

Descent Measures for Two Loci with Some Applications*

C. CLARK COCKERHAM

*Department of Statistics, North Carolina State University,
Raleigh, North Carolina 27607*

AND

B. S. WEIR

Department of Mathematics, Massey University, Palmerston North, New Zealand

Received July 25, 1972

For any four genes, two at each of two loci, in a population, a 15 component descent measure has been introduced. These components are the probabilities of the 15 possible arrangements on a set of initial gametes of those genes of which the four of interest are copies. Since identity by descent of genes is equivalent to their being copies of a single gene on an initial gamete, descent measures have inbreeding coefficients as special cases. The individual descent measure, defined for four genes on two uniting gametes can be evaluated for any pedigree by means of an algorithm developed here. If initial gametic frequencies are specified, descent measures allow genotypic frequencies and disequilibria functions at one and two loci to be found. The procedures are illustrated for selfing and for sib mating. Several applications of the descent measures are discussed.

We are concerned with establishing a set of parameters which will simplify the development of two-locus genetic models. Any such set should enable us to describe the joint effects of linkage and inbreeding, provide genotypic frequencies at two loci, characterize various two-locus disequilibria functions, and be amenable to straightforward evaluation. Such requirements are met by the set of descent measures we define and discuss in this paper. We illustrate their evaluation and use by applying them to selfing and to sib mating.

* Paper No. 3830 of the Journal Series of the North Carolina State University Agricultural Experiment Station, Raleigh, North Carolina. This investigation was supported in part by Public Health Service Grant GM 11546 from the Division of General Medical Sciences.

TWO-LOCUS DESCENT MEASURES

The general two-locus descent measure $\mathbf{X}(ab, a'b')$ is defined for two genes a, a' at one locus and two genes b, b' at a second locus. No restriction is implied on the total number of allelic states in a population at these two loci. The measure gives the probabilities of the various arrangements of a, a', b, b' on gametes in the initial ancestors or an initial population.

There are 15 ways in which the four genes could have been distributed on initial gametes, and the probabilities of these 15 events form the components of the vector \mathbf{X} . For any two genes, not necessarily at the same locus, we set

$$\begin{aligned} \delta(xy) &= 1 && \text{if } x \text{ and } y \text{ are copies of genes on one initial gamete,} \\ &= 0 && \text{otherwise,} \end{aligned}$$

so that the components of $\mathbf{X}(ab, a'b')$ may be written as ${}_{mn}X_{ij}^{kl}$ where

$$\begin{aligned} i &= \delta(aa'), & k &= \delta(ab), & m &= \delta(ab'), \\ j &= \delta(bb'), & l &= \delta(a'b'), & n &= \delta(a'b). \end{aligned}$$

In other words we consider the states of descent of all six pairings of the four genes. Although six two-valued suffices suggest 64 components, in fact arrangements such as ${}_{00}X_{11}^{11}$ are impossible and the 15 components are

$$\begin{aligned} \mathbf{X}' = & ({}_{11}X_{11}^{11}, {}_{00}X_{11}^{00}, {}_{00}X_{00}^{11}, {}_{11}X_{00}^{00}, {}_{01}X_{10}^{10}, {}_{10}X_{10}^{01}, {}_{10}X_{01}^{10}, {}_{01}X_{01}^{01}, \\ & {}_{00}X_{10}^{00}, {}_{00}X_{01}^{00}, {}_{00}X_{00}^{10}, {}_{00}X_{00}^{01}, {}_{10}X_{00}^{00}, {}_{01}X_{00}^{00}, {}_{00}X_{00}^{00}). \end{aligned}$$

These sum to one since the four genes must fall into one of these states.

The quartets of genes for which a descent measure is defined are identified principally by the gametes on which they are located, while gametes are specified either as coming from specific individuals or uniting to form a specific individual or a specific generation. Since four genes can be carried on two, three, or four gametes, we are led to define digametic, trigametic, and quadrigametic classes of measures, accordingly:

$$\begin{aligned} \mathbf{F}_A &= \mathbf{X}(ab, a'b' : ab, a'b' \text{ are located on gametes which unite to} \\ & \quad \text{form individual } A), \\ \theta_{BC} &= \mathbf{X}(ab, a'b' : ab, a'b' \text{ are located on two gametes from} \\ & \quad \text{individuals } B \text{ and } C, \text{ respectively}), \\ \gamma_{B,CD} &= \mathbf{X}(ab, a'b' : ab, a', b' \text{ are located on three gametes from} \\ & \quad \text{individuals } B, C \text{ and } D, \text{ respectively}), \\ \delta_{BC,DE} &= \mathbf{X}(ab, a'b' : a, b, a', b' \text{ are located on four gametes from} \\ & \quad \text{individuals } B, C, D, \text{ and } E, \text{ respectively}). \end{aligned}$$

These definitions must be adhered to rigidly. It is not true, for example, that $\delta_{BC,DE} = \delta_{DE,BC}$.

SUMMARY MEASURES

It is useful to summarize these measures in various ways and in doing so we point out several equivalences and the incorporation of inbreeding measures as well as develop summary notation. The above notation is deliberately similar to that used for two-locus inbreeding measures (Weir and Cockerham, 1969a)

TABLE I
Two-Gene-Pair and Gene-Pair Marginal Totals

$\delta(bb') = 1$		$\delta(bb') = 0$		
$\delta(aa') = 1$	$F_{11} = {}_{11}F_{11}^{11} + {}_{00}F_{11}^{00}$	$F_{10} = {}_{01}F_{10}^{10} + {}_{10}F_{10}^{01} + {}_{00}F_{10}^{00}$		$F_{1.}$
$\delta(aa') = 0$	$F_{01} = {}_{01}F_{01}^{10} + {}_{01}F_{01}^{01} + {}_{00}F_{01}^{00}$	$F_{00} = {}_{00}F_{00}^{11} + {}_{11}F_{00}^{00} + {}_{00}F_{00}^{10}$ $+ {}_{00}F_{00}^{01} + {}_{10}F_{00}^{00} + {}_{01}F_{00}^{00} + {}_{00}F_{00}^{00}$		$F_{0.}$
	$F_{.1}$	$F_{.0}$		1

$\delta(a'b') = 1$		$\delta(a'b') = 0$		
$\delta(ab) = 1$	$F^{11} = {}_{11}F_{11}^{11} + {}_{00}F_{00}^{11}$	$F^{10} = {}_{01}F_{10}^{10} + {}_{10}F_{01}^{10} + {}_{00}F_{00}^{10}$		$F^{1.}$
$\delta(ab) = 0$	$F^{01} = {}_{10}F_{10}^{01} + {}_{01}F_{01}^{01} + {}_{00}F_{00}^{01}$	$F^{00} = {}_{00}F_{00}^{00} + {}_{11}F_{00}^{00} + {}_{00}F_{00}^{10}$ $+ {}_{00}F_{00}^{01} + {}_{10}F_{00}^{00} + {}_{01}F_{00}^{00} + {}_{00}F_{00}^{00}$		$F^{0.}$
	$F^{.1}$	$F^{.0}$		1

$\delta(a'b) = 1$		$\delta(a'b) = 0$		
$\delta(ab') = 1$	${}_{11}F = {}_{11}F_{11}^{11} + {}_{11}F_{00}^{00}$	${}_{10}F = {}_{10}F_{10}^{01} + {}_{10}F_{01}^{10} + {}_{10}F_{00}^{00}$		${}_{1.}F$
$\delta(ab') = 0$	${}_{01}F = {}_{01}F_{10}^{10} + {}_{01}F_{01}^{01} + {}_{01}F_{00}^{00}$	${}_{00}F = {}_{00}F_{00}^{00} + {}_{00}F_{00}^{11} + {}_{00}F_{00}^{10}$ $+ {}_{00}F_{00}^{01} + {}_{00}F_{00}^{10} + {}_{00}F_{00}^{01} + {}_{00}F_{00}^{00}$		${}_{0.}F$
	${}_{.1}F$	${}_{.0}F$		1

which gave joint probabilities of identity by descent at two pairs of homologous genes. We recognize that identity by descent is equivalent to genes at the same locus being copies of genes on one initial gamete. The digametic descent measure \mathbf{F} should contain the information implied by the two-locus inbreeding coefficient as well as by the usual (Wright, 1922) one-locus coefficients.

In Table I are given the two-gene-pair marginal measures as well as the gene-pair marginal measures for each of the three sets of two-gene-pairs (aa' , bb'), (ab , $a'b'$), and (ab' , $a'b$). In the upper table the four components of the two-locus inbreeding coefficient are one set of marginal components of \mathbf{F} which we now refer to as the general (two-locus) individual descent measure. The gene-pair marginals $F_{1.}$ and $F_{.1}$ are the one-locus inbreeding coefficients for the A and B loci, and F_{11} is the marginal two-locus inbreeding coefficient. In the next table for the individual for which \mathbf{F} is defined $F^{1.}$ and $F^{.1}$ are the probabilities that parental gametes are copies of initial parental gametes and F^{11} is the joint probability for both parental gametes, while in the lower table ${}_{1.}F$ and ${}_{.1}F$ are the probabilities that recombinant gametes are copies of initial parental gametes and ${}_{11}F$ is the joint probability for both recombinant gametes. It seems appropriate to refer to $F^{1.}$ and $F^{.1}$ as parental descent coefficients and to ${}_{1.}F$ and ${}_{.1}F$ as recombinant descent coefficients. Then F^{11} and ${}_{11}F$ are the two-gamete parental and recombinant descent coefficients, respectively.

The descent arguments involve the same gametic pathways for $F_{1.}$ and $F_{.1}$ as well as for ${}_{1.}F$ and ${}_{.1}F$ and the equal values for each pair are denoted as F_1 and ${}_{1}F$, respectively. This implies also that $F_{10} = F_{01}$ and ${}_{10}F = {}_{01}F$. We use F^1 to refer to the average $(F^{1.} + F^{.1})/2$. While $F^{1.}$ and $F^{.1}$ may differ, for our present or any foreseen purposes the average suffices, and so do the following averages in some of which some of the measures may not be equal:

$$\begin{aligned} {}_{01}\tilde{F}_{10}^{10} &= ({}_{01}F_{10}^{10} + {}_{10}F_{10}^{01} + {}_{10}F_{01}^{10} + {}_{01}F_{01}^{01})/4, & {}_{00}\tilde{F}_{10}^{00} &= ({}_{00}F_{10}^{00} + {}_{00}F_{01}^{00})/2, \\ {}_{00}\tilde{F}_{00}^{10} &= ({}_{00}F_{00}^{10} + {}_{00}F_{00}^{01})/2, & {}_{10}\tilde{F}_{00}^{00} &= ({}_{10}F_{00}^{00} + {}_{01}F_{00}^{00})/2. \end{aligned}$$

We need, then, to be concerned with only a condensed vector of nine measures,

$$\tilde{\mathbf{F}}' = ({}_{11}F_{11}^{11}, {}_{00}F_{11}^{00}, {}_{00}F_{00}^{11}, {}_{11}F_{00}^{00}, 4 {}_{01}\tilde{F}_{10}^{10}, 2 {}_{00}\tilde{F}_{10}^{00}, 2 {}_{00}\tilde{F}_{00}^{10}, 2 {}_{10}\tilde{F}_{00}^{00}, {}_{00}F_{00}^{00}).$$

One additional set of summary measures, three-gene, is useful. These are given in Table II for the three genes ($aa'b$). Similar measures apply to each of the four sets of three genes, and we use the average

$${}_{1}F_1^1 = ({}_{.1}F_{1.}^{1.} + {}_{.1}F_{.1}^{1.} + {}_{1.}F_{1.}^{.1} + {}_{.1}F_{.1}^{.1})/4.$$

It is sometimes convenient to use the following summary measures, actually only eight since the nine condensed measures sum to one,

$$\mathbf{F}^{*'} = (F_{11}^{11}, F_{11}, F^{11}, {}_{11}F, {}_{1}F_1^1, F_1, F^1, {}_{1}F, 1),$$

where $F_{11}^{11} = {}_{11}F_{11}^{11}$. The linear transforms between $\tilde{\mathbf{F}}$ and \mathbf{F}^* are given in Appendix A.

TABLE II
Three-Gene Marginal Totals
($aa'b$)

$$\begin{aligned} .1F_{1.}^1 &= {}_{01}F_{10}^{10} + {}_{11}F_{11}^{11} \\ .0F_{1.}^0 &= {}_{10}F_{10}^{01} + {}_{00}F_{11}^{00} + {}_{00}F_{10}^{00} \\ .0F_{0.}^1 &= {}_{10}F_{01}^{10} + {}_{00}F_{00}^{11} + {}_{00}F_{00}^{10} \\ .1F_{0.}^0 &= {}_{01}F_{01}^{01} + {}_{11}F_{00}^{00} + {}_{01}F_{00}^{00} \\ \frac{{}_{00}F_{00}^0}{1} &= {}_{00}F_{00}^{00} + {}_{00}F_{01}^{00} + {}_{00}F_{00}^{01} + {}_{10}F_{00}^{00} \\ .1F_{1.}^1 + .0F_{1.}^0 &= F_{1.}, \quad .1F_{1.}^1 + .0F_{0.}^1 = F^1 \\ .1F_{1.}^1 + .1F_{0.}^0 &= .1F \end{aligned}$$

INITIAL ANCESTORS OR FOUNDER POPULATION

Two types of initial ancestors or populations will be considered. Each is characterized by gene frequencies p_i for alleles a_i , q_j for alleles b_j and gametic linkage disequilibrium Δ_{ij} between genes a_i and b_j or, alternatively, by gametic frequencies $p_{ij} = p_i q_j + \Delta_{ij}$. In one case the initial ancestors, say N of them, take the frequencies expected from randomly uniting a specific set of $2N$ gametes, which amounts to averaging over all randomizations of pairings of the gametes. For the other case the N initial ancestors are assumed to take the frequencies expected for a random sample from an infinite randomly mating population, which amounts to averaging over all random samples of N initial ancestors. The difference between the two cases of initial ancestors will become clear as we enumerate the initial frequencies, but it can be seen that the results for the second case are produced by letting the $2N$ gametes take the frequencies expected for a random sample from the infinite randomly mating population.

For digametic initial frequencies we use \mathcal{P}_{kl}^{ij} to denote the frequency with which a random pair of gametes is $a_i b_j$ for the first gamete and $a_k b_l$ for the second gamete. Since the argument is symmetrical for gametes, $\mathcal{P}_{kl}^{ij} = \mathcal{P}_{ij}^{kl}$. For trigametic initial frequencies let $\mathcal{P}_{k|l}^{ij}$ denote for three random gametes

the frequency with which the first is $a_i b_j$, a_k is on the second and b_l on the third, and again there is upper-lower symmetry, $\mathcal{P}_{k|l}^{ij} = \mathcal{P}_{ij}^{k|l}$. Finally, for four random gametes we use $\mathcal{P}_{k|l}^{i|j}$ for a specific ordering of a_i , b_j , a_k , b_l each on separate gametes, and there is allelic symmetry $\mathcal{P}_{k|l}^{i|j} = \mathcal{P}_{k|j}^{i|l} = \mathcal{P}_{i|j}^{k|l}$. All of these frequencies are for four genes.

We use the convention that an index replaced by a dot implies summation over that index, e.g.,

$$\mathcal{P}_{k.}^{ij} = \sum_l \mathcal{P}_{kl}^{ij}.$$

There are many equivalences among the various sums of \mathcal{P}_{kl}^{ij} , $\mathcal{P}_{k|l}^{ij}$, and $\mathcal{P}_{k|l}^{i|j}$ because adding over genes for any random gamete reduces by one the number of gametes involved in the frequency. The principal examples are, symmetrical arguments being omitted,

$$\begin{aligned} \mathcal{P}_{..}^{i.} &= \mathcal{P}_{.i.}^{i.} = \mathcal{P}_{.i|.}^{i.} = p_i, & \mathcal{P}_{..}^{ij} &= \mathcal{P}_{.i.}^{ij} = p_{ij} = p_i q_j + \Delta_{ij}, \\ \mathcal{P}_{.j.}^{i.} &= \mathcal{P}_{.i|j}^{i.} = \mathcal{P}_{..}^{i|j} = \mathcal{P}_{.i|j}^{i.} = \mathcal{P}_{.i.}^{i|j}, & \mathcal{P}_{k.}^{ij} &= \mathcal{P}_{k|.}^{ij}. \end{aligned}$$

We illustrate the difference between the two sets of initial ancestors for $\mathcal{P}_{.j.}^{i.}$. Let N_{ij} be the number of $a_i b_j$ gametes, N_i the number of a_i genes and N_j the number of b_j genes. Now $2N$ gametes provide $N(2N - 1)$ pairs of gametes, and for each pair there are two pairings of a with b , each on different gametes, giving a total of $2N(2N - 1)$ pairings. There are $N_i N_j - N_{ij}$ pairs of a_i and b_j genes on different gametes. Thus, the frequency,

$$\mathcal{P}_{.j.}^{i.} = \frac{N_i N_j - N_{ij}}{2N(2N - 1)} = p_i q_j - \frac{\Delta_{ij}}{2N - 1},$$

$$\frac{N_{i.}}{2N} = p_i, \quad \frac{N_{.j}}{2N} = p_j, \quad \frac{N_{ij}}{2N} = \mathcal{P}_{..}^{ij} = p_{ij} = p_i q_j + \Delta_{ij},$$

for an initial set of $2N$ gametes. Just let $N \rightarrow \infty$ for initial ancestors being random members from an infinite randomly mating population, and $\mathcal{P}_{.j.}^{i.} = p_i q_j$. Examples of the remaining frequencies are listed in Table III for a specific set of gametes. The first terms (letting $N \rightarrow \infty$) are those for a random sample of parents.

A comment on disequilibria is in order. A disequilibrium is the deviation of a frequency from that based on random association of genes and accounting for any lower order disequilibria. We shall designate these disequilibria as \mathcal{D} and index them according to the sum to which they apply. By referring to Table III we can quickly identify the two-gene disequilibria. For example,

TABLE III

Initial Di-, Tri-, Quadrigmatic Frequencies for a Set of 2N Gametes

$\mathcal{P}_{k.}^i = p_i p_k + \frac{p_i p_k}{2N - 1}$	$\mathcal{P}_{ij}^k = p_i p_j q_i + \frac{p_i p_j q_i - p_i \Delta_{kj} - p_k \Delta_{ij}}{2N - 1} - \frac{2(p_i \Delta_{kj} + p_k \Delta_{ij})}{(2N - 1)(2N - 2)}$
$\mathcal{P}_{i.}^i = p_i^2 - \frac{p_i(1 - p_i)}{2N - 1}$	$\mathcal{P}_{ij}^i = p_i^2 q_i - \frac{p_i q_i(1 - p_i) + 2p_i \Delta_{ij}}{2N - 1} + \frac{2\Delta_{ij}(1 - 2p_i)}{(2N - 1)(2N - 2)}$
$\mathcal{P}_{.i}^i = p_{ij}$	$\mathcal{P}_{kij}^i = p_{ij} p_k q_i + \frac{p_{ij}(2p_k q_i - \Delta_{ki})}{2N - 1} + \frac{2p_{ij}(p_k q_i - 4_{ki})}{(2N - 1)(2N - 2)}$
$\mathcal{P}_{.j}^i = p_i q_j - \frac{\Delta_{ij}}{2N - 1}$	$\mathcal{P}_{ij}^i = p_{ij} p_i q_i - \frac{p_{ij}[q_i(1 - 2p_i) + \Delta_{ij}]}{2N - 1} - \frac{2p_{ij}[q_i(1 - p_i) + 4_{ij}]}{(2N - 1)(2N - 2)}$
$\mathcal{P}_{k.}^i = p_k p_{ij} + \frac{p_k p_{ij}}{2N - 1}$	$\mathcal{P}_{ij}^i = p_{ij} p_i q_j - \frac{p_{ij}[q_j(1 - p_i) + p_i(1 - q_j) + \Delta_{ij}]}{2N - 1} + \frac{2p_{ij}[(1 - p_i)(1 - q_j) - 4_{ij}]}{(2N - 1)(2N - 2)}$
$\mathcal{P}_{i.}^i = p_i p_{ij} - \frac{p_{ij}(1 - p_i)}{2N - 1}$	$\mathcal{P}_{kij}^i = p_i p_k q_i q_i - a^* + \frac{2p_i p_k q_i q_i - 4a - b^*}{2N - 1} - \frac{3(p_i p_k q_i q_i + 2a + b)}{(2N - 1)(2N - 2)(2N - 3)}$
$\mathcal{P}_{ki}^i = p_{ij} p_{ki} + \frac{p_{ij} p_{ki}}{2N - 1}$	$\mathcal{P}_{ij}^i = p_i^2 q_i q_i - \frac{p_i q_i q_i(1 - 2p_i) + 2p_i c^*}{2N - 1} - \frac{p_i q_i q_i(1 - p_i) - (2 - 8p_i)c - 2\Delta_{ij} \Delta_{ii}}{(2N - 1)(2N - 2)}$
$\mathcal{P}_{.i}^i = p_{ij} p_{ii} + \frac{p_{ij} p_{ii}}{2N - 1}$	$+ \frac{3[p_i q_i q_i(1 - p_i) + (2 - 4p_i)c + 2\Delta_{ij} \Delta_{ii}]}{(2N - 1)(2N - 2)(2N - 3)}$
$\mathcal{P}_{ij}^i = p_{ij}^2 - \frac{p_{ij}(1 - p_{ij})}{2N - 1}$	$\mathcal{P}_{ij}^i = p_i^2 q_i^2 - \frac{p_i q_i(p_i + q_i - 2p_i q_i + 4\Delta_{ij})}{2N - 1} + \frac{p_i q_i(1 - p_i) + 2\Delta_{ij}(2p_i + 2q_i - 8p_i q_i + 4_{ij})}{(2N - 1)(2N - 2)}$
	$- \frac{3[p_i q_i(1 - p_i)(1 - q_i) + 2\Delta_{ij}((1 - 2p_i)(1 - 2q_i) - 4_{ij})]}{(2N - 1)(2N - 2)(2N - 3)}$

* $a = p_i q_i \Delta_{kj} + p_k q_i \Delta_{ij} + p_i q_j \Delta_{ki} + p_k q_j \Delta_{ii}$; $b = \Delta_{ij} \Delta_{ki} + \Delta_{ii} \Delta_{kj}$; $c = q_i \Delta_{ii} + q_i \Delta_{ij}$.

$\mathcal{D}_{k.}^{i.} = p_i p_k + \mathcal{D}_{k.}^{i.}$ and $\mathcal{D}_{k.}^{i.} = p_i p_k / (2N - 1)$. Some other two-gene disequilibria are

$$\mathcal{D}_{i.}^{i.} = -\frac{p_i(1-p_i)}{2N-1}, \quad \mathcal{D}_{..}^{ij} = \Delta_{ij}, \quad \mathcal{D}_{.j}^{i.} = -\frac{\Delta_{ij}}{2N-1}.$$

To find three-gene disequilibria we account for the two-gene disequilibria

$$\mathcal{D}_{k.}^{ij} = p_i p_k q_j + p_i \mathcal{D}_{k.}^{j.} + q_j \mathcal{D}_{k.}^{i.} + p_k \mathcal{D}_{..}^{ij} + \mathcal{D}_{k.}^{ij},$$

and

$$\mathcal{D}_{k.}^{ij} = \frac{p_i \Delta_{jk} + p_k \Delta_{ij}}{2N-1} = \mathcal{D}_{i.}^{kj}$$

or

$$\mathcal{D}_{i.}^{ij} = -\frac{\Delta_{ij}(1-2p_i)}{2N-1}.$$

Four-gene disequilibria are again the remainders after accounting for the lower order ones and will be considered in more detail later.

We note that with an initial set of gametes all disequilibria obtain, some due to finiteness, some due to gametic linkage disequilibria Δ 's, and some due to both. When initial members are a random sample the only disequilibrium is gametic

$$\mathcal{D}_{..}^{ij} = \Delta_{ij};$$

all other two gene and higher order ones being zero.

GENOTYPIC FREQUENCIES

The expression of genotypic and other frequencies is a function of the descent measures and the conditions assumed for the initial ancestors. The descent measures characterize the descension of the genes, at most two at each of two loci, for an individual or for the population of which it is a random member, taking into account size of the population, mating pattern, and so on. Excluded are disturbing forces such as mutation and selection.

For frequencies or disequilibria in subsequent generations corresponding to those for the ancestors we shall replace \mathcal{P} by P and \mathcal{D} by D . Corresponding equivalences and properties of symmetry remain for P and D .

For an individual under question characterized by the descent measures \mathbf{F} we write the genotypic frequency as P_{kl}^{ij} for the zygote formed by the union of male gamete $a_i b_j$ with female gamete $a_k b_l$. For a_i , a_k and b_j , b_l not alike in

state (as implied by the distinct subscripts) we know that four distinct initial genes are involved and these may have been arranged on two, three, or four gametes. With probability ${}_{00}F_{00}^{11}$ they were arranged $a_i b_j$ and $a_k b_l$ so that the genotypic frequency includes the term $\mathcal{P}_{kl}^{ij} {}_{00}F_{00}^{11}$. With probability ${}_{11}F_{00}^{00}$ they were arranged $a_i b_l$ and $a_k b_j$ and, consequently, the term $\mathcal{P}_{kj}^{il} {}_{11}F_{00}^{00}$. For the summary component ${}_{00}\tilde{F}_{00}^{10} = {}_{00}F_{00}^{10} + {}_{00}F_{00}^{01}$ the arrangement was $a_i b_j$ and a_k and b_l on separate gametes with probability ${}_{00}F_{00}^{10}$ and $a_k b_l$ and a_i and b_j on separate gametes with probability ${}_{00}F_{00}^{01}$ leading to the term ${}_{00}\tilde{F}_{00}^{10} \frac{1}{2} (\mathcal{P}_{kl}^{ij} + \mathcal{P}_{ij}^{kl})$. By such arguments we find the frequencies for double homozygotes and single heterozygotes, as well as double heterozygotes, and display them in Table IV. To find trigametic frequencies $P_{k|l}^{ij}$, one simply replaces \mathbf{F} by γ , the appropriate trigametic descent measure, in Table IV. Correspondingly, for quadri-gametic frequencies one uses δ , the quadrigametic descent measures.

DISEQUILIBRIUM FUNCTIONS

In order to arrive at disequilibrium functions we first need sums of the genotypic frequencies, and we use the same convention as for \mathcal{P} . Any of these sums may be found by appropriate summations of the terms in Table IV. It is more enlightening, however, to develop some of them directly.

Summation over all but one index provides the gene frequencies as before,

$$P_{..}^i = p_i, \quad P_{..}^j = q_j.$$

There are three types of gene-pair sums $P_{k.}^i$, $P_{..}^{ij}$ and $P_{.i}^i$ each with its counterpart. For $P_{k.}^i$ the alleles are distinct and, thus, nonidentical by descent with probability $(1 - F_1)$ and

$$P_{k.}^i = (1 - F_1) \mathcal{P}_{k.}^i = (1 - F_1) p_i p_k \frac{2N}{2N - 1},$$

while for $P_{.i}^i$ the genes are identical by descent with probability F_1 and a_i with frequency p_i and nonidentical with probability $(1 - F_1)$ and $a_i a_i$ with frequency $\mathcal{P}_{.i}^i$,

$$P_{.i}^i = F_1 p_i + (1 - F_1) \mathcal{P}_{.i}^i = p_i^2 + p_i(1 - p_i) \left(F_1 \frac{2N}{2N - 1} - \frac{1}{2N - 1} \right).$$

With respect to $P_{..}^{ij}$ the two genes came from an original gamete with probability F^1 and two distinct gametes with probability $(1 - F^1)$,

$$P_{..}^{ij} = F^1 p_{ij} + (1 - F^1) \mathcal{P}_{.j}^i = p_i q_j + \Delta_{ij} \left(F^1 \frac{2N}{2N - 1} - \frac{1}{2N - 1} \right).$$

TABLE IV
Genotypic Frequencies as Functions of the General Descent Measures

	F_{II}^{11}	${}_{00}F_{II}^{00}$	${}_{00}F_{ij}^{11}$	${}_{11}F_{00}^{00}$	$4{}_{01}F_{10}^{10}$	$2{}_{00}F_{ij}^{00} + \mathcal{P}_{ij}^j$	$2{}_{00}F_{ij}^{10}$	$2{}_{10}F_{00}^{00}$	F_{00}^{00}
P_{ij}^{ij}	\mathcal{P}_{ij}^{ij}	\mathcal{P}_{ij}^i	\mathcal{P}_{ij}^{ij}	\mathcal{P}_{ij}^{ij}	$\frac{1}{2}(\mathcal{P}_{ij}^{ij} + \mathcal{P}_{ij}^j)$	$\frac{1}{2}(\mathcal{P}_{ij}^i + \mathcal{P}_{ij}^j)$	\mathcal{P}_{ij}^{ij}	\mathcal{P}_{ij}^{ij}	\mathcal{P}_{ij}^{ij}
P_{ii}^{ij}	0	0	\mathcal{P}_{ii}^{ij}	\mathcal{P}_{ii}^{ij}	$\frac{1}{4}(\mathcal{P}_{ii}^{ij} + \mathcal{P}_{ij}^j)$	$\frac{1}{2}\mathcal{P}_{ij}^i$	$\frac{1}{2}(\mathcal{P}_{ij}^{ij} + \mathcal{P}_{ij}^{ii})$	$\frac{1}{2}(\mathcal{P}_{ij}^{ij} + \mathcal{P}_{ij}^{ii})$	\mathcal{P}_{ij}^{ij}
P_{kj}^{ij}	0	0	\mathcal{P}_{kj}^{ij}	\mathcal{P}_{kj}^{ij}	$\frac{1}{4}(\mathcal{P}_{kj}^{ij} + \mathcal{P}_{ij}^i)$	$\frac{1}{2}\mathcal{P}_{kj}^i$	$\frac{1}{2}(\mathcal{P}_{kj}^{ij} + \mathcal{P}_{ij}^{kj})$	$\frac{1}{2}(\mathcal{P}_{kj}^{ij} + \mathcal{P}_{ij}^{kj})$	\mathcal{P}_{kj}^{ij}
P_{ki}^{ij}	0	0	\mathcal{P}_{ki}^{ij}	\mathcal{P}_{ki}^{ij}	0	0	$\frac{1}{2}(\mathcal{P}_{ki}^{ij} + \mathcal{P}_{ij}^{ki})$	$\frac{1}{2}(\mathcal{P}_{ki}^{ij} + \mathcal{P}_{ij}^{ki})$	\mathcal{P}_{ki}^{ij}

Using the same type of argument for $P_{.j}^i$ but replacing F^1 with ${}_1F$ leads to

$$P_{.j}^i = p_i q_j + \Delta_{ij} \left({}_1F \frac{2N}{2N-1} - \frac{1}{2N-1} \right).$$

The difference

$$P_{..}^{ij} - P_{.j}^i = (F^1 - {}_1F) \Delta_{ij} \frac{2N}{2N-1}$$

is provided entirely by the parental and recombinant descent coefficients and the initial gametic disequilibrium. All finite corrections disappear since all \mathcal{D} are zero except $\mathcal{D}_{..}^{ij} = \Delta_{ij}$ when the initial ancestors are members of an infinite random mating population.

Before proceeding to three-gene sums we shall consider two-gene disequilibria. For alleles these are

$$D_{k.}^i = -F_1 p_i p_k + (1 - F_1) \mathcal{D}_{k.}^i, \quad D_{i.}^k = F_1 p_i (1 - p_i) + (1 - F_1) \mathcal{D}_{i.}^k,$$

which are the usual results for inbreeding. One can obtain any D for homozygotes from that for heterozygotes in a simple manner. As an example consider the substitution of i for k in $D_{k.}^i$. Substitute $-(1 - p_i)$ for p_k , and substitute the index i for k .

The other two types of two-gene disequilibria are

$$D_{..}^{ij} = \Delta_{ij} \left(F^1 \frac{2N}{2N-1} - \frac{1}{2N-1} \right), \quad D_{.j}^i = \Delta_{ij} \left({}_1F \frac{2N}{2N-1} - \frac{1}{2N-1} \right).$$

For the remaining sums and disequilibria we shall assume initial ancestors to be random and avoid the cumbersome finite corrections for a specific set of gametes. Direct arguments for three-gene sums require the three-gene descent measures,

$$\begin{aligned} P_{k.}^{ij} &= {}_0F_0^1 p_{ij} p_k + {}_1F_0^0 p_{kj} p_i + {}_0F_0^0 p_i p_k q_j \\ &= (1 - F_1) p_i p_k q_j + (F^1 - {}_1F_1^1) p_k \Delta_{ij} + ({}_1F - {}_1F_1^1) p_i \Delta_{kj}, \\ P_{i.}^{ij} &= {}_1F_1^1 p_{ij} + ({}_0F_0^1 + {}_1F_0^0) p_i p_{ij} + {}_0F_1^0 p_i q_j + {}_0F_0^0 p_i^2 q_j \\ &= F_1 p_i q_j + (1 - F_1) p_i^2 q_j + ({}_1F + F^1 - 2 {}_1F_1^1) p_i \Delta_{ij} + {}_1F_1^1 \Delta_{ij}. \end{aligned}$$

The three-gene disequilibrium, after accounting for two-gene disequilibria, is found to be

$$D_{k.}^{ij} = -{}_1F_1^1 (p_i \Delta_{kj} + p_k \Delta_{ij}).$$

By performing our substitution rule of $-(1 - p_i)$ for p_k ,

$$D_{i.}^{jj} = {}_1F_1^1(1 - 2p_i) \Delta_{ij}.$$

Switching k and i in $D_{k.}^{ij}$ leaves it unchanged, and we have three-gene symmetry $D_{k.}^{ij} = D_{i.}^{kj}$. This does not hold for the P 's, however,

$$P_{k.}^{ij} - P_{i.}^{kj} = (F^1 - {}_1F)(p_k \Delta_{ij} - p_i \Delta_{kj}),$$

the difference being due to two-gene disequilibria which are taken into account in finding the three-gene D 's.

The four-gene frequencies are just those given in Table IV. To find four-gene disequilibria we account for all lower order ones including two-gene-pair ones (e.g., $D_{k.}^i, D_{.i}^j$)

$$\begin{aligned} P_{kl}^{ij} = & p_i p_k q_j q_l + D_{k.}^i D_{.l}^j + D_{.i}^{jj} D_{kl} + D_{.i}^i D_{k.}^j \\ & + p_i D_{kl}^j + p_k D_{.l}^{ij} + q_j D_{ki}^i + q_l D_{k.}^{ij} + q_j q_l D_{k.}^i \\ & + p_i p_k D_{.l}^j + p_i q_j D_{ki}^i + p_k q_l D_{.i}^{ij} + p_k q_j D_{.l}^i + p_i q_l D_{k.}^j + D_{kl}^{ij}. \end{aligned}$$

After some tediousness,

$$\begin{aligned} D_{kl}^{ij} = & \eta_{11} p_i p_k q_j q_l + \eta^{11} \Delta_{ij} \Delta_{kl} + {}_{11}\eta \Delta_{il} \Delta_{kj} \\ & + F_{11}^{11} [p_i q_j \Delta_{kl} + p_k q_l \Delta_{ij} + p_i q_l \Delta_{kj} + p_k q_j \Delta_{il} - \Delta_{ij} \Delta_{kl} - \Delta_{kj} \Delta_{il}], \end{aligned}$$

where

$$\eta_{11} = F_{11} - (F_1)^2, \quad \eta^{11} = F^{11} - (F^1)^2, \quad {}_{11}\eta = {}_{11}F - ({}_1F)^2.$$

The term η_{11} was called identity disequilibrium by Weir and Cockerham (1969a). The corresponding measures η^{11} and ${}_{11}\eta$ could be called parental and recombinant disequilibria, respectively, all three being descent disequilibrium functions for two-gene-pair descent measures. By performing our substitution rule,

$$\begin{aligned} D_{il}^{ij} = & -\eta_{11} p_i (1 - p_i) q_j q_l + (\eta^{11} + {}_{11}\eta) \Delta_{ij} \Delta_{il} \\ & + F_{11}^{11} [(2p_i - 1)(q_j \Delta_{il} + q_l \Delta_{ij}) - 2\Delta_{ij} \Delta_{il}], \\ D_{kj}^{ij} = & -\eta_{11} p_i p_k q_j (1 - q_j) + (\eta^{11} + {}_{11}\eta) \Delta_{ij} \Delta_{kj} \\ & + F_{11}^{11} [(2q_j - 1)(p_i \Delta_{kj} + p_k \Delta_{ij}) - 2\Delta_{ij} \Delta_{kj}], \\ D_{ij}^{ij} = & \eta_{11} p_i (1 - p_i) q_j (1 - q_j) + (\eta^{11} + {}_{11}\eta) \Delta_{ij}^2 \\ & + F_{11}^{11} [(1 - 2p_i)(1 - 2q_j) - 2\Delta_{ij}] \Delta_{ij}. \end{aligned}$$

To summarize these results we note first that all two-gene disequilibria are functions of the corresponding two-gene marginal descent measures and three-gene disequilibria of the three-gene marginal descent measure, while four-gene disequilibria involve in addition to the four-gene descent measure F_{11}^{11} the three two-gene-pair descent disequilibrium functions. If initial ancestors are random and in linkage equilibrium, all $\Delta_{ij} = 0$, then there is only two-gene allelic disequilibrium due to F_1 and four-gene disequilibrium due to the identity disequilibrium, η_{11} . If we include as linkage disequilibrium all those parts other than that due to identity by descent of alleles, then it is dependent on and strongly influenced by initial disequilibrium as well as the other descent measures to be explored.

A few other things should be noted. All D 's sum to zero over any index, e.g.,

$$\sum_j D_{..}^{ij} = 0, \quad \sum_k D_{k.}^{ij} = 0, \quad \sum_l D_{kl}^{ij} = 0.$$

When there are only two alleles, a_i and a_k , for example,

$$p_i = 1 - p_k, \quad \Delta_{ij} = -\Delta_{kj}, \quad D_{..}^{ij} = -D_{..}^{kj}.$$

If further, $q_j = 1 - q_l$ then $\Delta_{ij} = \Delta_{kl} = -\Delta_{il} = -\Delta_{kj}$ with corresponding equalities among the nonallelic two-gene disequilibria.

While we have been able to express all deviations from random associations of genes in terms of D 's and relate these to descent measures and initial conditions there are other comparisons often of interest. One is the difference between double heterozygotes

$$\begin{aligned} 2(P_{kl}^{ij} - P_{il}^{kj}) &= 2(F^{11} - {}_{11}F)(\Delta_{ij}\Delta_{kl} - \Delta_{il}\Delta_{kj}) \\ &+ 2(F^1 - {}_1F)(p_i q_j \Delta_{kl} + p_k q_l \Delta_{ij} - p_i q_l \Delta_{kj} - p_k q_j \Delta_{il}). \end{aligned}$$

The term $(\Delta_{ij}\Delta_{kl} - \Delta_{il}\Delta_{kj})$ is zero if either locus has only two alleles. The difference $2\sum_{k \neq i, l \neq j} (P_{kl}^{ij} - P_{kj}^{il})$ is actually equivalent to $2(P_{..}^{ij} - P_{.j}^{i.}) = 2(F^1 - {}_1F)\Delta_{ij}$ and is the result for two alleles at each of the loci.

Of interest also are the deviations of genotypic frequencies from products of one-locus frequencies or of gametic frequencies or of recombinant frequencies. These can be found in terms of gene frequencies and D 's and translated into F 's and Δ 's by removing those terms for the marginal products, e.g., $P_{k.}^i P_{.l}^j = (p_i p_k + D_{k.}^i)(q_j q_l + D_{.l}^j)$. These will be studied in more detail in terms of examples.

For tri- and quadrigametic disequilibrium functions simply replace F by γ and δ , respectively.

ALGORITHMS FOR FINDING THE DESCENT MEASURES

Some familiarity and understanding of the descent measures is accomplished from algorithms for their computation from pedigrees. The procedure for finding the inbreeding coefficient F_1 is well known. The procedures for F_{11} , F_{10} , F_{01} , F_{00} were developed in detail by Weir and Cockerham (1969a). To this list we wish to add the other descent measures. We work with the summary components.

The arguments for the parental descent coefficient F^1 and the recombinant descent coefficient ${}_1F$ are interrelated, and they must be considered together. Any F measure for an individual is the same as the θ measure, called coancestry, between random gametes from each of the parents by definition. Then, referring to Fig. 1,

$$F_A = \theta_{BC}.$$

For F_A^1 the argument is the average of that for each of the gametes, one from B and one from C . The gamete will be parental with probability $(1 + \lambda)/2$, where λ is the linkage parameter, in which case the probability of having descended

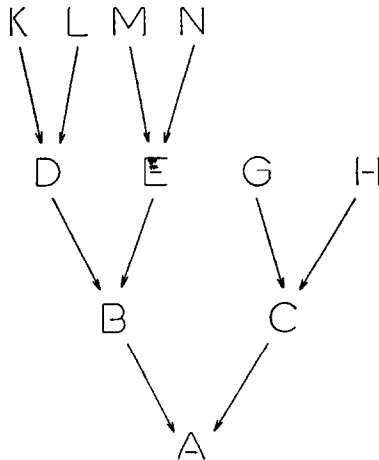


FIGURE 1

from an initial gamete is just F_B^1 or F_C^1 . The gamete will be recombinant with probability $(1 - \lambda)/2$ and the appropriate measure is ${}_1F_B$ or ${}_1F_C$. Consequently,

$$F_A^1 = \theta_{BC}^1 = \frac{1 + \lambda}{2} \frac{F_B^1 + F_C^1}{2} + \frac{1 - \lambda}{2} \frac{{}_1F_B + {}_1F_C}{2}. \quad (1)$$

In the case of two random gametes from the same individual

$$\theta_{BB}^1 = [(1 + \lambda)/2]F_B^1 + [(1 - \lambda)/2]{}_1F_B.$$

For the recombinant descent coefficient the argument is always for an a gene on one gamete with a b gene on another gamete and, thus, involves genes from different ancestors whether recombination has occurred or not;

$${}_1F_A = {}_1\theta_{BC} = \frac{1}{2}({}_1\theta_{BG} + {}_1\theta_{BH}) = \frac{1}{4}({}_1\theta_{DG} + {}_1\theta_{DH} + {}_1\theta_{EG} + {}_1\theta_{EH}) \quad (2)$$

and expands back in the pedigree in the same manner until a common ancestor becomes involved. For two gametes from the same individual

$${}_1\theta_{BB} = \frac{F_B^1 + {}_1F_B}{2} \quad (3)$$

the argument being for two genes on the same gamete half the time and on different gametes half the time. For expansions of F^0 and ${}_0F$ simply substitute the index.

For three-gene probabilities we note that three genes, two at one locus and one at the other locus, may be carried on two or three gametes. We illustrate the expansion of digametic three-gene measures with ${}_1F_{1A}^1 = {}_1\theta_{1BC}^1$. The arguments are symmetrical for any three genes and any two individuals, that is for an a and b gene from one individual and an a or b gene from the other with equal probability.

To keep records straight we let

$${}_1F_{1A}^1 = {}_1\theta_{1BC}^1 = \frac{1}{2}({}_1\theta_{1B,C}^1 + {}_1\theta_{1C,B}^1),$$

where ${}_1\theta_{1B,C}^1$ is for a random gamete from B and a random gene (a or b) from C , and vice versa for ${}_1\theta_{1C,B}^1$. Bringing in the parents of the second individual simply averages the function, that is

$${}_1\theta_{1B,C}^1 = {}_1\theta_{1B,(GH)}^1 = \frac{1}{2}({}_1\theta_{1B,G}^1 + {}_1\theta_{1B,H}^1).$$

The gamete from individual B will be parental and from D or E each with probability $(1 + \lambda)/4$ and recombinant with probability $(1 - \lambda)/2$.

$$\begin{aligned} {}_1\theta_{1B,C}^1 &= {}_1\theta_{1(DE),C}^1 \\ &= [(1 + \lambda)/4]({}_1\theta_{1D,C}^1 + {}_1\theta_{1E,C}^1) + [(1 - \lambda)/4]({}_1\gamma_{1DC,E}^1 + {}_1\gamma_{1EC,D}^1) \\ &= [(1 + \lambda)/8]({}_1\theta_{1D,G}^1 + {}_1\theta_{1D,H}^1 + {}_1\theta_{1E,G}^1 + {}_1\theta_{1E,H}^1) \\ &\quad + [(1 - \lambda)/8]({}_1\gamma_{1DG,E}^1 + {}_1\gamma_{1DH,E}^1 + {}_1\gamma_{1EG,D}^1 + {}_1\gamma_{1EH,D}^1). \end{aligned}$$

The notation for the trigametic three-gene measure ${}_1\gamma_{1DC,E}^1$ implies two genes

at one locus from D and C and a gene at the other locus from E . Averaging the expansions for ${}_1\theta_{1B,C}^1$ and ${}_1\theta_{1C,B}^1$ produces

$$\begin{aligned} {}_1\theta_{1BC}^1 &= [(1 + \lambda)/8]({}_1\theta_{1DG}^1 + {}_1\theta_{1DH}^1 + {}_1\theta_{1EG}^1 + {}_1\theta_{1EH}^1) \\ &\quad + [(1 - \lambda)/16]({}_1\gamma_{1DG,E}^1 + {}_1\gamma_{1DH,E}^1 + {}_1\gamma_{1EG,D}^1 + {}_1\gamma_{1EH,D}^1 \\ &\quad + {}_1\gamma_{1GD,H}^1 + {}_1\gamma_{1GE,H}^1 + {}_1\gamma_{1HD,G}^1 + {}_1\gamma_{1HE,G}^1). \end{aligned} \quad (4)$$

Such expansions are continued until common ancestry is encountered,

$${}_1\theta_{1BB}^1 = [(1 + \lambda)/4]F_B^1 + \frac{1}{2}{}_1F_{1B}^1 + [(1 - \lambda)/4]{}_1F_B^1. \quad (5)$$

The argument is simpler for the trigametic three-gene measure ${}_1\gamma_{1DE,G}^1$. As above, bringing in the parents of each individual just averages the function

$$\begin{aligned} {}_1\gamma_{1DE,G}^1 &= {}_1\gamma_{1(KL)E,G}^1 = \frac{1}{2}({}_1\gamma_{1KE,G}^1 + {}_1\gamma_{1LE,G}^1) \\ &= \frac{1}{4}({}_1\gamma_{1KM,G}^1 + {}_1\gamma_{1KN,G}^1 + {}_1\gamma_{1LM,G}^1 + {}_1\gamma_{1LN,G}^1), \end{aligned}$$

and the parents of G may also be brought in. Expansions involving common ancestors are given in Appendix B.

For the three two-gene-pair and four-gene identity measures, the expansions follow the same pattern as those given by Weir and Cockerham (1969a) for F_{00} until common ancestry becomes involved. We now need to consider digametic, trigametic, and quadrigametic cases. Taking account of the parents of B and C

$$\begin{aligned} F_A^{11} &= \theta_{BC}^{11} = \theta_{B(GH)}^{11} = [(1 + \lambda)/4](\theta_{BG}^{11} + \theta_{BH}^{11}) + [(1 - \lambda)/4](\gamma_{B,GH}^{11} + \gamma_{B,HG}^{11}) \\ &= [(1 + \lambda)^2/16](\theta_{DG}^{11} + \theta_{EG}^{11} + \theta_{DH}^{11} + \theta_{EH}^{11}) \\ &\quad + [(1 - \lambda)^2/8](\gamma_{D,\overline{GH}}^{11} + \gamma_{E,\overline{GH}}^{11} + \gamma_{G,\overline{ED}}^{11} + \gamma_{H,\overline{ED}}^{11}) \\ &\quad + [(1 - \lambda)^2/16](\delta_{DE,GH}^{11} + \delta_{ED,GH}^{11} + \delta_{DE,HG}^{11} + \delta_{ED,HG}^{11}), \end{aligned} \quad (6)$$

where $\gamma_{D,\overline{GH}}^{11} = (\gamma_{D,GH}^{11} + \gamma_{D,HG}^{11})/2$. Taking account of the parents of D and E

$$\begin{aligned} \gamma_{D,EG}^{11} &= \gamma_{(KL),EG}^{11} = [(1 + \lambda)/4](\gamma_{K,EG}^{11} + \gamma_{L,EG}^{11}) + [(1 - \lambda)/4](\delta_{KL,EG}^{11} + \delta_{LK,EG}^{11}) \\ &= [(1 + \lambda)/8](\gamma_{K,MG}^{11} + \gamma_{K,NG}^{11} + \gamma_{L,MG}^{11} + \gamma_{L,NG}^{11}) \\ &\quad + [(1 - \lambda)/8](\delta_{KL,MG}^{11} + \delta_{KL,NG}^{11} + \delta_{LK,MG}^{11} + \delta_{LK,NG}^{11}), \\ \delta_{DE,GH}^{11} &= \delta_{(KL)E,GH}^{11} = \frac{1}{2}(\delta_{KE,GH}^{11} + \delta_{LE,GH}^{11}) \\ &= \frac{1}{4}(\delta_{KM,GH}^{11} + \delta_{KN,GH}^{11} + \delta_{LM,GH}^{11} + \delta_{LN,GH}^{11}), \end{aligned}$$

and the parents of G and H may also be brought in.

While these three general expansions have been written for the two-gene-pair

component X^{11} of each measure, they also apply to the other two-gene-pair and the four-gene identity measures. Each X^{11} is simply replaced by X_{11} , ${}_{11}X$, or X_{11}^{11} . The notational ordering of the genes as pointed out in the introduction must be adhered to strictly. The connection between the joint parental and recombinant quadrigenetic descent measures is ${}_{11}\delta_{DH,GE} = \delta_{DE,GH}^{11}$.

Measures involving common ancestors require special attention. We give them for θ_{BB} here and list the remainder in Appendix B.

$$\begin{aligned}\theta_{11BB} &= \frac{(1+\lambda)^2}{8} F_B^1 + \frac{1-\lambda^2}{2} {}_1F_{1B}^1 + \frac{(1-\lambda)^2}{8} {}_1F_B + \frac{1+\lambda^2}{4} F_{11B}^{11} \\ \theta_{11BB} &= \frac{1+\lambda^2}{4} (1 + F_{11B}) + \frac{1-\lambda^2}{2} F_{1B} \\ \theta_{BB} &= \frac{(1+\lambda)^2}{8} (F_B^1 + F_B^{11}) + \frac{1-\lambda^2}{2} {}_1F_{1B}^1 + \frac{(1-\lambda)^2}{8} ({}_1F_B + {}_{11}F_B) \\ {}_{11}\theta_{BB} &= \frac{(1+\lambda)^2}{8} (F_B^1 + {}_{11}F_B) + \frac{1-\lambda^2}{2} {}_1F_{1B}^1 + \frac{(1-\lambda)^2}{8} ({}_1F_B + F_B^{11}).\end{aligned}\tag{7}$$

With these algorithms, and given the measures among the initial ancestors, one can compute the various measures for any pedigree. For most purposes the initial measure will be taken to be ${}_{00}F_{00}^{11} = 1$.

For individuals of populations that go to fixation we note the limiting values of the measures,

$$F_{1(\infty)} = F_{11(\infty)} = 1, \quad F_{(\infty)}^1 = {}_1F_{(\infty)} = F_{(\infty)}^{11} = {}_{11}F_{(\infty)} = {}_1F_{1(\infty)}^1 = F_{11(\infty)}^{11}.$$

The key to the final results is then $F_{(\infty)}^1$.

If linkage is complete, $\lambda = 1$, then $F_{(t)}^1 = F_{(t)}^{11} = 1$ and the other measures are the same as the one-locus inbreeding coefficient,

$$F_{11(t)}^{11} = {}_1F_{1(t)}^1 = {}_{11}F_{(t)} = F_{11(t)} = {}_1F_{(t)} = F_{1(t)}.$$

EXAMPLES OF SELFING AND SIB MATING

Self-fertilization provides the simplest illustrations of the use of the measures. We begin in the zero generation with ${}_{00}F_{00}^{11} = 1$. Some F 's remain zero. Since all four genes at any time must trace to genes on two original gametes, the tri- and quadrigenetic components of \mathbf{F} are zero in all generations,

$${}_{00}F_{00}^{00} = 2 {}_{00}\tilde{F}_{10}^{00} = 2 {}_{00}\tilde{F}_{00}^{10} = 2 {}_{10}\tilde{F}_{00}^{00} = {}_0F_0^0 = 0.$$

Applying the algorithms the usual results are obtained for the identity by descent measures ($t = 0, 1, 2, \dots$ indexes generations)

$$\begin{aligned} F_{1(t)} &= 1 - \left(\frac{1}{2}\right)^t, \\ F_{11(t)} &= 1 - 2\left(\frac{1}{2}\right)^t + ((1 + \lambda^2)/4)^t, \\ \eta_{11(t)} &= ((1 + \lambda^2)/4)^t - \left(\frac{1}{2}\right)^t. \end{aligned}$$

The other two-gene descent measures must be considered jointly:

$$\begin{aligned} F_{1(t+1)}^1 &= \theta_{1(t)}^1 = [(1 + \lambda)/2]F_{1(t)}^1 + [(1 - \lambda)/2]{}_1F_{1(t)}, \\ {}_1F_{1(t+1)} &= {}_1\theta_{1(t)} = \frac{1}{2}F_{1(t)}^1 + \frac{1}{2}{}_1F_{1(t)}. \end{aligned}$$

Solving these two equations and noting that $F_{1(0)}^1 = 1, {}_1F_{1(0)} = 0$

$$F_{1(t)}^1 = \frac{1}{2 - \lambda} + \left(\frac{\lambda}{2}\right)^t \frac{1 - \lambda}{2 - \lambda}, \quad {}_1F_{1(t)} = \frac{1}{2 - \lambda} \left[1 - \left(\frac{\lambda}{2}\right)^t\right].$$

From these all other measures may be found directly

$$\begin{aligned} F^{11} &= F_{11} + F^1 - F_1, \quad {}_{11}F = F_{11} + {}_1F - F_1, \\ F_{11}^{11} &= (F_{11} + F^1 + {}_1F - 1)/2. \end{aligned}$$

The genotypic frequencies are summarized next for random ancestors,

$$\begin{aligned} P_{ij}^{ij} &= p_{ij}[F_{11} + F_{10}(p_i + q_j) + F_{00} p_{ij}] - {}_{00}F_{11}^{00} \Delta_{ij}, \\ 2 P_{il}^{ij} &= 2 F_{00} p_{ij} p_{il} + F_{10}(q_l p_{ij} + q_j p_{il}), \\ 2 P_{kl}^{ij} &= 2 {}_{00}F_{00}^{11} p_{ij} p_{kl} + 2 {}_{11}F_{00}^{00} p_{il} p_{kj}, \end{aligned}$$

and the final frequencies are

$$P_{ij(\infty)}^{ij} = p_i q_j + F_{1(\infty)}^1 \Delta_{ij}, \quad F_{1(\infty)}^1 = \frac{1}{2 - \lambda}.$$

We may also argue from an initial individual $\mathcal{P}_{22}^{11} = 1$ or a specific pair of gametes with gene frequencies of $\frac{1}{2}$ and $\Delta_{11} = \Delta_{22} = -\Delta_{12} = -\Delta_{21} = \frac{1}{4}$.

We simply note that homozygotes fall into parental and nonparental classes so that

$$\begin{aligned}
 P_{11}^{11}(t) &= P_{22}^{22}(t) = \frac{1}{2} F_{11}^{11}(t) \\
 &= \frac{1}{2} \left[\frac{1}{2-\lambda} - \frac{1}{2-\lambda} \left(\frac{\lambda}{2}\right)^{t+1} - \left(\frac{1}{2}\right)^t + \frac{1}{2} \left(\frac{1+\lambda^2}{4}\right)^t \right], \\
 P_{12}^{12}(t) &= P_{21}^{21}(t) = \frac{1}{2} {}_{00}F_{11}^{00}(t) \\
 &= \frac{1}{2} \left[\frac{1-\lambda}{2-\lambda} + \frac{1}{2-\lambda} \left(\frac{\lambda}{2}\right)^t - \left(\frac{1}{2}\right)^t + \frac{1}{2} \left(\frac{1+\lambda^2}{4}\right)^t \right], \\
 2 P_{22}^{11}(t) &= {}_{00}F_{00}^{11}(t) = \frac{1}{2} \left[\left(\frac{1+\lambda^2}{4}\right)^t + \left(\frac{\lambda}{2}\right)^t \right], \\
 2 P_{21}^{12}(t) &= {}_{11}F_{00}^{00}(t) = \frac{1}{2} \left[\left(\frac{1+\lambda^2}{4}\right)^t - \left(\frac{\lambda}{2}\right)^t \right], \\
 2 P_{12}^{11}(t) &= {}_{01}F_{10}^{10}(t) = \frac{1}{2} \left[\left(\frac{1}{2}\right)^t - \left(\frac{1+\lambda^2}{4}\right)^t \right],
 \end{aligned}$$

the latter one being the same for all four single homozygotes.

We illustrate in terms of gametic linkage disequilibrium the difference in results for the two types of initial ancestor. For random ancestors

$$D_{..}^{ij}(t) = F_{(t)}^1 \Delta_{ij} = \frac{1}{2-\lambda} \left[1 + (1-\lambda) \left(\frac{\lambda}{2}\right)^t \right] \Delta_{ij}, \quad D_{..}^{ij}(\infty) = \frac{\Delta_{ij}}{2-\lambda},$$

while for an initial pair of gametes,

$$D_{..}^{ij}(t) = \left[\frac{\lambda}{2-\lambda} + \frac{2(1-\lambda)}{2-\lambda} \left(\frac{\lambda}{2}\right)^t \right] \Delta_{ij}, \quad D_{..}^{ij}(\infty) = \frac{\lambda}{2-\lambda} \Delta_{ij}$$

(Wright, 1933). The difference is due to the linkage disequilibrium among final subpopulations from different initial founders. For random parents one-half of the original disequilibrium is fixed even with free recombination.

The mating scheme for sib mating is shown in Fig. 2. The algorithms are used to determine the summary components of the general descent measure $F_{(t)}$ in the t th generation.

Starting with the two-gene summary components, it is well known that

$$F_{1(t+2)}^1 = \frac{1}{4} + \frac{1}{2} F_{1(t+1)}^1 + \frac{1}{4} F_{1(t)}^1, \quad F_{1(0)}^1 = F_{1(1)}^1 = 0.$$

From Eq. (1)

$$F_{(t+2)}^1 = [(1+\lambda)/2] F_{(t+1)}^1 + [(1-\lambda)/2] F_{(t+1)}^1,$$

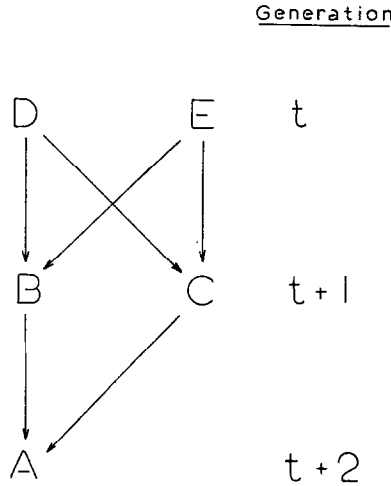


FIGURE 2

while from Eq. (2) by identifying the parents of *C* with those of *B*

$${}_1F_A = \frac{1}{4}({}_1\theta_{DD} + {}_1\theta_{DE} + {}_1\theta_{ED} + {}_1\theta_{EE}) = \frac{1}{2}({}_1\theta_{DD} + {}_1F_B),$$

since both members of each generation are equally inbred. Substituting the value of ${}_1\theta_{DD}$ from Eq. (3)

$${}_1F_A = \frac{1}{2}({}_1F_B + (F_D^1 + {}_1F_D)/2) \quad \text{or} \quad {}_1F_{(t+2)} = \frac{1}{2} {}_1F_{(t+1)} + \frac{1}{4} {}_1F_{(t)} + \frac{1}{4} F_{(t)}^1.$$

Separate equations for the parental and recombinant descent coefficients can now be found

$$F_{(t+2)}^1 = \frac{1}{4} + (\lambda/2)F_{(t+1)}^1 + (\lambda/4)F_{(t)}^1, \quad F_{(0)}^1 = 1, \quad F_{(1)}^1 = (1 + \lambda)/2,$$

$${}_1F_{(t+2)} = \frac{1}{4} + (\lambda/2){}_1F_{(t+1)} + (\lambda/4){}_1F_{(t)}, \quad {}_1F_{(0)} = {}_1F_{(1)} = 0.$$

The three-gene component ${}_1F_1^1$ requires knowledge of three other three-gene measures. From Eq. (4)

$${}_1F_{1A}^1 = [(1 + \lambda)/4]({}_1F_{1B}^1 + \theta_{1DD}^1) + [(1 - \lambda)/4]({}_1\gamma_{1DD,E}^1 + {}_1\gamma_{1DE,D}^1),$$

which leads us to introduce the notation

$${}_1Q_{1(t+1)}^1 = {}_1\theta_{1DD}^1 = {}_1\theta_{1EE}^1, \quad {}_1R_{1(t+1)}^1 = {}_1\gamma_{1DD,E}^1 = {}_1\gamma_{1EE,D}^1,$$

$${}_1S_{1(t+1)}^1 = {}_1\gamma_{1DE,D}^1 = {}_1\gamma_{1ED,E}^1,$$

so that

$${}_1F_{1(t+2)}^1 = [(1 + \lambda)/4]({}_1F_{1(t+1)}^1 + {}_1Q_{1(t+1)}^1) + [(1 - \lambda)/4]({}_1R_{1(t+1)}^1 + {}_1S_{1(t+1)}^1).$$

The transition equation for ${}_1Q_{1(t)}^1$ follows from Eq. (5) and those for ${}_1R_{1(t)}^1$ and ${}_1S_{1(t)}^1$ from Appendix B. They are listed in Appendix C. While two-gene summary components were expressed in terms of two-gene components only, three-gene transition equations require three-gene and two-gene components.

The methods for determining $F_{11(t)}$ are given in Weir and Cockerham (1969a) and the details in Cockerham and Weir (1968). The remaining three summary components are written as $\bar{F}'_{(t)} = (F_{11(t)}^{11}, F_{(t)}^{11}, {}_{11}F_{(t)})$ and from Eq. (6)

$$\begin{aligned} \bar{F}'_A = & [(1 + \lambda)^2/8](\bar{F}_B + \bar{\theta}_{DD}) + [(1 - \lambda^2)/2]\bar{\gamma}_{D,DE} \\ & + [(1 - \lambda)^2/8](\delta_{DE,DE} + \delta_{DE,ED}), \end{aligned}$$

which, after defining

$$\begin{aligned} \bar{Q}_{(t+1)} &= \bar{\theta}_{DD} = \bar{\theta}_{EE}, & \bar{R}_{(t+1)} &= \bar{\gamma}_{D,DE} = \bar{\gamma}_{E,DE}, \\ \bar{S}_{(t+1)} &= \bar{\gamma}_{D,EE} = \bar{\gamma}_{E,DD}, & \bar{W}_{(t+1)} &= \delta_{DD,EE} = \delta_{EE,DD}, \\ \bar{T}_{(t+1)} &= \frac{1}{2}(\delta_{DE,DE} + \delta_{DE,ED}) = \frac{1}{2}(\delta_{ED,ED} + \delta_{ED,DE}), \end{aligned}$$

is written as

$$\bar{F}'_{(t+2)} = [(1 + \lambda)^2/8](\bar{F}'_{(t+1)} + \bar{Q}_{(t+1)}) + [(1 - \lambda^2)/2]\bar{R}_{(t+1)} + [(1 - \lambda)^2/4]\bar{T}_{(t+1)}.$$

The equation for $\bar{Q}_{(t)}$ follows from Eqs. (7), and those for \bar{R} , \bar{S} , \bar{T} , and \bar{W} from Appendix B. They are all listed in Appendix C. Four-gene measure transitions involve two-, three-, and four-gene measures.

For the initial values ${}_{00}F_{00(0)}^{11} = 1$ the values of the measures introduced into the transition equations are given in Appendix D for $t = 1$, and the values of \mathbf{F}^* are given in Appendix E for the first three generations. The progress towards equilibrium is indicated, for $\lambda = 0.0, 0.5, 0.9, 1$, in Fig. 3.

From the final value, $F_{(\infty)}^1 = 1/(4 - 3\lambda)$, the final linkage disequilibria for random pairs of initial ancestors are found to be

$$D_{..(\infty)}^{ij} = \frac{A_{ij}}{4 - 3\lambda},$$

and the final genotypic frequencies are

$$P_{ij(\infty)}^{ij} = P_{..(\infty)}^{ij} = p_i q_j + \frac{A_{ij}}{4 - 3\lambda}.$$

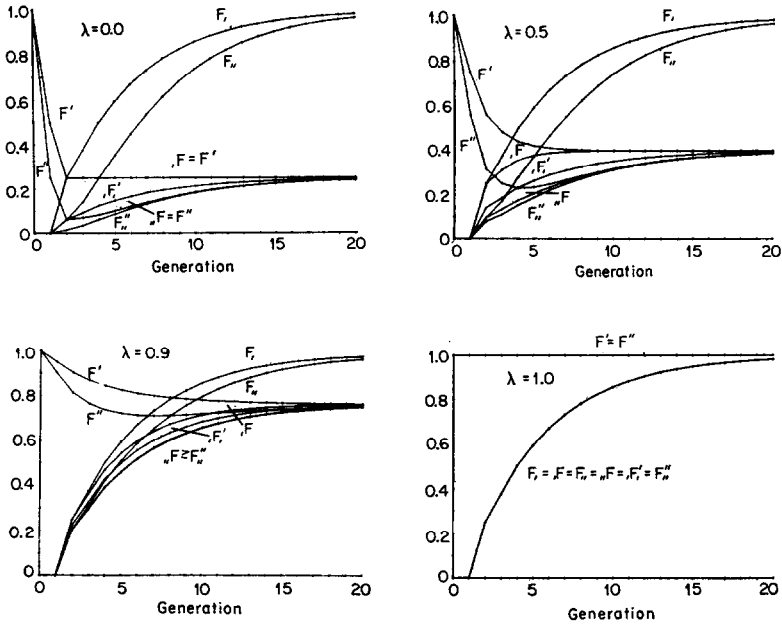


FIG. 3. Sib mating summary measures for $\lambda = 0.0, 0.5, 0.9, 1.0$.

However, for four specific initial gametes, $\mathcal{D}_{ij}^i = -\Delta_{ij}/3$ and

$$D_{..(\infty)}^{ij} = \frac{\lambda \Delta_{ij}}{4 - 3\lambda},$$

there being no linkage disequilibrium with free recombination. Thus, for the two initial genotypes $\mathcal{P}_{22}^{11} \times \mathcal{P}_{44}^{33}$, the final population is

$$P_{ii(\infty)}^{ii} = \frac{1}{4(4 - 3\lambda)} \quad P_{ij(\infty)}^{ij} = \frac{1 - \lambda}{4(4 - 3\lambda)} \quad \begin{matrix} i, j = 1, 2, 3, 4 \\ i \neq j. \end{matrix}$$

If instead the initial pair were double homozygotes $\mathcal{P}_{11}^{11} \times \mathcal{P}_{22}^{22}$ the final population would be

$$P_{11(\infty)}^{11} = P_{22(\infty)}^{22} = (2 - \lambda)/2(4 - 3\lambda) \quad P_{12(\infty)}^{12} = P_{21(\infty)}^{21} = (1 - \lambda)/(4 - 3\lambda)$$

as given by Haldane and Waddington (1931).

These final results do not depend on the constitution of the initial population other than a fixed set of gametes which is demonstrated in Appendix F.

DISCUSSION

To aid the development of two-locus genetic models we have defined a two-locus descent measure. For any four genes, two at each of two loci, the measure gives the probabilities of the possible arrangements of the genes of which these four are copies on a set of initial gametes. For an individual the descent measure is defined for four genes on two uniting gametes. It is the argument back to initial gametes which is the key to the whole development, and which points out the departure from other procedures (e.g. Kimura, 1963). When two genes at the same locus are copies of genes on one initial gamete, they are identical by descent so that the introduction of two-locus descent measures serves to extend the one-locus work of Malécot (1948) and the two-locus work of Haldane (1949) and Schnell (1961). It subsumes our earlier work (Cockerham and Weir, 1968; Weir and Cockerham, 1969a,b). Descent measures are defined as probabilities and simple probability arguments are used throughout.

While there are 15 possible arrangements of the antecedents of four genes, it was shown that there are nine distinct valued components of individual descent measures. Individual measures are digametic, and use is made of trigametic and quadrigametic measures.

The major use of individual descent measures is in the expression of genotypic frequencies. For systems other than random mating, the traditional approach has been to use mating type frequencies to find two-locus genotypic frequencies (e.g. Haldane and Waddington, 1931). This approach quickly becomes cumbersome as the number of alleles per locus increases and the degree of inbreeding decreases. Once descent measures have been evaluated for a pedigree, the structure of any generation can be related to that of an initial population. In particular the frequencies of various pairs of uniting gametes can, by use of individual descent measures, be expressed in terms of initial gametic frequencies. With the nine distinct components of the individual descent measure, all $mn(mn + 1)/2$ genotypic frequencies for m and n alleles, respectively, at two loci can be expressed in terms of mn gametic frequencies. In this work we have assumed that either the frequencies of initial gametes are specified or that they are drawn from an infinite random mating population of known composition. Genotypic frequencies in later generations can then be given as functions of descent measures which are specified by the pedigree.

With genotypic frequencies known, it is a simple matter to find disequilibria functions. In general these functions are viewed as departures of joint probabilities of events from those expected when events are independent. A complete listing of disequilibria is given. For example, such quantities as linkage disequilibrium, difference between coupling and repulsion double heterozygote frequencies, and deviations of two-locus genotypic frequencies from products

of corresponding one-locus or gametic frequencies may be expressed explicitly in terms of descent measures and initial gametic frequencies. It is then possible to give precise expressions for the dependence of events at two loci and discuss for example the interaction of linkage and inbreeding or the presence of disturbing forces such as selection in populations of known pedigree.

The evaluation of descent measures follows a well defined algorithm. It was found to be simpler to work with eight summary components of the components of the individual descent measure. Two of these were the one- and two-locus inbreeding coefficients. Evaluation of the latter has been completely documented (Weir and Cockerham, 1969a,b) and need not be considered further. Among the remaining components there are three classes: two are digenic (parental and recombinant descent coefficients, respectively), one trigenic, and three quadrigenic. All are digametic. Each class is evaluated separately and complete details for obtaining transition equations between successive generations are given. Trigametic and quadrigametic measures need to be used in these transition equations, all of which are linear and in a convenient form for computer work. They appear simpler than those resulting from the work of Narain (1966).

The use of descent measures was illustrated for selfing and for sib mating. Selfing at two loci has been studied extensively (e.g. Wright, 1933 and Narain, 1965) and we were able to duplicate known results. Previous results for sib mating (see Cockerham and Weir, 1968 for review) have been limited mainly to discussions of equilibrium populations for two alleles per locus. To find the individual descent measure we needed 30 linear one-step transition equations each involving two, four or six variables. We stress that once the descent measure is found, genotypic frequencies for any number of alleles per locus follow. Graphs of the summary components of the inbreeding measure were displayed in Fig. 3. Sib mating populations have substantially reached inbreeding equilibrium by the twentieth generation. Initial equality of F_1 and ${}_1F$ indicates that two nongametic genes, whether or not at the same locus, are equally likely to have arisen from one initial gamete. As linkage decreases these values diverge further. The point of intersection of F_1 and F^1 indicates the time at which genes at the same locus and genes on the same gamete are equally likely to have arisen from one initial gamete. This point becomes later in time as linkage increases. Until $\lambda = \frac{2}{3}$ there is greater ultimate probability that four uniting genes arose from two than one initial gamete. For λ greater than $\frac{2}{3}$ the situation is reversed.

It should be clear by now that the effects of linkage with inbreeding are of two varieties: recombination between the genes and linkage disequilibrium. The effects of the linkage parameter are entirely on the descent measures which in conjunction with the initial linkage disequilibrium and gene frequencies determine the genotypic frequencies. With no initial linkage disequilibrium only the one- and two-locus inbreeding coefficients are required. The result is

a transient one and is to increase the frequencies of double heterozygotes and double homozygotes by an amount proportional to the identity disequilibrium and to lower the frequencies of single heterozygotes correspondingly. It is transient because the identity disequilibrium goes to zero with fixation.

The principal effects of linkage disequilibrium can be abstracted out in terms of gametic linkage disequilibrium; this being what has been generally studied in the past. This disequilibrium is given entirely by F^1 and the initial disequilibrium. The parental descent measure is the important counterpart for two nonalleles on the same gamete of the inbreeding coefficient for two alleles, and the parental descent measure alone in addition to $F_{1(\infty)} = F_{11(\infty)} = 1$ is required for final population frequencies. The other important counterpart for nonalleles is the recombinant descent measure. The two together provide all of the information on gametic linkage disequilibrium within and among subpopulations. That within subpopulations is

$$D_w^{ij} = D_{..}^{ij} - D_{.j}^{i.} = (F^1 - {}_1F) \Delta_{ij}$$

for a random sample of founders and $\hat{D}_w^{ij} = (F^1 - {}_1F) \hat{\Delta}_{ij} 2N / (2N - 1)$ for a specific initial set of gametes, and in each case goes to zero with fixation, $F_{(\infty)}^1 = {}_1F_{(\infty)}$. We shall use hats in this comparative discussion to distinguish specific sets of gametes. When specific sets of gametes are considered to be a random sample from the infinite randomly mating population $\mathcal{E} \hat{\Delta}_{ij} = \Delta_{ij} (2N - 1) / 2N$ and $D_w^{ij} = \mathcal{E} \hat{D}_w^{ij}$ (\mathcal{E} denotes expectation or average over random samples).

The final gametic linkage disequilibrium is the total linkage disequilibrium among subpopulations at fixation. It is $D_{..(\infty)}^{ij} = F_{(\infty)}^1 \Delta_{ij}$ for the random founders and $\hat{D}_{..(\infty)}^{ij} = [F_{(\infty)}^1 2N / (2N - 1) - 1 / (2N - 1)] \hat{\Delta}_{ij}$ for the specific set of gametes.

It was necessary to introduce the idea of averaging over randomizations of the initial set of gametes to arrive at transient frequencies and without employing additional descent measures such as $F_{(a)}^1$ and $F_{(b)}^1$ in Appendix F. However, the results in Appendix F suggest that for any pedigree system of mating the final frequencies are the same for all randomizations or sets of initial parents for a specific set of initial gametes.

If we consider initial sets of parents to be random samples from the infinite randomly mating population, then we can decompose the total final linkage disequilibrium, $D_{..(\infty)}^{ij}$, into that, $\mathcal{E} \hat{D}_{..(\infty)}^{ij} = (F_{\infty}^1 - 1/2N) \Delta_{ij}$, among final subpopulations from the same initial ancestors and the remainder, $D_{..(\infty)}^{ij} - \mathcal{E} \hat{D}_{..(\infty)}^{ij} = \Delta_{ij} / 2N$, among final subpopulations from different initial ancestors. The latter is, of course, just the covariance of the genes in a sample of $2N$ gametes.

To illustrate also the connection between the two initial populations for frequencies, consider

$$\hat{P}_{..}^{ij} = \hat{p}_i \hat{q}_j + \left(F^1 \frac{2N}{2N-1} - \frac{1}{2N-1} \right) \Delta_{ij}.$$

If the initial gametes are a random sample then $\mathcal{E}\hat{p}_i \hat{q}_j = p_i q_j + \Delta_{ij}/2N$ so that $\mathcal{E}\hat{P}_{..}^{ij} = P_{..}^{ij} = p_i q_j + F^1 \Delta_{ij}$. While we have treated the two sets of founder conditions separately, one can always produce the results for a random sample of parents from those for a specific set of initial gametes by letting the set of gametes be a random sample from the parent population.

Fixation probabilities or frequencies are given simply by $P_{..}^{ij(\infty)}$ or $\hat{P}_{..}^{ij(\infty)}$.

Since we can express genotypic frequencies in terms of \mathbf{F}^* and the initial frequencies, we can, of course, apply these to a quantitative model of gene effects for two loci to produce means and variances. Thus, we can quantify the effects of inbreeding and linkage on inbreeding depression and on the genetic variance among individuals.

We have concentrated on the digametic descent measures \mathbf{F}^* and have mainly employed the trigametic γ^* and quadrigametic δ^* ones in the general algorithm. They, of course, provide important information. One application is in the determination of the variance of linkage disequilibrium. Other applications are to be explored.

Also, we have concentrated on the \mathbf{F}^* measures as they apply to the descension of neutral genes. It should be apparent, however, that these measures in conjunction with gene frequencies form a model for the decomposition of an arbitrary set of genotypic frequencies into various marginal and joint frequencies in terms of correlation and higher order parameters. For example, in the analysis of alleles the inbreeding coefficient is defined and utilized as the correlation of alleles within individuals, whatever are the causes of the correlation. Similarly, F^1 and ${}_1F$ may be adapted to serve as correlations of pairs of nonalleles. The other parameters are higher moment ones, three-gene, two-gene-pair, and four-gene. All together they provide a parametric model for the analysis of data on genotypic frequencies. Alternatively, the disequilibria functions may have certain advantages as a model. A complete analysis requires a complete classification of double heterozygotes, but this is a problem regardless of the model.

APPENDIX A

The relation between the condensed set of general measures,

$$\tilde{\mathbf{F}}' = (F_{11}^{11}, {}_{00}F_{11}^{00}, {}_{00}F_{00}^{11}, F_{00}^{00}, 4 {}_{01}\tilde{F}_{10}^{10}, 2 {}_{00}\tilde{F}_{10}^{00}, 2 {}_{00}\tilde{F}_{00}^{10}, 2 {}_{10}\tilde{F}_{00}^{00}, F_{00}^{00}),$$

and the set of summary measures,

$$\mathbf{F}^{*'} = (F_{11}^{11}, F_{11}, F^{11}, {}_{11}F, {}_1F_1^1, F_1, F^1, {}_1F, 1),$$

is $\mathbf{F}^* = Z\tilde{\mathbf{F}}$ or $\tilde{\mathbf{F}} = Z^{-1}\mathbf{F}^*$, where

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 1 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$Z^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -4 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 4 & -2 & 0 & 0 & -4 & 2 & 0 & 0 & 0 \\ 4 & 0 & -2 & 0 & -4 & 0 & 2 & 0 & 0 \\ 4 & 0 & 0 & -2 & -4 & 0 & 0 & 2 & 0 \\ -6 & 1 & 1 & 1 & 8 & -2 & -2 & -2 & 1 \end{bmatrix}$$

APPENDIX B

Expansions of Measures Involving Common Ancestors

We write the parents of individuals B, C, D , etc. as B_i, C_j, D_k , etc. ($i, j, k = 1, 2$), respectively. A subscript, such as i , in a measure implies that the measure is to be written for every value of the subscript. For example (primes denote distinct values of the same subscript)

$$\gamma_{B_i B_{i'}, C_j} = \gamma_{B_1 B_2, C_1} + \gamma_{B_2 B_1, C_1} + \gamma_{B_1 B_2, C_2} + \gamma_{B_2 B_1, C_2}.$$

There are three trigametic three-gene identity measures with common ancestors,

$$\begin{aligned} {}_1\gamma_{1BC, B} &= \frac{1}{8}({}_1\theta_{1B_i C_j}^1 + {}_1\gamma_{1B_i C_j, B_{i'}}^1), \\ {}_1\gamma_{1BB, C} &= \frac{1}{8}({}_1\theta_{B_i C_j}^1 + {}_1\gamma_{1B_i B_{i'}, C_j}^1), \\ {}_1\gamma_{1BB, B} &= \frac{1}{4}(F_B^1 + {}_1F_B) + \frac{1}{2} {}_1F_{1B}^1. \end{aligned}$$

We have already (Weir and Cockerham, 1969a) given an algorithm for finding X_{00} (and, hence, X_{11}) for the common ancestor cases. For the other two two-gene-pair and four-gene identity measures we introduce the notation

$$\bar{X} = \begin{bmatrix} X_{11}^{\text{II}} \\ X_{11}^{\text{I}} \\ {}_{11}X \end{bmatrix} \quad U = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

and list the expansions for common ancestors

$$\begin{aligned} \bar{\gamma}_{B,\overline{BC}} &= [(1 + \lambda)/16](U_1 \theta_{1B_i C_j}^1 + I \bar{\gamma}_{B_i, \overline{B_i C_j}}) \\ &\quad + [(1 - \lambda)/16](V \bar{\gamma}_{B_i, \overline{B_i C_j}} + U_1 \gamma_{1B_i C_j, B_i}^1), \\ \bar{\gamma}_{B,CC} &= [(1 + \lambda)/16](I \bar{\theta}_{B_i C_j} + I \bar{\gamma}_{B_i, C_j C_j'}) \\ &\quad + [(1 - \lambda)/16](I \bar{\gamma}_{C_j, B_i B_i'} + I \bar{\delta}_{B_i B_i', C_j C_j'}), \\ \bar{\gamma}_{B,BB} &= [(1 + \lambda)/8](U F_B^1 + I \bar{F}_B) + \frac{1}{2} U_1 F_{1B}^1 + [(1 - \lambda)/8](U_1 F_B + V \bar{F}_B), \\ \bar{\delta}_{BC,DB} &= (1/16)(V \bar{\gamma}_{B_i, D_k C_j} + I \bar{\delta}_{B_i D_k, B_i C_j}), \\ \bar{\delta}_{CB,BD} &= (1/16)(V \bar{\gamma}_{B_i, C_j D_k} + I \bar{\delta}_{C_j B_i, B_i D_k}), \\ \bar{\delta}_{BC,BD} &= (1/16)(U_1 \gamma_{1C_j D_k, B_i}^1 + I \bar{\delta}_{B_i C_j, B_i D_k}), \\ \bar{\delta}_{CB,DB} &= (1/16)(U_1 \gamma_{1C_j D_k, B_i}^1 + I \bar{\delta}_{C_j B_i, D_k B_i'}), \\ \bar{\delta}_{BB,CD} &= (1/16)(I \bar{\gamma}_{B_i, C_j D_k} + I \bar{\delta}_{B_i B_i', C_j D_k}), \\ \bar{\delta}_{CD,BB} &= (1/16)(I \bar{\gamma}_{B_i, C_j D_k} + I \bar{\delta}_{C_j D_k, B_i B_i'}), \\ \bar{\delta}_{BC,CB} &= (1/16)(V \bar{\theta}_{B_i C_j} + V \bar{\gamma}_{B_i, C_j C_j'} + V \bar{\gamma}_{C_j, B_i B_i'} + I \bar{\delta}_{B_i B_i', C_j C_j'}), \\ \bar{\delta}_{BC,BC} &= (1/16)(U_1 \theta_{B_i C_j} + U_1 \gamma_{1B_i B_i', C_j}^1 + U_1 \gamma_{1C_j C_j', B_i}^1 + I \bar{\delta}_{B_i C_j, B_i C_j'}), \\ \bar{\delta}_{BB,CC} &= (1/16)(I \bar{\theta}_{B_i C_j} + I \bar{\gamma}_{B_i, C_j C_j'} + I \bar{\gamma}_{C_j, B_i B_i'} + I \bar{\delta}_{B_i B_i', C_j C_j'}), \\ \bar{\delta}_{BB,BC} &= (1/16)(U_1 \theta_{1B_i C_j}^1 + (I + V) \bar{\gamma}_{B_i, B_i' C_j} + U_1 \gamma_{1B_i C_j, B_i'}^1), \\ \bar{\delta}_{BB,CB} &= (1/16)(U_1 \theta_{1B_i C_j}^1 + (I + V) \bar{\gamma}_{B_i, C_j B_i'} + U_1 \gamma_{1B_i C_j, B_i'}^1), \\ \bar{\delta}_{BB,BB} &= \frac{1}{8}(U F_B^1 + I \bar{F}_B) + \frac{1}{2} U_1 F_{1B}^1 + \frac{1}{8}(U_1 F_B + V \bar{F}_B). \end{aligned}$$

APPENDIX C

Sib Mating Transition Equations

The four three-gene identity measures required are related by the equations,

$$\begin{aligned} {}_1F_{1(t+1)}^1 &= [(1 + \lambda)/4]({}_1F_{1(t)}^1 + {}_1Q_{1(t)}^1) + [(1 - \lambda)/4]({}_1R_{1(t)}^1 + {}_1S_{1(t)}^1), \\ {}_1Q_{1(t+1)}^1 &= [(1 + \lambda)/4]F_{(t)}^1 + \frac{1}{2} {}_1F_{1(t)}^1 + [(1 - \lambda)/4]F_{(t)}^1, \\ {}_1R_{1(t+1)}^1 &= \frac{1}{4}({}_1F_{(t)}^1 + {}_1Q_{(t)}^1) + \frac{1}{2} {}_1S_{1(t)}^1, \\ {}_1S_{1(t+1)}^1 &= \frac{1}{4}({}_1F_{1(t)}^1 + {}_1Q_{1(t)}^1 + {}_1R_{1(t)}^1 + {}_1S_{1(t)}^1). \end{aligned}$$

The six sets of two-gene-pair and four-gene identity measures required are related by the equations,

$$\begin{aligned} \bar{F}_{(t+1)} &= [(1 + \lambda)^2/8]I(\bar{F}_{(t)} + \bar{Q}_{(t)}) + [(1 - \lambda^2)/2]I \bar{R}_{(t)} + [(1 - \lambda)^2/4]I \bar{T}_{(t)}, \\ \bar{Q}_{(t+1)} &= [(1 + \lambda)^2/8](U F_{(t)}^1 + I \bar{F}_{(t)}) + [(1 - \lambda^2)/2]U {}_1F_{1(t)}^1 \\ &\quad + [(1 - \lambda)^2/8](U {}_1F_{(t)} + V \bar{F}_{(t)}), \\ \bar{R}_{(t+1)} &= [(1 + \lambda)/8]U({}_1F_{1(t)}^1 + {}_1Q_{1(t)}^1) + [(1 + \lambda)/8]I(\bar{R}_{(t)} + \bar{S}_{(t)}) \\ &\quad + [(1 - \lambda)/8]U({}_1R_{1(t)}^1 + {}_1S_{1(t)}^1) + [(1 - \lambda)/8]V(\bar{R}_{(t)} + \bar{S}_{(t)}), \\ \bar{S}_{(t+1)} &= [(1 + \lambda)/8]I(\bar{F}_{(t)} + \bar{Q}_{(t)}) + \frac{1}{2}I \bar{R}_{(t)} + [(1 - \lambda)/4]I \bar{T}_{(t)}, \\ \bar{T}_{(t+1)} &= (1/16) U({}_1F_{(t)} + {}_1Q_{(t)}) + (1/16) V(\bar{F}_{(t)} + \bar{Q}_{(t)}) + (1/4) V \bar{R}_{(t)} \\ &\quad + \frac{1}{4}U {}_1S_{1(t)}^1 + \frac{1}{8}I(\bar{T}_{(t)} + \bar{W}_{(t)}), \\ \bar{W}_{(t+1)} &= \frac{1}{8}I(\bar{F}_{(t)} + \bar{Q}_{(t)}) + \frac{1}{2}I \bar{R}_{(t)} + \frac{1}{4}I \bar{T}_{(t)}, \end{aligned}$$

where V and U were defined in Appendix B.

APPENDIX D

Initial Values ($t = 1$) of Measures Required in Evaluation of \mathbf{F}^ for Sib Mating Three-Gene Measures*

$${}_1Q_{1(1)}^1 = (1 + \lambda)/4 \quad {}_1R_{1(1)}^1 = 0 \quad {}_1S_{1(1)}^1 = 0$$

Two-Gene-Pair and Four-Gene Measures

$$\bar{Q}_{(t)} = \begin{bmatrix} (1 + \lambda)^2/8 \\ (1 + \lambda)^2/4 \\ (1 + \lambda^2)/4 \end{bmatrix} \quad \bar{R}_{(t)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \bar{S}_{(t)} = \begin{bmatrix} 0 \\ (1 + \lambda)/4 \\ 0 \end{bmatrix}$$

$$\bar{T}_{(t)} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{8} \end{bmatrix} \quad \bar{W}_{(t)} = \begin{bmatrix} 0 \\ \frac{1}{4} \\ 0 \end{bmatrix}$$

APPENDIX E

Values in First Three Generations of Inbreeding Measures F for Sib Mating*

	$t = 1$	$t = 2$	$t = 3$
F_{11}^{11}	0	$(1 + \lambda)^4/64$	$(1/512)(18 + 34\lambda + 57\lambda^2 + 44\lambda^3 + 28\lambda^4 + 10\lambda^5 + \lambda^6)$
F_{11}	0	$(2 + 3\lambda^2 + 2\lambda^3 + \lambda^4)/32$	$(1/256)(36 + 16\lambda^2 + 22\lambda^3 + 17\lambda^4 + 4\lambda^5 + \lambda^6)$
F^{11}	$(1 + \lambda)^2/4$	$(1 + \lambda)^4/16$	$(1/256)(18 + 34\lambda + 71\lambda^2 + 62\lambda^3 + 48\lambda^4 + 20\lambda^5 + 3\lambda^6)$
${}_{11}F$	0	$(2 + 3\lambda^2 + 2\lambda^3 + \lambda^4)/32$	$(1/256)(18 + 18\lambda + 15\lambda^2 + 16\lambda^3 + 19\lambda^4 + 8\lambda^5 + 2\lambda^6)$
${}_1F_1^1$	0	$(1 + \lambda)^2/16$	$(1/64)(6 + 7\lambda + 8\lambda^2 + 3\lambda^3)$
F_1	0	$\frac{1}{4}$	$\frac{3}{8}$
F^1	$(1 + \lambda)/2$	$(1 + \lambda)^2/4$	$\frac{1}{8}(2 + 2\lambda + 3\lambda^2 + \lambda^3)$
${}_1F$	0	$\frac{1}{4}$	$(2 + \lambda)/8$

APPENDIX F

Final Recombinant Gametic Frequencies from a Fixed Set of Initial Gametes with Sib Mating

A gamete received by an individual in generation t has three possible origins. With probability $F_{(t)}^1$ it carries genes which were on one initial gamete. It carries genes which were on two different gametes in one initial parent with

probability $F_{a(t)}^1$, and with probability $F_{b(t)}^1$ it carries two genes from gametes in two different parents. They sum to one, $F_{(t)}^1 + F_{a(t)}^1 + F_{b(t)}^1 = 1$, and each satisfies the recurrence relation

$$F_{(t+3)}^1 = [(2 + \lambda)/2]F_{(t+2)}^1 - (\lambda/4)F_{(t+1)}^1 - (\lambda/4)F_t^1.$$

Their initial values differ ($F_{(0)}^1 = 1$, $F_{a(0)}^1 = F_{b(0)}^1 = 0$; $F_{(1)}^1 = (1 + \lambda)/2$, $F_{a(1)}^1 = (1 - \lambda)/2$, $F_{b(1)}^1 = 0$) and lead to the following solutions for $\lambda \neq 0$:

$$F_{(t)}^1 = \frac{1}{4 - 3\lambda} + \frac{1 - \lambda}{(4 - 3\lambda)(r_1 - r_2)} \{(1 - \lambda + 3r_1)r_1^t - (1 - \lambda + 3r_2)r_2^t\},$$

$$F_{a(t)}^1 = \frac{1 - \lambda}{4 - 3\lambda} + \frac{1 - \lambda}{(4 - 3\lambda)(r_1 - r_2)} \{(1 - r_1)r_1^t - (1 - r_2)r_2^t\},$$

$$F_{b(t)}^1 = \frac{2(1 - \lambda)}{4 - 3\lambda} - \frac{1 - \lambda}{(4 - 3\lambda)(r_1 - r_2)} \{(2 - \lambda + 2r_1)r_1^t - (2 - \lambda + 2r_2)r_2^t\},$$

where $r_1 = \frac{1}{4}(\lambda + (\lambda(4 + \lambda))^{1/2})$, $r_2 = \frac{1}{4}(\lambda - (\lambda(4 + \lambda))^{1/2})$. For $\lambda = 0$; $F_{(t)}^1 = F_{a(t)}^1 = \frac{1}{2}$, $F_{b(t)}^1 = \frac{1}{2}$ for $t \geq 2$. There are twice as many (8) gametes in the class for F_b^1 as for F_a^1 so that all recombinant gametes have the same final frequency. The two classes of recombinant gametes do differ in frequencies in the early stages of descension.

REFERENCES

- COCKERHAM, C. C. AND WEIR, B. S. 1968. Sib mating with two linked loci, *Genetics* **60**, 629-640.
- HALDANE, J. B. S. 1949. The association of characters as a result of inbreeding and linkage, *Ann Eugenics* **15**, 15-23.
- HALDANE, J. B. S., AND WADDINGTON, C. H. 1931. Inbreeding and linkage, *Genetics* **10**, 357-374.
- KIMURA, M. 1963. A probability method for treating inbreeding systems, especially with linked genes, *Biometrics* **19**, 1-17.
- MALÉCOT, G. 1948. "Les mathématiques de l'hérédité," Masson, Paris.
- NARAIN, P. 1965. Homozygosity in a selfed population with an arbitrary number of linked loci, *J. Genet.* **59**, 254-266.
- NARAIN, P. 1966. Effect of linkage on homozygosity of a population under mixed selfing and random mating, *Genetics* **54**, 363-374.
- SCHNELL, F. W. 1961. Some general formulations of linkage effects in inbreeding, *Genetics* **46**, 947-957.
- WEIR, B. S., AND COCKERHAM, C. C. 1969a. Pedigree mating with two linked loci, *Genetics* **61**, 923-940.
- WEIR, B. S., AND COCKERHAM, C. C. 1969b. Group inbreeding with two linked loci, *Genetics* **63**, 711-742.
- WRIGHT, S. 1922. Coefficients of inbreeding and relationship, *Amer. Natur.* **56**, 330-338.
- WRIGHT, S. 1933. Inbreeding and recombination, *Proc. Nat. Acad. Sci. U.S.A.* **19**, 420-433.