p_2^2 A + 2p_1p_2 D - p_1^2 A) + f(p_2 A - p_1 A)\\ &= (1-f)\overline{Y}_0 + f\overline{Y}_1,\end{aligned} \qquad 3.10.1$$

where $\overline{Y}_0$ is the mean value with random mating ($f = 0$) and $\overline{Y}_1$ is the mean value with complete homozygosity ($f = 1$).

Notice that this can be written as

$$\overline{Y} = \overline{Y}_0 + (\overline{Y}_1 - \overline{Y}_0)f, \qquad 3.10.2$$

illustrating once again, as was shown in Section 3.6, that the phenotype is a linear function of f. Equation 3.10.2 is the same as 3.6.1.

We now write the expression for the contribution of this locus to the total population variance.

$$\begin{aligned}V_f &= (1-f)(p_1^2A^2 + 2p_1p_2 D^2 + p_2^2 A^2) + f(p_1A^2 + p_2 A^2) - \overline{Y}^2\\ &= (1-f)(V_0 + \overline{Y}_0^2) + f(V_1 + \overline{Y}_1^2) - [(1-f)\overline{Y}_0 + f\overline{Y}_1]^2\\ &= (1-f)V_0 + fV_1 + f(1-f)(\overline{Y}_0 - \overline{Y}_1)^2,\end{aligned} \qquad 3.10.3$$

where V_0 and V_1 are the variance with random mating ($f = 0$) and complete inbreeding ($f = 1$). This formula comes from Wright (1951).

This shows that, unlike the mean, the variance is not a linear function of f but is quadratic.

Later, in considering the effects of selection, we shall see that the rate of gene frequency change depends on the additive component of the gene effect. Therefore it is of special interest to examine the effect of inbreeding on the mean and variance of a locus without dominance.

Without dominance there is no effect of inbreeding on the mean, as can be seen by letting $D = 0$ in 3.10.1 or 3.6.1. $\bar{Y}_0 = \bar{Y}_1 = A(p_2 - p_1)$,

$$V_0 = p_1^2 A^2 + p_2^2 A^2 - \bar{Y}_0^2 = 2p_1 p_2 A^2, \qquad 3.10.4$$

$$V_1 = p_1 A^2 + p_2 A^2 - \bar{Y}_1^2 = 4p_1 p_2 A^2 = 2V_0. \qquad 3.10.5$$

Thus, with no dominance the total variance is

$$V_f = (1 - f)V_0 + fV_1 = V_0(1 + f), \qquad 3.10.6$$

which in this case is linear rather than quadratic. A population within which there is some consanguineous mating has an increase in genetic variance proportional to f, provided there is no dominance.

If there is subdivision of the population into inbred strains, we can measure the variance within and between such strains. From 3.10.4. we see that the variance of a randomly mating population for a nondominant locus is proportional to the amount of heterozygosity, measured by $2p_1 p_2$. In a strain of inbreeding coefficient f the heterozygosity is reduced by a fraction f; hence the variance is reduced by this amount.

The effect of any inbreeding which is of a type that divides the population into isolated groups will be to decrease the variance within groups, to increase the variance between groups, and increase the total variance. This happens, for example, in the development of a series of inbred lines of plants or livestock.

These conclusions apply only to the special case of no dominance. With varying degrees of dominance explicit formulae for variance within and between lines become more difficult (Robertson, 1952; Wright, 1952). This is discussed in Chapter 7.

A case of special interest arises when there are recessive alleles of low frequency. In this circumstance the variance within a subpopulation increases during the early stages of the inbreeding process, despite the fact that the population is becoming more homozygous. The reason is that inbreeding brings out previously hidden recessive factors which now can contribute to the phenotypic variance. This increase will continue until it is eventually offset by the increasing homogeneity in gene content from individual to individual within the subpopulation.

3.11 The Inbreeding Effect of a Finite Population

As was mentioned in the introduction to this chapter, there is a decrease of heterozygosity in a finite population even if there is random mating within the group. Each generation may be regarded as being made up of $2N$ gametes drawn from the previous generation and which combine to make the N individuals of this generation. The gene frequency will therefore change somewhat, the amount depending on the smallness of the population. The change will have a variance of $p(1-p)/2N$ where p is the frequency of the gene under consideration in the parents. We have assumed here that any particular successful gamete is equally likely to have come from any one of the parents, so that the situation is analogous to binomial sampling of $2N$ gametes (see A.5.5).

It might not seem obvious that random changes in gene frequency, which can be in either direction, will on the average cause a net decrease in heterozygosity and increase in homozygosity. One way to visualize this is to regard each of the $2N$ genes at a locus in the N parents as individually labeled. The effect of random processes in drawing (with replacement) from these is that some will be omitted entirely while others will be drawn more than once. Thus there will be a certain amount of identity next generation. If the process continues long enough, all the genes will be descended from a single individual gene and complete autozygosity will be attained.

The distribution of the probabilities of various gene frequencies during this process is a difficult problem. On the other hand, the average change in heterozygosity is uniform and easily derived. It can be expressed as a function of the inbreeding coefficient.

Consider first a population with completely random mating, including self-fertilization. As before, we consider a single locus. Imagine that the progeny are produced by drawing random pairs of gametes from an infinite pool to which each parent had contributed equally (or, if the pool is finite, the drawing is with replacement). Two gametes then have a chance $1/2N$ of carrying identical genes, since the N diploid parents have $2N$ genes at this locus. Two gametes have a chance of $1 - 1/2N$ of carrying different parental genes.

In the first case, the probability of the genes being identical is of course 1. In the second case the probability of their being identical is f_{t-1}, the inbreeding coefficient of an average individual in the previous generation. The reason for this is that the two alleles were drawn at random from the parent generation, and since the parents were the result of random mating, the probability of any two alleles being identical is the same as that for two in the same zygote; the latter is, by definition, f_{t-1}. Therefore,

$$f_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) f_{t-1}. \qquad \text{3.11.1}$$

Recalling that H_t, the heterozygosity at time t, is $H_0(1 - f_t)$, we obtain by substitution in 3.11.1

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1} = \left(1 - \frac{1}{2N}\right)^t H_0 . \qquad \text{3.11.2}$$

The result is the very simple one; despite the complexities in the changes in gene frequencies, the average heterozygosity decreases by a fraction $1/2N$ each generation.

As stated earlier, this formula assumes completely random mating (that is, random combination of gametes) including the possibility of self-fertilization. Note that, when $N = 1$, the results agree with the formula for self-fertilization, as expected (3.8.1 and 3.8.2).

No Self-fertilization We now assume that gametes are combined at random, but with the restriction that two uniting gametes cannot come from the same parent. We let f_t be the inbreeding coefficient at time t, as before, and let g_t be the coefficient of consanguinity of two different randomly chosen individuals in generation t. The inbreeding coefficient in generation t is clearly the same as the consanguinity coefficient in generation $t - 1$, since mating is at random. To get the consanguinity coefficient in generation t, we note that the two chosen genes (one from each individual) have come from the same individual in the previous generation with probability $1/N$ and from different individuals with probability $1 - 1/N$. In the first case the probability of identity is $(1 + f_{t-1})/2$ as explained earlier (see, for example, 3.4.1 and 3.8.1); in the second case it is g_{t-1}. Putting these together, we obtain

$$\begin{aligned} f_t &= g_{t-1}, \\ g_t &= \frac{1}{2N}(1 + f_{t-1}) + \frac{N-1}{N} g_{t-1}. \end{aligned} \qquad \text{3.11.3}$$

Substituting the first into the second gives

$$f_{t+1} = \frac{1}{2N}(1 + f_{t-1}) + \frac{N-1}{N} f_t ,$$

which, after going back one generation and rearranging, is

$$f_t = f_{t-1} + (1 - 2f_{t-1} + f_{t-2})/2N. \qquad \text{3.11.4}$$

Separate Sexes If the individuals in generation $t - 1$ consist of N_m males and N_f females adding up to a total of N individuals, only a slight modification is required. The two sexes necessarily make the same contribution to later generations, since each fertilization event involves one maternal and one paternal gamete. Therefore the probability that two genes in different

individuals in generation t are both derived from a male in generation $t-1$ is $1/4$; and that they came from the same male is $1/4N_m$. Likewise, the probability of their coming from the same female is $1/4N_f$. Then the probability that the two genes came from the same individual in generation $t-1$ is

$$\frac{1}{4N_m} + \frac{1}{4N_f} = \frac{1}{N_e}, \qquad 3.11.5$$

where N_e is the *effective* number of individuals in generation $t-1$.

This means that equations 3.11.3 and 3.11.4 are correct with separate sexes; it is only necessary to use N_e instead of N in the formula. Notice that when $N_m = N_f = N/2$, then $N_e = N$. When the two sexes are equally frequent, the actual number and the effective number are the same.

The effective population number is used with a wider definition than this single example would suggest. If there are fluctuations in the population number from time to time, or if the distribution of number of progeny per parent is nonbinomial, or if there is any other kind of deviation from the idealized model that we have assumed, then it is conventional to define for that population an effective number. The effective number then, as in this example, is the size of an ideally behaving population that would have the same homozygosity increase as the observed population. We shall return to this subject in Section 3.13 and in Chapter 7.

Notice that, when the number of males and females is equal, there is no distinction (as far as rate of homozygosity change caused by random gene-frequency drift) between a population with separate sexes and a hermaphroditic population without self-fertilization.

Returning to equation 3.11.4 we can make the same substitution as before, $H_t = H_0(1-f_t)$, and write an equation for the heterozygosity in successive generations. This yields

$$H_t = \frac{N-1}{N} H_{t-1} + \frac{1}{2N} H_{t-2}. \qquad 3.11.6$$

Notice that when $N = 2$ equations 3.11.4 and 3.11.6 reduce to the equations for sib-mating (3.8.4 and 3.8.5) as expected. This comparison may be misleading in one regard, however. Part of the increase in homozygosity in larger populations is due to the fact that different members of the population leave different numbers of descendants. The general formula takes this into account, and with larger N it turns out that about half of the increase in homozygosity is caused by the restricted number of parents and the other half by differential contribution of offspring. Equation 3.11.6 assumes under these circumstances that the distribution of progeny-number is binomial, as follows from the assumption that each progeny gene has an equal chance of having come from any parent. However, as is seen, the formula is still

correct in the limiting case of sib mating, despite the fact that the progeny number per parent is necessarily constant in this case.

The ultimate ratio by which H_t decreases each generation is given by the larger root of the quadratic equation

$$\lambda^2 - \left(\frac{N-1}{N}\right)\lambda - \frac{1}{2N} = 0. \qquad 3.11.7$$

This equation is easily obtained by setting $H_t/H_{t-1} = H_{t-1}/H_{t-2} = \lambda$ in 3.11.6.

The relevant solution to the equation is

$$\lambda = \frac{N - 1 + \sqrt{N^2 + 1}}{2N}, \qquad 3.11.8$$

or approximately, unless N is very small,

$$1 - \lambda \sim \frac{1}{2N + 1}, \qquad 3.11.9$$

or, when N is large,

$$1 - \lambda \sim \frac{1}{2N}. \qquad 3.11.10$$

Hence, in a moderately large population the average heterozygosity is reduced by about $1/2N$ per generation whether there are separate sexes or not, provided that the two sexes are equal in frequency.

The accuracy of the approximation 3.11.9 for small N can be seen by comparison with the data in Table 3.1.2. The last column in this table gives the proportion by which the heterozygosity is reduced, which is to be compared with the approximation $1/(2N + 1)$. Within 6 generations the approximation is quite good: with $N = 2$, $1/(2N + 1) = 1/5$ or .20 whereas the exact answer is .19. So if the population is anywhere near a steady state, 3.11.9 is very good even for quite small numbers.

The results of this section were all obtained first by Wright (1931) by his method of path coefficients. A summary of this method and its applications to problems in population genetics is given in his Galton Lecture (Wright 1951).

3.12 Hierarchical Structure of Populations

At the beginning of this chapter we emphasized that inbreeding can occur under two quite different circumstances. In both cases there is a decrease in heterozygosity measured by f, but some of the consequences are quite different.

In the first case we can have a large population within which isolated consanguineous matings occur. The inbreeding coefficient measures the average decrease in heterozygosity and equations 3.2.1—3.2.3 give the expected genotypic frequencies for given gene frequencies. As soon as random mating occurs the inbreeding coefficient returns to zero. An example is the case of partial self-fertilization discussed in Section 3.8; as soon as self-fertilization is prevented, the original heterozygosity is restored. (We should perhaps mention that here, as throughout this chapter, we are ignoring the effects of selection and mutation.)

On the other hand, there may be inbreeding because of restriction of population number, even though mating is at random within the population. The average heterozygosity within the population is reduced and the individuals become more closely related, both measured by f (i.e., f_I and f_{IJ}). But, as emphasized in Section 3.11, the increased homozygosity within the population is not due to departure from Hardy–Weinberg ratios, but to changes in the gene frequencies. The change is such as to make the value of $2p_1p_2$ (or $\Sigma p_i p_j$ with multiple alleles) decrease in proportion to f. The formulae 3.2.1–3.2.3 hold only in the sense of giving the genotype frequencies averaged over a whole series of such populations.

In this situation the loss of heterozygosity within the population is permanent and could be restored only by crossing with other populations. Repeated self-fertilization and sib mating constitute extreme cases of small populations.

There are circumstances in which the two effects are combined. For example, we might inquire about the inbreeding coefficient of an animal whose parents were sibs in an isolated population of effective size N. His homozygosity will be greater than if his parents were sibs in a large population. We now derive a procedure to handle this situation. The conclusions were first reached by Wright (1943, 1951) by a different method.

Let S be a subpopulation derived by isolating a finite number of individuals from a large total population T. For example, S could be a breed or a strain isolated from a foundation stock T. Or, S could be one of a series of geographically isolated subpopulations of a large population T. Let I be an individual within subpopulation S.

We now define f_{IS} as the probability that two homologous genes in I are derived from the same gene in a common ancestor within the subpopulation. Let f_{ST} be the probability that two homologous genes, chosen at random from the subpopulation, are both descended from a gene in the subpopulation. We let f_{IT} be the overall probability of identity in individual I.

The probability of nonidentity is then the product of two terms: $(1 - f_{IS})$, the probability that the two genes do not both come from a gene in a known common ancestor in the population, and $(1 - f_{ST})$, the probability that if the

two genes are randomly chosen from within the population they will not be identical because of some more remote relationship. Therefore,

$$1 - f_{IT} = (1 - f_{IS})(1 - f_{ST}), \qquad 3.12.1$$

or

$$f_{IT} = f_{ST} + (1 - f_{ST})f_{IS}. \qquad 3.12.1a$$

For example, what is the inbreeding coefficient of a child whose parents were cousins on an island whose population is descended from a shipwreck 10 generations ago? Assume for simplicity that there were 50 survivors, equally divided between the two sexes, and that the population has remained of this size and sex distribution since that time. From equation 3.11.2 (accurate enough, although it would be better to use $2N + 1$ instead of $2N$) we have

$$1 - f_{ST} = \left(1 - \frac{1}{100}\right)^{10} = 0.9045$$

and

$$1 - f_{IS} = 1 - \frac{1}{16} = 0.9375,$$

1/16 being the inbreeding coefficient of a child of a cousin marriage. From 3.12.1,

$$1 - f_{IT} = (0.9045)(0.9375) = 0.848$$

and

$$f_{IT} = 0.152,$$

compared with 0.0625 for cousin marriage in an infinite population.

Students of animal breeding will enjoy Wright's (1951 and earlier) application of these methods to the history of Shorthorn cattle. He showed that there was a substantial increase in the inbreeding coefficient f_{IT} of British Shorthorns, almost entirely due to f_{ST} and hardly at all because of consanguineous matings within the breed (f_{IS}).

Similar analyses are possible in human isolates if accurate pedigree records are available. If they are not, it is sometimes possible to get reasonably satisfactory information from marriage records. Since a person's name is inherited as if it were linked to the father's Y chromosome, marriages between persons of the same surname (isonymous) can be used as indications of consanguinity. Using this procedure on the Hutterite population, Crow and Mange (1965) were able to show that f_{ST} in this population is quite appreciable, about 4%, but that f_{IS} is not significantly different from 0.

In other words, the increased homozygosity is due almost entirely to the small effective size of the population and not at all to nonrandom marriage within the isolate.

Notice that 3.12.1 may be written as

$$1 - f_{IS} = \frac{H_0(1 - f_{IT})}{H_0(1 - f_{ST})}, \qquad 3.12.2$$

which shows that $1 - f_{IS}$ is a measure of the heterozygosity of an individual in the subpopulation relative to that of an individual derived from random mating in the same subpopulation.

As was mentioned earlier, the inbreeding coefficient can be described in terms of correlation as well as in terms of gene identity. In fact Wright's original definition and derivation was through correlation analysis. One advantage of the correlation interpretation, as opposed to the probabilistic, is that negative values have a meaning. This is especially useful in this section. F_{IS} is the correlation between homologous genes in an individual relative to genes chosen at random from his subpopulation. F_{IT} is the correlation between homologous genes in an individual relative to the whole population. F_{ST} is the correlation between randomly chosen genes in the subpopulation relative to the total population.

F_{ST} is necessarily positive, but the others need not be. For example, if there is specific avoidance of matings between related individuals within the subpopulation, F_{IS} may be negative.

Wright defined f_{ST} as the correlation between two random gametes from the same subpopulation and derived the relation

$$f_{ST} = \frac{V_p}{\bar{p}(1 - \bar{p})}, \qquad 3.12.3$$

where $\bar{p}$ and V_p are the mean and variance of the gene A among the subpopulations. We can derive this as follows, using 3.12.1.

Consider a pair of alleles, A and A', and let p be the frequency of A. Then the frequency of a heterozygote in the subgroup with frequency p is $2p(1 - p)(1 - f_{IS})$. Thus the frequency of heterozygotes in the total population is

$$H = E\{2p(1 - p)(1 - f_{IS})\}$$

where $E\{\ \}$ designates taking the expectation or average over all subpopulations. If the variation of gene frequencies among subgroups is independent of f_{IS}, then

$$H = 2(1 - f_{IS})E\{p(1 - p)\}.$$

But

$$E\{p(1-p)\} = E\{p - p^2\}$$
$$= \bar{p} - \bar{p}^2 - V_p$$

(see A.2.2 or 2.9.2). Thus

$$H = 2(1 - f_{IS})(\bar{p} - \bar{p}^2 - V_p). \qquad 3.12.4$$

On the other hand, from the definition of f_{IT},

$$H = 2(1 - f_{IT})\bar{p}(1 - \bar{p}). \qquad 3.12.5$$

Equating these two expressions gives

$$(1 - f_{IT}) = (1 - f_{IS})\left(1 - \frac{V_p}{\bar{p}(1 - \bar{p})}\right).$$

Comparing this with 3.12.1 gives us the desired expression, 3.12.3.

Notice that if there is no inbreeding in the subpopulation ($f_{IS} = 0$), then

$$H = 2\bar{p}(1 - \bar{p}) - 2V_p. \qquad 3.12.6$$

This, although expressed in terms of heterozygotes rather than homozygotes, is the same as we described as Wahlund's principle in 2.9.2.

Nei (1965) extended this procedure to cover the case of multiple alleles and has shown that

$$f_{ST} = \frac{-\mathrm{Cov}_{ij}}{\bar{p}_i \bar{p}_j}, \qquad 3.12.7$$

where $\bar{p}_i$ and $\bar{p}_j$ are the mean frequencies of the alleles A_i and A_j and Cov_{ij} is the covariance of their frequencies. (See problem 22, Chapter 1.)

Nei and Imaizumi (1966) applied these formulae to study the differentiation of *ABO* blood group gene frequencies among the prefectures of Japan. They obtained 6 values of f_{ST} that were in close agreement, suggesting that the local differentiation of gene frequencies was mainly random.

Equation 3.12.1. can easily be extended to include subdivisions of a subdivision in a hierarchy. For example,

$$(1 - f_{IT}) = (1 - f_{IR})(1 - f_{RS})(1 - F_{ST}), \qquad 3.12.8$$

where R is a subpopulation of S which is a subpopulation of T. Then f_{RS} is the probability that two randomly chosen homologues in R are identical because of common ancestry within R and f_{ST} is the probability that two randomly chosen genes in S are identical because of common ancestry within S.

3.13 Effective Population Number

In our discussion of random drift in gene frequencies and the resulting decrease in heterozygosity in a finite population we have assumed, in addition to random mating, that the expected number of progeny is the same for each individual. That is to say, we regarded a successful gamete as being equally likely to have come from any individual in the parent generation. We also assumed that the population size remains constant from generation to generation. Neither of these conditions is likely to be met in nature.

In equation 3.11.5 we introduced the concept of effective population number by noting that, if the numbers of males and females differ, we can replace N in the formulae such as 3.11.6 by N_e, the effective number, where

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f}$$

or

$$N_e = \frac{4N_m N_f}{N_m + N_f}. \qquad 3.13.1$$

Notice that the value of N_e is influenced much more by the smaller than by the larger of N_m and N_f. For example, if $N_m = 1$ and $N_f = 100$, N_e is about 4; so, in highly polygynous species the number of males is much more important than the number of females in determining the amount of random drift in the population. N_e is proportional to the harmonic mean of N_m and N_f, and the harmonic mean is more strongly influenced by the smaller values.

We can also use the concept of effective population number when the true size varies with time. From 3.11.2 we note that the heterozygosity after t generations, when N varies from generation to generation, is

$$\begin{aligned}\frac{H_t}{H_0} &= \left(1 - \frac{1}{2N_0}\right)\left(1 - \frac{1}{2N_1}\right)\left(1 - \frac{1}{2N_2}\right)\cdots\left(1 - \frac{1}{2N_{t-1}}\right) \\ &= \prod_{i=0}^{t-1}\left(1 - \frac{1}{2N_i}\right).\end{aligned} \qquad 3.13.2$$

We then ask: What population of constant size would have the same decrease in heterozygosity over the time period involved? We call this number the effective population number.

To find this, we write

$$\frac{H_t}{H_0} = \left(1 - \frac{1}{2N_e}\right)^t = \prod_{i=0}^{t-1}\left(1 - \frac{1}{2N_i}\right), \qquad 3.13.3$$

which can be solved for N_e.

Notice that if the N_i's are fairly large and t is small, this equation is roughly

$$1 - \frac{t}{2N_e} = 1 - \sum \frac{1}{2N_i}$$

or

$$\frac{1}{N_e} = \frac{1}{t} \sum \frac{1}{N_i}. \qquad 3.13.4$$

So, if the population size fluctuates, the effective population number is roughly the harmonic mean of the various values.

In Chapter 2 we showed that with varying population growth rates at different times the arithmetic mean is appropriate as a summarizing value if fitness is measured by the Malthusian parameter, m (1.2.5), and the geometric mean is used if fitness is measured by w (1.1.5). Here we find that still a third average, the harmonic mean, serves best as a single representative value.

Finally, if the individuals in the population do not have the same expected number of progeny, the effective number will be less than the census number. If the population size is constant

$$N_e = \frac{4N - 2}{\sigma_k^2 + 2} \qquad 3.13.5$$

where σ_k^2 is the variance in the number of progeny per parent. This is derived, along with more general formulae, later in Section 7.6.

This formula assumes complete random mating, including the possibility of self-fertilization. Notice that when the number of progeny per parent has a binomial distribution with mean 2 (the mean must be 2 in a sexual population that is neither increasing nor decreasing) this is equivalent to drawing a sample of $2N$ gametes randomly from the N parents. The binomial variance is

$$\sigma_k^2 = 2N\left(\frac{1}{N}\right)\left(1 - \frac{1}{N}\right) = \frac{2(N-1)}{N}.$$

Substituting this into 3.13.5 gives

$$N_e = N,$$

as it should. In this "ideal" population the actual and effective numbers are the same.

Notice that the ideal population model when $N_e = N$ does not imply a constant number of progeny per parent, but a randomly varying number. If the number has greater than binomial variance the effective number is smaller

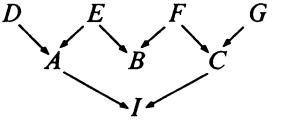
than the census number; if the variance is less than binomial, the effective number is larger. An extreme case is that where each parent is constrained to have the same number of progeny, as in many livestock or laboratory animal breeding systems. If $\sigma_k^2 = 0$ in 3.13.5 then $N_e = 2N - 1$, almost twice the census number. So, with random mating and binomial progeny distribution, about half the reduction in heterozygosity is due to consanguinity among mates; the other half is caused by variable numbers of progeny.

In most populations in nature the effective number is less than the census number. For a discussion of laboratory and census data on this question, see Crow and Morton (1955).

The concept of effective population number and the formulae in this section are all due to Sewall Wright (see especially his 1931 and 1938*b* papers). For extensions to more complicated situations see Crow (1954), Crow and Morton (1955), and Kimura and Crow (1963*a*). We shall return to this subject in Section 7.6.

3.14 Problems

1. In Table 3.1.1, if the initial frequencies are in Hardy–Weinberg ratios, show that in generation t, $D_t = p(1 - 2^{-t}q)$, where p and q are the frequencies of the dominant and recessive genes.
2. In deriving 3.8.19 we rather arbitrarily subtracted $2(1 - S)H_0/(2 - S)$ from both sides of the equation. Show, by equating H_t to H_{t-1}, that this is an equilibrium value (and that therefore the procedure was not arbitrary).
3. Is individual I inbred?



4. The algorithm for computing the inbreeding (or consanguinity) coefficient (3.4.1) permits no individual to be counted twice in the same path. Invent an algorithm that can dispense with the $(1 + f_A)$ term by permitting the path to pass through some individuals twice. (Figure 3.4.3 provides a good example.)
5. What is the rate of decrease in heterozygosity from mother to daughter in mother-son mating in honeybees? Remember that the male bee is haploid, being derived parthenogenetically from the mother. Thus the rule is the same as for the X chromosome in ordinary species. See Figure 3.14.1.

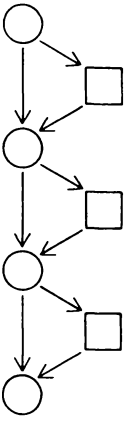


Figure 3.14.1. Repeated mother-son mating in honeybees.

6. Compare the results of parent-offspring mating for an autosomal locus with that for a sex-linked locus (or for bees) in problem 5. See Figure 3.14.2.

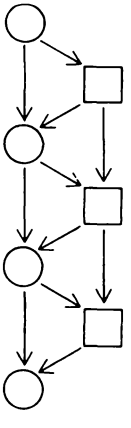


Figure 3.14.2. Repeated parent-offspring mating for autosomal locus.

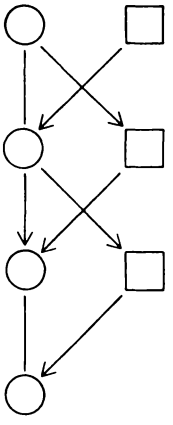


Figure 3.14.3. Repeated sib (or half-sib), mating in honeybees.

7. Show that continued sib mating leads to the same rate of decrease in heterozygosity in honeybees (or for an X-linked locus) as in ordinary diploid inheritance. See Figure 3.14.3.
8. Artificial insemination in bees requires pooled sperm from several males. If these males are brothers, will this alter the results of problem 7?
9. Show that, with partial self-fertilization, when self-fertilized plants produce fewer progeny than cross-fertilized the gene frequency does not change (although, of course, the rate of increase in homozygosity is less).
10. Is the rate of decrease in heterozygosity by random drift the same for a hermaphroditic species, such as earthworms, where individuals produce both eggs and sperm but do not self-fertilize, as for a bisexual species with equal numbers of males and females?
11. Show that with continued sib mating the heterozygosity, H_t, is given by $H_t = H_{t-1} - (1/8)H_{t-3}$.
12. How would you expect inbreeding to affect the rate of approach to linkage equilibrium?
13. Generalize the formula for maximum avoidance of inbreeding to a population of size 8. What is the value of λ?
14. Figure 3.14.4 gives an example of a mating system that reaches an equilibrium level of heterozygosity rather than approaching 0. Show that the eventual heterozygosity alternates between the original amount and 2/3 this amount. (You might enjoy generalizing this to two generations of sib mating between outcrosses. The inbreeding coefficient then cycles through the values 0, 5/14, and 3/7 at equilibrium.)

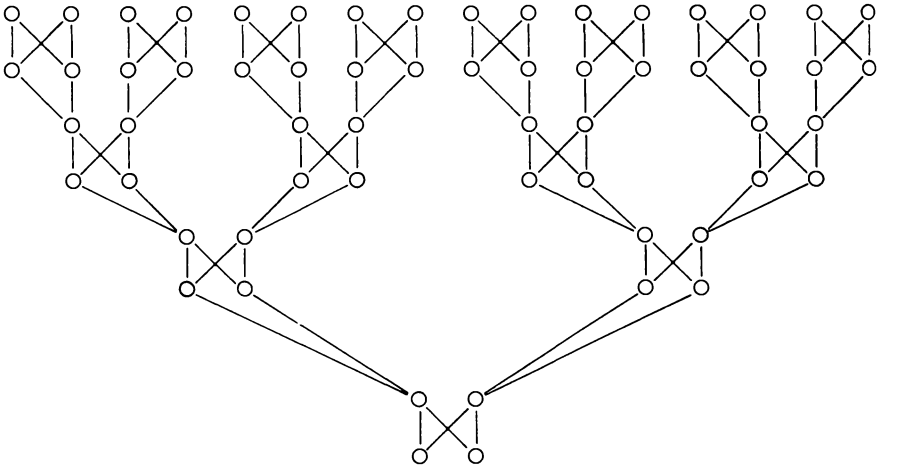


Figure 3.14.4. A mating system in which sib-mating is alternated with outcrossing.

15. Show that, starting with a randomly mating monoecious population of size N, the inbreeding coefficient after t generations is $1 - [(2N - 1)/2N]^t$.
16. What is the mean number of backcross generations required to separate a gene from the selected locus if the recombination frequency is 0.1?
17. If 10 generations of backcrossing are carried out, what is the probability that the genes are not yet separated?
18. How many generations of backcrossing must an experimenter carry out if he wants a probability of at least 0.95 that the genes have separated?
19. Show that the variance of the number of backcross generations that a gene will remain linked (see 3.8.21 for the mean number) is $(1 - r)/r^2$. *Hint*: Note that when $Z = 1 + y + y^2 + y^3 + \cdots$,

$$\frac{d^2Z}{dy^2} - \frac{dZ}{dy} = 1 + 4y + 9y^2 + 16y^3 + \cdots.$$

20. Show that with maximum avoidance of inbreeding in a population of size 4,

$$h_t = \tfrac{1}{2}h_{t-1} + \tfrac{1}{4}h_{t-2} + \tfrac{1}{8}h_{t-3}.$$

21. Comparing this to the formula for sib mating, which is maximum avoidance in a population of size 2 (3.8.5), and that for self-fertilization (size 1), you can probably guess what the generalization is to populations of size 8, 16, etc. Show this for a population of size 8. (See Problem 13)
22. Show that the variance of x in 3.8.20 is $1/t^2$.